NBER WORKING PAPER SERIES

AI-POWERED TRADING, ALGORITHMIC COLLUSION, AND PRICE EFFICIENCY

Winston Wei Dou
Itay Goldstein
Yan Ji

Working Paper 34054
http://www.nber.org/papers/w34054

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025

We thank Tobias Adrian, Kerry Back, Snehal Banerjee, Hui Chen, Jean-Edouard Colliard, Will Cong, Antoine Didisheim, Itamar Drechsler, Maryam Farboodi, Slava Fos, Paolo Fulghieri, Joao Gomes, Mark Grinblatt, Ming Guo, Wei Jiang, Chris Jones, Scott Joslin, Joe Harrington, Larry Harris, Zhiguo He, 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, Paul Romer, Nick Roussanov, Tom Sargent, Antoinette Schoar, Hyun-Song Shin, Rob Stambaugh, Eric Talley, Anton Tsoy, Stijn Van Nieuwerburgh, Dimitri Vayanos, Laura Veldkamp, Jiang Wang, Neng Wang, Xian Wu, Liyan Yang, Jacob Yunger, David Zhang, and seminar and conference participants at AsianFA, ASU Sonoran Winter Finance Conference, BIS, BIS Meeting of Heads of Financial Stability, BI-SHoF Conference, Boston College, CFTRC, CFEA, CICF, CMU, Columbia, Cubist Systematic Strategies (Point72), CUFE, Duke/UNC Asset Pricing Conference, EFA, Fed Board, FINRA, FIRS, Florida International University, FMA Asia/Pacific Conference, Frankfurt School of Finance and Management, Fudan, George Mason, Harvard University, HKU, HKUST, HK Conference for Fintech and AI, IESE Barcelona Workshop on AI in Finance, 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, NTU Conference on AI for Finance, NYU/Penn Law and Finance Conference, OECD, Olin Finance Conference at WashU, OSU, Oxford, PKU, PKU/PHBS Sargent Institute Macro-Finance Workshop, QES Global Quant and Macro Investing Conference, QRFE Workshop on Market Microstructure, Fintech and AI, Renmin University, Rice University, Shanghai Jiao Tong University (SAIF & Antai), SFS Cavalcade North America, SHUFE, Toronto Macro/Finance Conference, Tsinghua (PBCSF & SEM), UCLA, UIC Finance Conference, UIUC, UT Austin, University of Houston, University of Macau, University of Mannheim, University of Miami, University of Minnesota, University of Toronto, University of Zurich, USC, WashU, Western University, WFA, and Wharton for their comments. Dou is grateful for the financial supports from the Golub Faculty Scholar Award at Wharton. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

AI-Powered Trading, Algorithmic Collusion, and Price Efficiency
Winston Wei Dou, Itay Goldstein, and Yan Ji
NBER Working Paper No. 34054
July 2025
JEL No. D43, G10, G14, L13

ABSTRACT

The integration of algorithmic trading with reinforcement learning, termed AI-powered trading, is transforming financial markets. Alongside the benefits, it raises concerns for collusion. This study first develops a model to explore the possibility of collusion among informed speculators in a theoretical environment. We then conduct simulation experiments, replacing the speculators in the model with informed AI speculators who trade based on reinforcement-learning algorithms. We show that they autonomously sustain collusive supra-competitive profits without agreement, communication, or intent. Such collusion undermines competition and market efficiency. We demonstrate that two separate mechanisms are underlying this collusion and characterize when each one arises.

Winston Wei Dou
University of Pennsylvania
The Wharton School
and NBER
wdou@wharton.upenn.edu

Yan Ji
Hong Kong University of Science
and Technology (HKUST)
jiy@ust.hk

Itay Goldstein
University of Pennsylvania
The Wharton School
and NBER
itayg@wharton.upenn.edu

A data appendix is available at http://www.nber.org/data-appendix/w34054
A SSRN Link is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452704

1 Introduction

The integration of algorithmic trading with reinforcement learning (RL) algorithms, often termed AI-powered trading, has the potential to reshape financial markets and poses new regulatory challenges. While traditional algorithmic trading relies on static, hardcoded rules defined by humans, RL-based trading algorithms autonomously optimize their strategies through self-learning, trial-and-error interactions with the market and adapt in real time based on observed outcomes. Such adoption of AI algorithms in trade execution has recently gained significant momentum and its future progression seems unavoidable.[1]

One of the most pressing regulatory concerns related to the adoption of AI is the risk of AI collusion. As we discuss in the literature review below, AI collusion has been a concern in areas outside financial markets and it poses particular risks in financial markets. We define AI collusion as a scenario where autonomous, self-interested RL algorithms independently learn to coordinate their trading in a way that secures supra-competitive profits, without explicit agreements, communication, or pre-programmed intent. Such algorithmic collusion could benefit a small group of sophisticated speculators equipped with advanced technologies, while harming broader market participants by undermining competition, liquidity, and market efficiency.

What makes AI collusion particularly challenging to regulators is that it falls outside the scope of existing antitrust enforcement frameworks,[2] which focus on detecting explicit communication or evidence of shared intent (e.g., Harrington, 2018; Massarotto, 2025). This focus reflects the prevailing view that communication is important for humans to sustain collusion.[3] As a result, AI collusion, despite yielding similar anti-competitive outcomes, remains largely unaddressed under current law. This legal gap is particularly salient in financial markets, where the boundary between illegal communication used for manipulation and lawful communication necessary for enhancing market efficiency and stability is inherently difficult to define and detect. But before evaluating these issues, we need a better understanding of whether AI collusion in securities trading can arise in the first place, given the unique nature and structure of the financial market, and if so, then how it is affected by the parameters of the market.

In this article, we show that AI collusion in securities trading can robustly arise. Our analysis starts with a model to analyze the possibilities of collusion in equilibrium without considering AI agents. We then conduct simulation experiments with RL algorithms trading in an environment similar to the model and explore the patterns of collusion they achieve. We show that there are two fundamentally distinct algorithmic mechanisms through which collusion is achieved across a range of market environments: one based on price-trigger strategies, and the other driven by over-pruning bias in learning. We systematically characterize the conditions under which each mechanism prevails. Both algorithmic mechanisms underlying AI collusion have counterparts in economic theory and can

[1] For example, the Securities and Exchange Commission (SEC) recently approved Nasdaq's RL-based, AI-driven order type; major digital platforms have begun deploying RL trading bots; and leading hedge funds and trading powerhouses are increasingly adopting AI for trading.

[2] While securities trading is primarily governed by securities laws, Section 1 of the Sherman Act applies to collusive practices that suppress competition in financial markets. Overlap arises when manipulative conduct has anti-competitive effects, triggering dual enforcement by the Department of Justice (DOJ) and SEC.

[3] This is rooted in historical case studies and experimental research on human tacit collusion (e.g., Levine, Palfrey and Plott, 1991; Genesove and Mullin, 2001; Fonseca and Normann, 2012; Charness et al., 2014; Cooper and Kühn, 2014).

be interpreted through game-theoretic equilibrium concepts. We analyze the resulting AI collusive equilibrium using extensive simulations and provide heuristic justification for how these algorithmic mechanisms operate.

Theoretical Benchmarks. We start by developing a model that incorporates key ingredients of trading in financial markets. We provide theoretical analysis of this model as benchmark, and then use it as basis for our simulation experiments with RL algorithms. Our model builds on the influential framework of Kyle (1985), in which an informed speculator trades against noise traders, and a market maker sets prices to minimize pricing errors based on the information gleaned from the total order flow. We start from the static framework of Kyle (1985) and extend it in the following ways that are critical for the exploration of collusion in trading.

First, instead of a single informed speculator operating in a one-period market, we consider oligopolistic informed speculators who trade repeatedly across periods, with each period involving a different short-lived asset.[4] At the beginning of each period, each informed speculator receives a private signal about the fundamental value of that period's short-lived asset, which is realized at the end of the period. Clearly, the introduction of multiple oligopolistic speculators who interact repeatedly reflects realistic market settings, such as quantitative hedge funds and proprietary trading firms engaged in trading that happens at increasingly higher frequencies. These features are essential for studying how collusion may arise in financial markets.

Second, we introduce a continuum of atomistic, information-insensitive investors who trade the asset and collectively generate a downward-sloping demand curve within each trading period (similar to Kyle and Xiong, 2001; Vayanos and Vila, 2021). These investors, such as retail traders using technical analysis or institutional investors seeking hold-to-maturity positions for hedging short-term risks, are typically unresponsive to real-time information about the asset's fundamental value. Instead, they trade against the current price and in the direction of the asset's expected long-term value. While we do not endogenize the behavior of these information-insensitive investors, they need not be behaviorally biased. They may find it optimal not to pay attention to short-term fluctuations or may behave this way for institutional reasons. As discussed above, these traders resemble different types of investors in real-world markets. This feature, together with the next one, injects inefficiency into the pricing mechanism of Kyle (1985), which as we show, is a critical element for a key mechanism for collusion.

Third, trading occurs through the market maker who sets the market price and holds inventory to clear the market. The market maker observes the total order flow from informed speculators and noise traders, along with the deterministic order flow schedule of information-insensitive investors as a function of price. Given this information, the market maker sets the market price optimally to minimize a weighted average of inventory costs and pricing errors. Hence, unlike in Kyle (1985), inventory costs play a role in the pricing mechanism, which is a realistic feature of financial markets. Having the information-insensitive traders alongside this concern for inventory costs is what injects inefficiency to the price, which will be important for the analysis of collusion.

We analyze the theoretical model and generate some novel results about the possibility of collusion

[4]Our repeated trading setup is distinct from dynamic trading frameworks with a long-lived asset traded over multiple rounds within each period (e.g., Kyle, 1985; Holden and Subrahmanyam, 1992; Rostek and Weretka, 2015).

in financial market trading. We first consider two theoretical benchmarks to characterize the steady-state behavior of informed speculators: the non-collusive Nash equilibrium benchmark and the perfect cartel benchmark. The non-collusive Nash equilibrium refers to the one-shot Nash equilibrium of the stage game, in which no one can profitably deviate. Here, every speculator is trading to maximize current trading profit, not taking into account the effect on the profit of others. In contrast, the perfect cartel represents the outcome in which informed speculators act collectively as a monopolist to maximize joint profits. Relative to the non-collusive Nash equilibrium, informed speculators in the cartel trade less aggressively on the information about asset, as this enables them to make higher collective profits. A collusive equilibrium, if it arises, would lie between these two benchmarks. We define such an equilibrium by two features: (i) informed speculators earn supra-competitive profits that exceed those obtained in the non-collusive Nash equilibrium by trading less aggressively on signals; and (ii) each speculator has the option to deviate for unilateral one-period gains, with such deviations imposing losses on others.

A collusive trading equilibrium can be sustained as a subgame perfect Nash equilibrium in our framework as a result of price-trigger strategies. When market prices are sufficiently informative, speculators can imperfectly infer others' trades from market price movements, enabling tacit coordination. Specifically, speculators trade less aggressively on their information, knowing that a deviation to a more aggressive strategy is likely to lead to the price shooting over the trigger, which will lead the other speculators to punish them by reverting to the aggressive strategy in the non-collusive Nash equilibrium. This form of collusion was introduced by Green and Porter (1984) and Abreu, Pearce and Stacchetti (1986). Importantly, the viability of this equilibrium hinges critically on high price informativeness, which is a central concept in the context of financial markets. We show that sustaining a collusive Nash equilibrium via price-trigger strategies becomes impossible when noise trading risk is high or when information-insensitive investors are only weakly present. Intuitively, high noise trading risk leads to low price informativeness, weakening the effectiveness of prices as monitoring devices. Moreover, when the information-insensitive investors are not prominent, speculators must trade conservatively on private signals to preserve information rents, reducing price informativeness and rendering prices ineffective for detecting deviations, regardless of the level of noise trading risk. This characterization of when price-trigger collusive equilibria are possible is novel in the literature on financial-market trading.

Other possibilities of collusive equilibria arise outside the concept of a Nash equilibrium. Specifically, following the concept of experience-based equilibrium (Fershtman and Pakes, 2012), informed speculators may trade less aggressively on their private signals because of a learning bias that leads them to undervalue the payoff from aggressive trading. The bias persists because learning is based solely on realized outcomes along the equilibrium path, while off-path strategies are insufficiently revisited or updated. As a result, the learning process reinforces outcome evaluations that are internally consistent with observed on-path data but fails to correct for underexplored off-path strategies. We show that such equilibria exist for the entire parameter space, not depending on how prominent noise traders or information-insensitive traders are.

Algorithmic Mechanisms That Lead to AI Collusion. In the main part of the paper, we examine whether informed speculators, each governed by an independent and self-interested AI algorithm,

can reach the collusive outcomes described above without explicit agreements, communication, or pre-programmed intent. To do so, we run simulation experiments using autonomous, model-free Q-learning algorithms that replace the informed speculators in the theoretical framework. Unlike their theoretical counterparts, these algorithms rely on RL to determine how to trade on private signals, rather than on rationality or strategic foresight. Q-learning serves as a foundational basis for many RL algorithms that have significantly advanced the AI field. It is valued for its simplicity, transparency, and economic interpretability.

We provide additional details about the algorithms in Section 2. As described there, in each period, an algorithm selects an action (i.e., quantity traded) based on the state it faces (more on the state variables below). It stores and updates estimated values for each state-action pair, including both optimal and suboptimal actions. These values are referred to as estimated Q-values, and together they form the estimated Q-matrix over the discrete state and action space. At the start of each period, the algorithm observes the realized state and uses it to update one cell in the Q-matrix corresponding to the state-action pair from the previous period. The realized state may depend on both the prior state and action. The update is a weighted average of past experience and new information, incorporating both the reward just received and the estimated continuation value based on the newly realized state. At the end of each period, the algorithm selects an action according to a standard exploration-exploitation rule. Exploitation involves choosing the action with the highest estimated Q-value for the current state, while exploration involves selecting a random action. The interplay between exploration and exploitation is a defining feature of RL algorithms and is critical for effective learning. Typically, the likelihood of exploration gradually declines to zero, while that of exploitation increases toward one. In our simulation experiments, the Q-learning algorithms use a state vector that includes (i) the lagged market price, (ii) the lagged fundamental value, and (iii) the current fundamental value. Because market prices are endogenously determined through interactions among algorithms, noise traders, information-insensitive investors, and the pricing rule, the system is highly complex and does not yield easily predictable outcomes.

Our simulation experiments show that AI collusion arises across a wide range of market parameters and RL hyperparameters. It emerges through two distinct algorithmic mechanisms, each corresponding to one of the two theoretical collusion mechanisms discussed above and occurring in a different region of the market parameter space. One mechanism is based on price-trigger strategies, closely approximating the collusive Nash equilibrium sustained by such strategies. The other results from a learning bias that leads to the over-pruning of aggressive strategies, aligning with the collusive experience-based equilibrium. We refer to the former as AI collusion driven by "artificial intelligence," and the latter as AI collusion driven by "artificial stupidity." We elaborate below on the conditions under which each mechanism arises and explain how it emerges.

Similar to the predictions from the theoretical model described above, our simulation experiments show that an AI collusive equilibrium sustained by price-trigger strategies emerges robustly in environments with low noise trading risk and a significant presence of information-insensitive investors. In such environments, the lagged price, as an endogenous state variable, is highly informative about whether all algorithms traded conservatively in the previous period, which is a key requirement in the theory for sustaining price-trigger strategies. Given that the RL algorithms do not know whom they are playing against or how their payoffs are generated — they simply track

states, their own actions, and their own realized payoffs — it is natural to ask how they converge to an outcome that closely resembles the one predicted by the fully rational model.

The intuition is as follows. After the exploration-intensive phase, the algorithms assign higher estimated Q-values to aggressive strategies, where they trade strongly on news about the fundamental, as these strategies yield much higher payoffs when played against opponents who randomly choose to trade aggressively. Hence, as the system transitions into the exploitation-intensive phase, the algorithms consistently select aggressive trading strategies when they trade against each other, and prices move strongly with fundamentals as a result. This leads the estimated Q-values of aggressive strategies to gradually decline, as they converge towards their non-collusive Nash equilibrium levels when the aggressive strategy is commonplace among the algorithms. At the same time, occasional but ongoing exploration reveals to the algorithms that conservative trading strategies yield higher estimated Q-values than aggressive ones in states where lagged prices respond only moderately to lagged fundamentals. As a result, the algorithms gradually converge to adopting conservative strategies when others do the same, mirroring collusive behavior. A feedback loop reinforces this outcome: in these states, all algorithms select conservative strategies during exploitation, which causes similar states to recur, where lagged prices respond only moderately to fundamentals. Finally, for this pattern to amount to price-trigger collusive behavior, a form of "punishment" following large price responses to fundamentals is needed. Indeed, we observe that all algorithms shift to aggressive trading following such a price response. This occurs because the algorithms recognize the pattern that when prices respond strongly to fundamentals, trading aggressively is still the best option. Overall, the trading behavior thus exhibits mostly conservative trading with moderate price reactions but there are occasional reversions to punishment phases characterized by aggressive trading behavior. This pattern emerges even though the algorithms lack the strategic sophistication of the fully rational informed speculators in the model. This is why we refer to it as AI collusion through "artificial intelligence."

Importantly, the convergence to this pattern relies on the informativeness of prices. In environments with high noise trading risk or a limited presence of information-insensitive investors, both of which result in low price informativeness, the price-trigger mechanism breaks down. This is because the link between the fundamental value, the action, and the price becomes too noisy for the RL process to reliably distinguish patterns where conservative behavior leads to moderate price responses to fundamentals from those where aggressive behavior leads to strong responses. Interestingly, we find that across a wide range of such settings, AI collusion still emerges, but through a learning bias that systematically over-prunes aggressive strategies.

The intuition for this particular form of learning bias lies in the asymmetry of estimated Q-value updates in response to noise trading shocks, a feature inherent to RL due to its reliance on exploitation. When noise traders happen to trade in the same direction as the algorithm's trade, algorithms submitting aggressive trades incur large losses, which become more severe as noise trading risk increases. The algorithm then sharply lowers the estimated Q-value of that strategy, treating it as a very poor action. This discourages the algorithm from revisiting the strategy, thereby locking in the downward bias on its estimated value. Conversely, when noise traders happen to trade in the opposite direction of the algorithm's trade, the algorithm earns large profits and may initially overestimate the Q-value. However, because exploitation leads to frequent reuse of strategies

with high estimated Q-values, the algorithm continually revisits this action, allowing the estimated Q-value to be eventually corrected through sufficient further updates. In environments where trading outcomes are dominated by random noise rather than informed trading, this asymmetry in the exploitation process becomes especially pronounced and cannot be effectively corrected through exploration. Aggressive trading strategies, being more exposed to noise trading shocks, are more likely to be prematurely pruned. This asymmetry causes the algorithm to develop a biased value system that consistently favors conservative trading strategies. Given the nature of competition among algorithms, they end up collectively benefiting from this bias, leading to a collusive outcome. This is why we refer to the behavior as AI collusion through "artificial stupidity."

The pervasiveness of AI collusion in our simulation experiments has first-order implications for market outcomes, and so is highly relevant for the purpose of market regulation. We show that a greater extent of collusion, characterized by higher supra-competitive profits for the algorithms, leads to lower market liquidity, lower price informativeness, and higher mispricing, regardless of which algorithmic mechanism the AI collusion is based on. To better understand the drivers of AI collusion, we conduct extensive simulation experiments by varying different market parameters. In price-trigger AI collusion, collusion capacity increases when the number of informed speculators falls, noise trading risk decreases, or the subjective discount factor increases. In contrast, over-pruning AI collusion shows different patterns: fewer informed speculators have similar effects, but lower noise trading risk reduces collusion, and the subjective discount factor has little impact. These results align with the underlying algorithmic mechanisms explained above and are strongly consistent with what we would expect given their theoretical underpinnings. We also examine the role of RL hyperparameters, including the weight put on recent experience relative to past information in the updating process and the rate of exploration decay. Across a broad range of values, collusive behavior and supra-competitive profits remain robust under both algorithmic mechanisms for AI collusion.

Contributions and Related Literature. This article uncovers the economic foundations and algorithmic mechanisms of AI collusion in securities trading, focusing on its effects on price formation and market efficiency. These issues are central to current regulatory uncertainty, as AI represents a fundamentally different form of intelligence. Unlike humans, whose decisions reflect logic, emotion, and beliefs about others' beliefs, AI relies on pattern recognition and optimization. As a result, existing frameworks based on human behavior may not capture the strategic dynamics or equilibrium behavior of AI traders, highlighting the need to study the algorithmic behavior — or "psychology" — of machines (Goldstein, Spatt and Ye, 2021).

Our work follows recent work on AI collusion in retail markets (e.g., Calvano et al., 2020, 2021; Johnson, Rhodes and Wildenbeest, 2023). The financial-market setting is fundamentally different as it exhibits asymmetric information, noise trading, and a price-setting mechanism that is facilitated by market makers who consider the details of the environment. Hence, we extend the simulation-based AI experimental framework from the retail-market environment to the financial-market environment by replacing assumptions of near-perfect information and fixed demand curves with a setting characterized by substantial asymmetric information and strategically adaptive demand curves shaped by market makers' price discovery. As discussed above, our setting is characterized by two key parameters: the level of noise trading risk and the extent of information-insensitive investor

presence. We identify two distinct algorithmic mechanisms through which AI collusion can occur and systematically characterize when each of them arises as a function of these parameters and others. Our model results present novel contribution to the theoretical literature on financial-market trading, and our simulation-based experimental results have no parallel in the emerging literature on AI collusion mentioned above.

While the price-trigger collusion is shown in the simpler setting of Calvano et al. (2020, 2021), we show that it only holds when there is small noise trading risk and strong information-insensitive investor presence. Yet, when this mechanism fails, AI collusion typically arises through a distinct channel: a learning bias driven by the over-pruning of aggressive strategies. In the terminology of Calvano et al. (2020), this latter channel may not count as collusion, as it arises from a learning bias, even though it satisfies the two defining features of a collusive equilibrium described above. However, it is important to note that it is equally robust and has largely the same implications for trading profits, market liquidity, price informativeness, and mispricing. Hence, understanding how and when the two mechanisms emerge is of equally high importance. Others have studied algorithmic mechanisms, which generate supra-competitive profits without the punishment trigger (e.g., Waltman and Kaymak, 2008; Hansen, Misra and Pai, 2021; Abada and Lambin, 2023; Asker, Fershtman and Pakes, 2024; Banchio and Mantegazza, 2024; Dolgopolov, 2024; Lambin, 2024), but they largely rely on simplifying restrictions on the algorithmic capacity. We provide a detailed discussion of these works in Online Appendix 1. On the other hand, the over-pruning bias we uncover arises in a highly sophisticated environment, complementing the range of parameters where price-trigger collusion arises, and points to a pervasive feature of the RL framework: the asymmetric effect of exploitation, whereby adverse and beneficial shocks influence learning differently.

There are a few early works that investigate the effects that related algorithms may have on financial or money markets (e.g., Marimon, McGrattan and Sargent, 1990; Routledge, 1999, 2001). However, they either explore adaptive learning algorithms or more basic RL algorithms than ours. They do not develop implications such as we develop here regarding collusion and its effects on market efficiency. A related contemporaneous work, Colliard, Foucault and Lovo (2025), studies interactions among Q-learning algorithms but focuses on stateless AI market makers. In contrast, we study AI-powered informed speculators using Q-learning with endogenous state variables, such as past prices. Unlike them, we uncover the different algorithmic mechanisms that drive AI collusion and characterize when they dominate. Cartea et al. (2022*b*) also analyze stateless RL in market making using a multi-armed bandit algorithm.

Furthermore, our paper contributes to the rapid growing literature on the impact of AI and big data on the efficiency and functioning of financial markets (e.g., Goldstein, Spatt and Ye, 2021; Farboodi and Veldkamp, 2023, for literature review). Recent studies theoretically examine how data abundance and advances in information processing technologies affect price informativeness and market liquidity (e.g., Dugast and Foucault, 2018; Farboodi and Veldkamp, 2020; Dugast and Foucault, 2024). Another strand of the literature demonstrates that advanced machine learning techniques can effectively extract predictive signals or latent economic structures from high-dimensional public data, which are otherwise difficult to detect using traditional methods (e.g., Kaniel et al., 2023; Cao et al., 2024; Chen, Kelly and Xiu, 2024; Gao, Xiong and Yuan, 2024; Kelly, Malamud and Zhou, 2024; Chen et al., 2025). In contrast, our paper focuses on understanding the behavior of AI agents that replace

humans. We examine the resulting AI equilibrium, shaped by algorithmic interactions, highlighting the importance of examining this equilibrium when assessing the overall impact of AI adoption on market efficiency.

Finally, our paper is closely related to the literature on imperfect competition in financial markets. Rostek and Yoon (2024) provide a recent review of the theory of imperfectly competitive financial markets, covering influential early contributions such as Kyle (1989) and Vayanos (1999), which focus on non-collusive equilibria. Theoretical works that study the effect of coordination or collusion among major market participants on market microstructure dynamics under a repeated game framework include Dutta and Madhavan (1997), Carlin, Lobo and Viswanathan (2007), and Hörner, Lovo and Tomala (2018), among others. The common feature of these models is that supra-competitive trading profits are sustained through the threat of punishment. In addition to these models that focus on market microstructure dynamics, there are other papers that theoretically analyze collusion in financial markets due to FinTech (e.g., Cong and He, 2019). Several papers provide supporting evidence for collusion in financial markets across various settings (e.g., Christie and Schultz, 1994, 1995; Christie, Harris and Schultz, 1994; Chen and Ritter, 2000; Dou, Wang and Wang, 2023; Bryzgalova, Pavlova and Sikorskaya, 2025; Lehar and Parlour, 2025). We provide new theoretical results characterizing when collusion can be sustained in financial-market environments and contribute further by showing that autonomous, self-interested AI-powered trading algorithms can learn to coordinate, even without any agreement, communication, or intention.

2 AI-Powered Trading Algorithms

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.[5] Second, it is widely adopted in practice. Third, it is valued for its simplicity, transparency, and economic interpretability.

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 vector s, which evolves according to a Markov process. 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 agents,

[5] Q-learning and other dynamically sophisticated RL algorithms take into account the possibility that actions lead to state transitions and internalize that an action taken in a given state can affect future states and rewards. In contrast, multi-armed bandit algorithms, a class of stateless RL methods, are not dynamically sophisticated: they do not incorporate any notion of state and therefore ignore the possibility that actions influence future decision-making environments.

and $s, s' \in S$ represent the states in the current and the next period, respectively, with S denoting the set of states. The state vector s may depend on agent-specific conditions and private signals faced by each agent i, for all i. 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 factor.

Equation (2.1) represents the recursive formulation of dynamic control problems (e.g., Bellman, 1954; Ljungqvist and Sargent, 2012). It characterizes behavior along the equilibrium path, where the optimal value function $V_i(s)$ depends only on the current state s. In contrast, the Q function, denoted by $Q_i(s, x_i)$, extends the value function to each possible state-action pair, allowing evaluation of outcomes not only on the equilibrium path but also for counterfactual or off-path actions. By definition, the Q-function value for a given (s, x_i) corresponds to the right-hand side of 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, Q-learning estimates the Q-matrix when the conditional distribution $\mathbb{E}[\cdot|s, x_i]$ and off-equilibrium observations (s, x_i) are limited. By design, the Q-learning algorithm addresses both challenges simultaneously: it uses the law of large numbers to learn the underlying distribution from repeated experiences, while its trial-and-error experiments generates counterfactual outcomes for state-action pairs that may not occur along the equilibrium path.

Agent i's iterative experimentation begins with an arbitrary initial estimated Q-matrix, $\widehat{Q}_{i,0}$, and recursively updates it from $\widehat{Q}_{i,t}$ to $\widehat{Q}_{i,t+1}$ in iteration $t+1$ as follows:

$$\widehat{Q}_{i,t+1}(s_t, x_{i,t}) = (1-\alpha)\underbrace{\widehat{Q}_{i,t}(s_t, x_{i,t})}_{\text{Past knowledge}} + \alpha\underbrace{\left[\pi_{i,t} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t}(s_{t+1}, x')\right]}_{\text{New information from experimentation}}, \tag{2.4}$$

where $\alpha \in [0, 1]$ captures the forgetting rate.[6] Upon agent i choosing action $x_{i,t}$ in state s_t and observing the payoff $\pi_{i,t}$, the update from $\widehat{Q}_{i,t}$ to $\widehat{Q}_{i,t+1}$ at the pair $(s_t, x_{i,t})$ occurs immediately after

[6]The forgetting rate α determines how quickly past experiments are discounted. For consistent learning, α must decay to zero to ensure convergence of the estimated Q-matrix $\widehat{Q}_{i,t}$ as t grows large. A smaller α improves asymptotic accuracy but slows convergence, reflecting a higher learning capacity at the expense of greater computational cost.

the next state s_{t+1} is drawn from the Markov transition distribution at the beginning of iteration $t+1$, conditional on the state s_t, the chosen action $x_{i,t}$, and the collective actions of all agents in iteration t.

Equation (2.4) indicates that for agent i in iteration $t+1$, 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} at the beginning of iteration $t+1$.

It is crucial to note that the forgetting rate α plays a significant role in the Q-learning algorithm, balancing past knowledge with new information from experimentation. 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. This weakens the applicability of the law of large numbers, leading to substantial bias in estimating $\mathbb{E}[\cdot|s, x_i]$ for future payoffs and continuation Q-values. Conversely, a smaller α slows the decay, preserving more past information and reducing bias. However, this requires significantly more iterations to achieve convergence, increasing computational costs.

2.3 Experimentation

Upon state s_t being realized at the beginning of iteration t, agent i chooses an action $x_{i,t}$, at the end of the iteration, according to either an exploitation or exploration mode, as follows:

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

To determine the mode, we employ the simple ε-greedy method. As outlined in equation (2.6), after the state s_t is realized in iteration t, agent i follows either the exploration or exploitation mode with exogenous probabilities ε_t and $1-\varepsilon_t$, respectively. In the exploitation mode, agent i chooses its action to maximize the estimated Q-value based on s_t in iteration t, given by $x_{i,t} = \operatorname{argmax}_{x \in \mathcal{X}} \widehat{Q}_{i,t}(s_t, x)$. In contrast, 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.[7] Exploration enables the algorithm to experiment with actions that appear suboptimal under the estimated Q-values, $\widehat{Q}_{i,t}$, in iteration t. Sufficient exploration is crucial for accurately approximating the true Q-matrix and mitigating learning bias, as it ensures that all state-action pairs are sufficiently sampled, especially in complex environments. However, this comes with a tradeoff: while it enhances learning accuracy, it also increases computational burden and introduces noise, which can hinder convergence in multi-agent settings. To manage this tradeoff, the exploration probability ε_t is set to decrease monotonically toward zero as t increases.

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, model-based synchronous learning updates all (s, x) pairs simultaneously in each iteration, assuming precise knowledge of the environment's structure, such as transition probabilities and payoff functions (e.g., Asker, Fershtman and Pakes, 2022, 2024). Model-based approaches are typically vulnerable to misspecification.

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' price discovery. 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 deviations from the Kyle (1985) baseline framework. First, we consider a setting with oligopolistic informed speculators in a repeated trading environment, 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.[8] 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.

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

[7] For simplicity, we use a uniform distribution, though smarter choices could improve Q-learning.

[8] 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; Rostek and Weretka, 2015).

set $\overline{v} \equiv \sigma_v \equiv 1$ for simplicity.[9] 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 best 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 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. Although y_t becomes observable after all trades from informed speculators and noise traders are submitted in period t, its components cannot be separately identified, making it impossible to distinguish the informed order flow from the noise trading flow. 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.[10] 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.[11]

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

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

[10] 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 2.1. 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).

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

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

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

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).[12] 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).

Interpreting the Model through a Specific Market Setting. Our model reflects realistic market environments, particularly those involving quantitative hedge funds and proprietary trading firms operating at increasingly short horizons. While the theoretical framework applies broadly to real-world settings, we anchor our simulation experiments in a concrete example to illustrate the economic relevance of AI-driven trading algorithms. In each period t, a new short-lived security is introduced and traded, such as a close-to-maturity option or futures contract. These contracts expire at the end of the period, with their payoff equal to the fundamental value v_t. Close-to-maturity derivatives are among the most actively traded across the maturity spectrum, making them a natural focal point for studying algorithmic trading behavior.

Below, we elaborate on each of the four types of market participants in this concrete real-world example. First, informed speculators, such as quantitative hedge funds and proprietary trading firms, 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.[13] These informed speculators

[12] Further discussions are provided in Online Appendix 2.1.

[13] Conceptually, "private signals" here include not only information derived from proprietary sources but also predictive

typically operate with two teams: (i) a 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 article are those employed by the execution team to convert private signals into strategic trading orders. These algorithms operate after the research team has generated the signals and focus on optimizing execution based on their informational content.

Second, 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.

Third, noise traders, by contrast, make trading decisions unrelated to fundamental information or technical analysis. Instead, their trades are driven by factors such as liquidity needs, portfolio rebalancing, market sentiment, or rumors.

Fourth, 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. Their primary function in our model is to support price discovery by integrating information from other market participants' trading behaviors 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,

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.

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. That is, each informed speculator maximizes its current-period payoff given others' actions, without considering how current actions may affect future payoffs or behavior. In this equilibrium, no one can profitably deviate. Details are in Online Appendix 2.1.

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 2.1.

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 via price-trigger strategies and the collusive experience-based equilibrium via 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. With sufficiently high price informativeness, informed speculators can imperfectly infer order flows of others from market prices, enabling tacit collusion.[14] Price-trigger collusion was introduced by Green and Porter (1984) and Abreu, Pearce and Stacchetti (1986).[15] We formalize this theoretical concept below.

[14]Under certain conditions, prices can even reveal others' private values in equilibrium (Rostek and Weretka, 2012), and act as a sufficient statistic for inferring others' behavior following unilateral deviations (Rostek and Yoon, 2021).

[15]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

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.*

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 $\Omega \equiv \{(\eta, \omega, T) \ : \ \eta \in [0,1],\ \omega \in [0, \bar{\omega}],\ T \in \mathbf{N}\}$.

We now explain why it is sufficient, without loss of generality, to restrict attention to the strategy space Ω with a sufficiently large but finite upper bound $\bar{\omega}$ in our analysis of the collusive Nash equilibrium. First, Sannikov and Skrzypacz (2007, Lemma 3) show that a tail test with a bang-bang property is the optimal mechanism for maximizing expected continuation payoffs while maintaining incentives against single-period deviations. Building on this insight, we focus on price-trigger strategies that serve as tail tests with bang-bang properties in our trading setting and support the collusive Nash equilibrium. Second, for a price-trigger strategy to be effective in deterring deviations (that is, to function as a powerful tail test), the associated test must have non-negligible test size (type I error). This requirement, grounded in the Neyman-Pearson lemma, implies that if the test size (type I error) is too small, the strategy becomes ineffective at detecting deviations and thus fails to discipline behavior. In particular, as long as the upper bound $\bar{\omega}$ is set sufficiently high, the test size becomes nearly zero for all $\omega > \bar{\omega}$,[16] making the strategy incapable of sustaining tacit collusion at such high ω values. Therefore, no meaningful strategy is omitted by focusing on Ω with a sufficiently large but finite $\bar{\omega}$. Additional technical details are provided in Online Appendix 2.1.

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

studies (e.g., Opp, Parlour and Walden, 2014; Dou, Ji and Wu, 2021*a*,*b*; Chen et al., 2023, 2024).

[16] e.g., $1 - \Phi(\bar{\omega}) < 10^{-15}$ when $\bar{\omega} = 8$.

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).[17] Specifically, an experience-based equilibrium is characterized by: (i) a recurrent Markovian state 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 under the true distribution, generated by optimal strategies on the equilibrium path, align with on-path evaluations. Crucially, this condition applies only to on-path outcomes. Players' beliefs or evaluations about off-path outcomes need not align with expected discounted cash flows under the true distribution, allowing for significant biases. In sum, as long as on-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 in Abreu, Milgrom and Pearce, 1991; 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 discovery, with

[17] See also Fudenberg and Kreps (1988), Fudenberg and Kreps (1995), Cho, Williams and Sargent (2002), and Cho and Sargent (2008) for influential early contributions.

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 2.3.

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. In particular, such a collusive experience-based equilibrium can 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 2.4.

3.5 The Impact of Collusive Informed Trading on Market Efficiency

To assess, based on the simulation experimental outcomes, whether informed AI speculators engage in tacitly collusive trading through price-trigger strategies or learning bias, we derive testable theoretical properties of the collusive equilibrium corresponding to 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:*

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

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

The detailed proof is provided in Online Appendix 2.5.

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

Next, 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:

$$\begin{aligned} &\rho \downarrow, \quad \sigma_u \uparrow, \quad \textit{or} \quad I \uparrow \quad \Longrightarrow \quad \Delta^C \downarrow \quad (\textit{i.e., collusion capacity} \downarrow) \\ &\quad \Longrightarrow \quad \mathcal{I}^C/\mathcal{I}^M \uparrow, \quad \mathcal{L}^C/\mathcal{L}^M \uparrow, \quad \textit{and} \quad \mathcal{E}^C/\mathcal{E}^M \downarrow \quad (\textit{i.e., market efficiency} \uparrow), \end{aligned} \tag{3.8}$$

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:

$$\begin{aligned} &\sigma_u \downarrow, \quad \textit{or} \quad I \uparrow \quad \Longrightarrow \quad \Delta^C \downarrow \quad (\textit{i.e., collusion capacity} \downarrow) \\ &\quad \Longrightarrow \quad \mathcal{I}^C/\mathcal{I}^M \uparrow, \quad \mathcal{L}^C/\mathcal{L}^M \uparrow, \quad \textit{and} \quad \mathcal{E}^C/\mathcal{E}^M \downarrow \quad (\textit{i.e., market efficiency} \uparrow), \end{aligned} \tag{3.9}$$

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 2.6.

4 Simulation Experiments on AI Trading Algorithms

As a proof of concept, this section presents simulation experiments to test whether informed AI speculators, equipped with autonomous model-free Q-learning algorithms, can achieve and sustain collusive behavior under asymmetric information and an adaptive asset demand curve that endogenously responds to their trading strategies. We specifically examine whether such collusive behavior by AI speculators can arise without explicit agreement, communication, or pre-programmed intent.

4.1 Algorithms as Experimental Subjects

Informed AI Speculators. We now analyze the behavior of the algorithms as experimental subjects. Specifically, these experiments replace the theoretical agents, referred to as "informed speculators" in the model, as detailed in Section 3, with Q-learning algorithms, as described in Section 2.

The dimensionality of the state 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.[18] 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.[19] In this setup, informed AI speculators make trading decisions in period t based on the current private signal v_t and a one-period memory of the previous fundamental value v_{t-1} and price p_{t-1}. 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 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 4.11).

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

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

[19] 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}.

$\widehat{Q}_{i,0}$ 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 $\widehat{p}_t(y)$ 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 among (i) AI speculators using Q-learning with lagged prices as endogenous state variables, (ii) an adaptive market maker learning from historical data, and (iii) randomness from noise traders and stochastic asset values makes it extremely difficult, if not impossible, to prove general results on convergence or long-run behavior. As in prior work (e.g., Calvano et al., 2020), our simulation-based approach is well-suited to study algorithmic behavior, strategic interaction, and resulting equilibrium outcomes. First, no general convergence results exist for environments of this complexity, let alone closed-form characterizations of their asymptotic behavior.

Second, although stochastic approximation theorems can, in principle, establish convergence in certain simplified settings, they are generally not applicable to settings of this complexity. Moreover, they rely on strict regularity conditions for the algorithms, such as decaying hyperparameters over time, which are rarely satisfied in practice. For example, hyperparameters are often held constant in real-world applications. As a result, the steady-state behavior observed through numerical convergence may be more practically relevant than the theoretical limit derived under idealized conditions.[20] In real-world applications, particularly in robotics and securities trading, RL algorithms operating in multi-agent environments face several practical challenges. These include the absence of theoretical guarantees on convergence and decriptions of equilibrium properties, the need for costly exploration, the inherently slow pace of learning, and the high cost and limited availability of real-world data. These factors make real-time training impractical. Consequently, training RL algorithms in simulation-based synthetic environments has become a widely adopted approach in practice. This aligns closely with the spirit of our simulation-based experiments. For example, hedge funds often use simulated financial markets to train RL-based execution strategies before deploying them in live trading, just as autonomous vehicles are first trained in virtual environments using simulated data before operating in the real world.

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 regarded as intractable, the

[20]Simulation-based algorithmic experiments fundamentally differ from numerical solutions of theoretical equilibria (e.g., Kubler and Schmedders, 2005; Dou et al., 2023; Duarte, Duarte and Silva, 2024; Hansen, Khorrami and Tourre, 2024).

mathematical proofs would provide little insight into why or how algorithms reach a collusive equilibrium. This is because such analyses typically rely on stochastic approximation methods, which focus on verifying high-level regularity conditions and technical details.[21]

To complement our simulation-based experiments across various trading environment specifications in the general model, we provide clear intuitions and heuristic justifications for the numerical convergence of multiple informed AI speculators using Q-learning algorithms, as well as for the steady-state properties of the resulting AI trading equilibrium, within a simplified model. The results are presented in Sections 5 and 6, with heuristic justifications provided in Online Appendix 3.

4.2 Numerical Specifications

We detail the numerical setup of our simulations, including the discretization of state and action spaces, Q-matrix initialization, parameter selection, and convergence criteria.

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 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.[22] 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 $N(0, \sigma_u^2)$.[23] Specifically, for each informed AI speculator

[21] Recent studies have established convergence of Q-learning algorithms to (collusive) Nash equilibria in simplified models, typically in 2×2 Prisoner's Dilemma settings (e.g., Cartea et al., 2022*a*; Possnig, 2024). These proofs rely heavily on existing stochastic approximation results and focus on technical verification with little intuitive explanation of the algorithmic mechanisms behind convergence.

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

[23] 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.

$i = 1, \cdots, I$, 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, their values are unknown to both the informed AI speculators and 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, we examine the implications of varying these parameters.

Second, "preference parameters" include the subjective discount factor for informed AI speculators, ρ, and the market maker's weight on the pricing error term, θ. We set ρ at a relatively high level, $\rho = 0.95$, to reflect the high-frequency trading environment. We examine the implications of varying ρ values in Section 6. We fix $\theta \equiv 0.1$ 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 $\sigma_v = 1$.[24] We set $\iota = 0.1$, $n_x = 15$, and $n_p = 31$.[25] 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 Online Appendix 4.12, 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 4.13, we consider two-tier Meta Q-learning algorithms that enable informed AI speculators to learn the optimal α for the lower-tier agent as part of the upper-tier agent's optimal decision.

[24] 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.

[25] 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 flow will result in a change in price p_t over the grid defined by $\mathbb{P}$.

Criterion for Numerical Convergence. Each experiment contains $N_{sim} = 1,000$ independent simulation sessions. We adopt a stringent criterion for convergence: all informed AI speculators' optimal strategies must remain unchanged for $1,000,000$ consecutive periods within a single session, and all N_{sim} sessions must continue running until each meets this convergence condition. The number of periods required for convergence varies substantially across experiments, depending on parameter values, and can also differ significantly across sessions within the same experiment. Across simulation experiments, convergence occurs within a range of approximately 20 million to 50 billion periods.[26]

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. Building on the theoretical benchmarks in Sections 3.3 and 3.4, Section 5.1 illustrates the exploration-exploitation tradeoff in RL algorithms that underpins the two algorithmic mechanisms driving AI equilibria, and Section 5.2 provides an overview of simulation results across various cases defined by different levels of σ_u and ξ. Section 5.3 presents simulation experiments in trading environments with a strong presence of information-insensitive investors (ξ is large relative to θ). In contrast, Section 5.4 focuses on simulation experiments in environments with a minimal presence of 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 Distinct Algorithmic Mechanisms behind AI Collusion

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.

Which algorithmic mechanism prevails, and consequently which type of AI equilibrium emerges, depends on the effectiveness of the exploration-exploitation tradeoff in the RL algorithm. Similar to the bias-variance tradeoff in supervised learning and high-dimensional statistics, this tradeoff aims to balance pruning the action space and reducing outcome variability. In RL, 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 received rewards, thereby lowering variability in the estimation of the optimal action. Similar to shrinkage techniques in

[26] 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.

supervised learning and high-dimensional statistics, exploitation narrows the choice space to improve convergence speed and reduce variance, although it may introduce some bias.

Drawing on the theoretical results establishing the existence of collusive equilibria sustained by two distinct economic mechanisms, as summarized in Propositions 3.1 and 3.2, the type of AI equilibrium reached by the system of RL algorithms after convergence depends on two key factors: the risk of noise trading flows, captured by σ_u, and the presence of information-insensitive investors, measured by ξ. Together, these parameters determine the informativeness of market prices, which is shaped by the underlying economic structure and, in turn, affects the effectiveness of the exploration-exploitation tradeoff.

AI collusion through the price-trigger-strategy mechanism becomes the dominant steady state when the exploration-exploitation tradeoff functions effectively, guiding learning without introducing significant bias. In this setting, a system of algorithms autonomously learns to sustain a collusive AI equilibrium that approximates a Nash equilibrium, even though each algorithm unilaterally maximizes its own trading profit. Crucially, each algorithm not only learns how the state vector (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, the information obtained from occasional exploration is more reliable. This allows exploitation to better focus on optimal trading strategies, while any bias introduced by exploitation can be effectively corrected through exploration. Further intuition is provided in Section 5.5.

AI collusion through the over-pruning learning bias mechanism emerges as the dominant steady state when the exploration-exploitation tradeoff fails to effectively guide the estimation of optimal trading strategies, resulting in significant bias. 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 exploration can be misleading, causing exploitation to become trapped in unilaterally suboptimal strategies that are collectively supra-competitive. In such cases, the significant bias introduced by exploitation cannot be effectively corrected through exploration. Further intuition is provided in Section 5.5.

To illustrate how over-pruning bias arises from an imbalance between exploration and exploitation, consider environments with low ξ or high σ_u, where market prices and trading profits are predominantly driven by noise trading shocks u_t. In these settings, the behavior of RL algorithms depends critically on how they process feedback from such shocks. Exploitation introduces asymmetries into the learning process, depending on whether a shock is adverse or beneficial. An adverse noise trading

shock moves in the same direction as the informed AI speculator's trade, causing substantial trading losses and sharply reducing the estimated Q-value of the chosen action. In contrast, a beneficial shock moves in the opposite direction, generating significant trading profits and potentially inflating the estimated Q-value, though this overestimation is more likely to be corrected over time through continued and repeated exploitation.

More precisely, following an adverse noise trading shock, the algorithm often classifies the chosen strategy as a "disastrous action," assigning it a significantly low Q-value. Exploitation then discourages the algorithm from revisiting this strategy in subsequent iterations, reinforcing the downward bias and preventing correction for such off-equilibrium-path actions. In contrast, following a beneficial shock, the algorithm tends to label the strategy as a "fantastic action" and assigns it an inflated Q-value. Because exploitation promotes repeated use of high-Q-value strategies, the algorithm continues to select and update this strategy, eventually correcting any initial overestimation. Aggressive strategies, by their nature, are more exposed to noise trading shocks, making them especially vulnerable to this asymmetric learning dynamic. As a result, they tend to be persistently undervalued and prematurely pruned from the set of candidate optimal strategies, reinforcing the over-pruning bias. Consequently, informed AI speculators gravitate toward more conservative trading strategies, consistent with the collusive behavior described in Definitions 3.1 and 3.3.

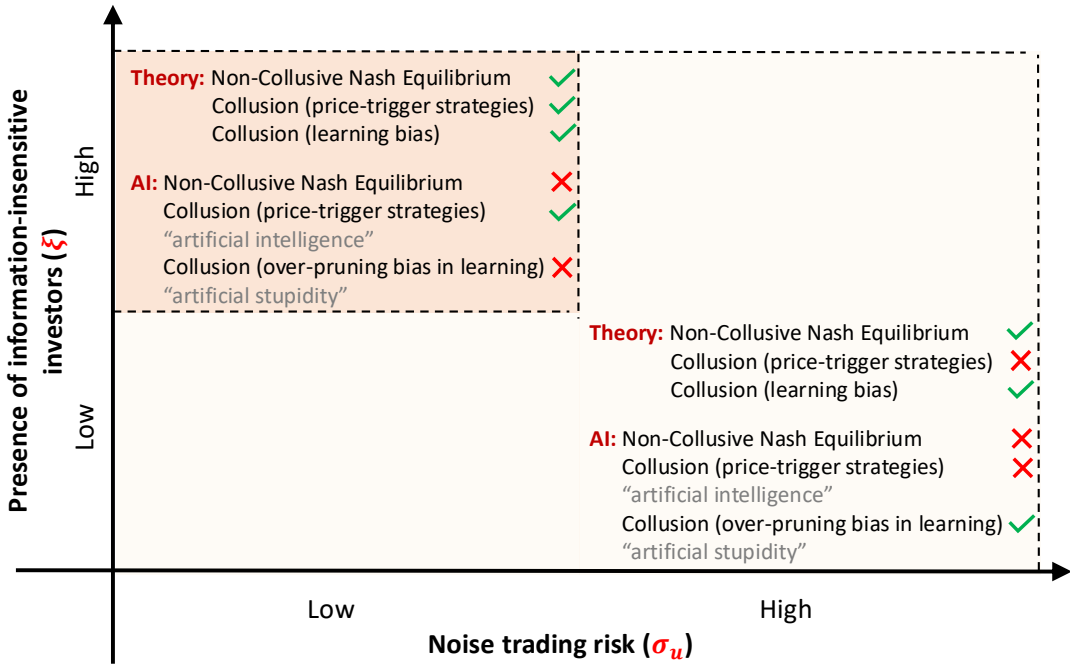
One way to interpret the asymmetric effect of exploitation is that it effectively makes RL algorithms risk-averse to randomness in their rewards. In decision theory, risk aversion arises from the asymmetric impact of adverse and beneficial shocks. Similarly, in RL, exploitation discourages revisiting 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 exposed to noise trading shocks — to be prematurely pruned from the set of potential optimal strategies, reinforcing over-pruning bias in learning. As a result, the algorithm behaves as if it were risk-averse, opting against aggressive strategies that expose profits to high risk.

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 intuitions behind the AI collusive equilibrium in Section 5.5 and heuristic justifications in Online Appendix 3. 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 via price-trigger strategies and a collusive experience-based equilibrium via learning bias can theoretically be achieved by informed speculators in such environments, as established in Propositions 3.1 and 3.2. However, in our simulations, 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.

(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



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.

3.1 and 3.2. Consistent with these theoretical benchmarks, simulations show that multiple informed AI speculators using Q-learning 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, simulations demonstrate that multiple informed AI speculators using Q-learning 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.

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

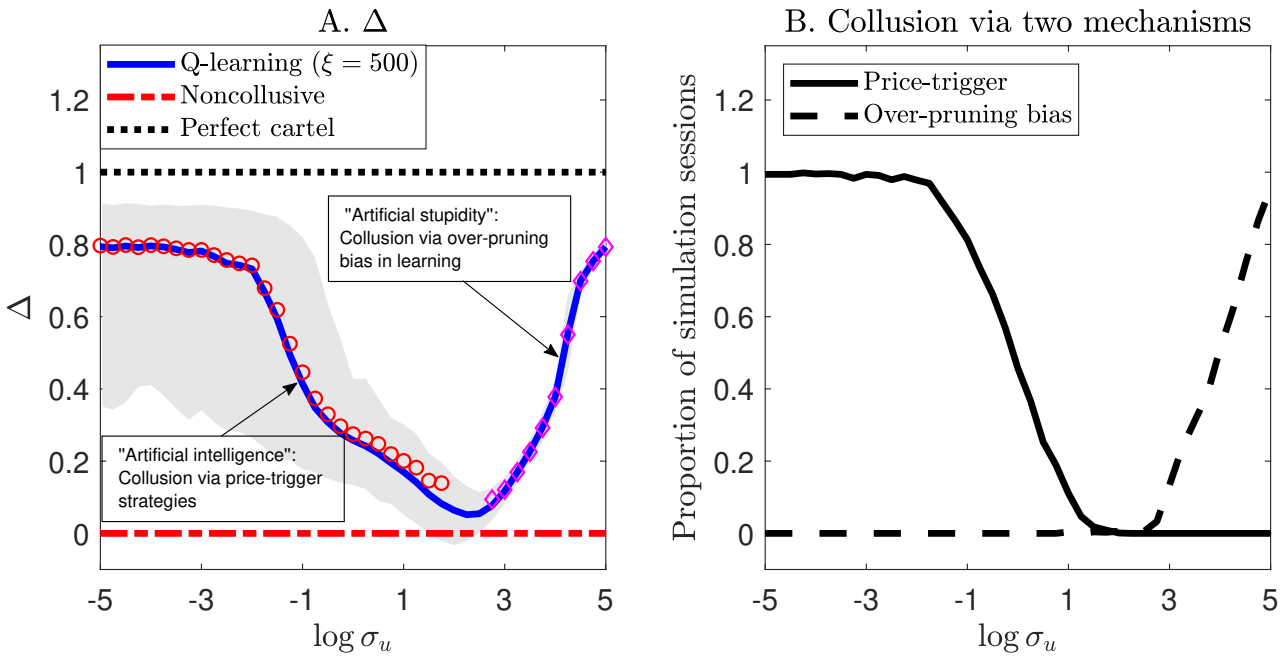


Figure 2: Two distinct mechanisms behind AI collusion.

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.[27]

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 when σ_u is either high or low. However, the algorithmic mechanisms underlying these AI collusion patterns 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.[28] The results show that when σ_u is low, nearly all simulation sessions converge to an AI collusive equilibrium sustained by price-trigger strategies,

[27] 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 4.6.

[28] Additional details on the classification are provided in Online Appendix 4.5.

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 AI collusion decreases, while the proportion converging to over-pruning learning bias AI collusion 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 sustained by 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 intuition and heuristic justification detailed in Section 5.5. 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, with further intuition and heuristic justification detailed in Section 5.5.

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 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 pre-programmed 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.[29] We then show that in high noise trading risk scenarios, informed AI speculators still sustain collusive, supra-competitive trading profits, but through a different mechanism: over-pruning bias in learning.[30] 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.

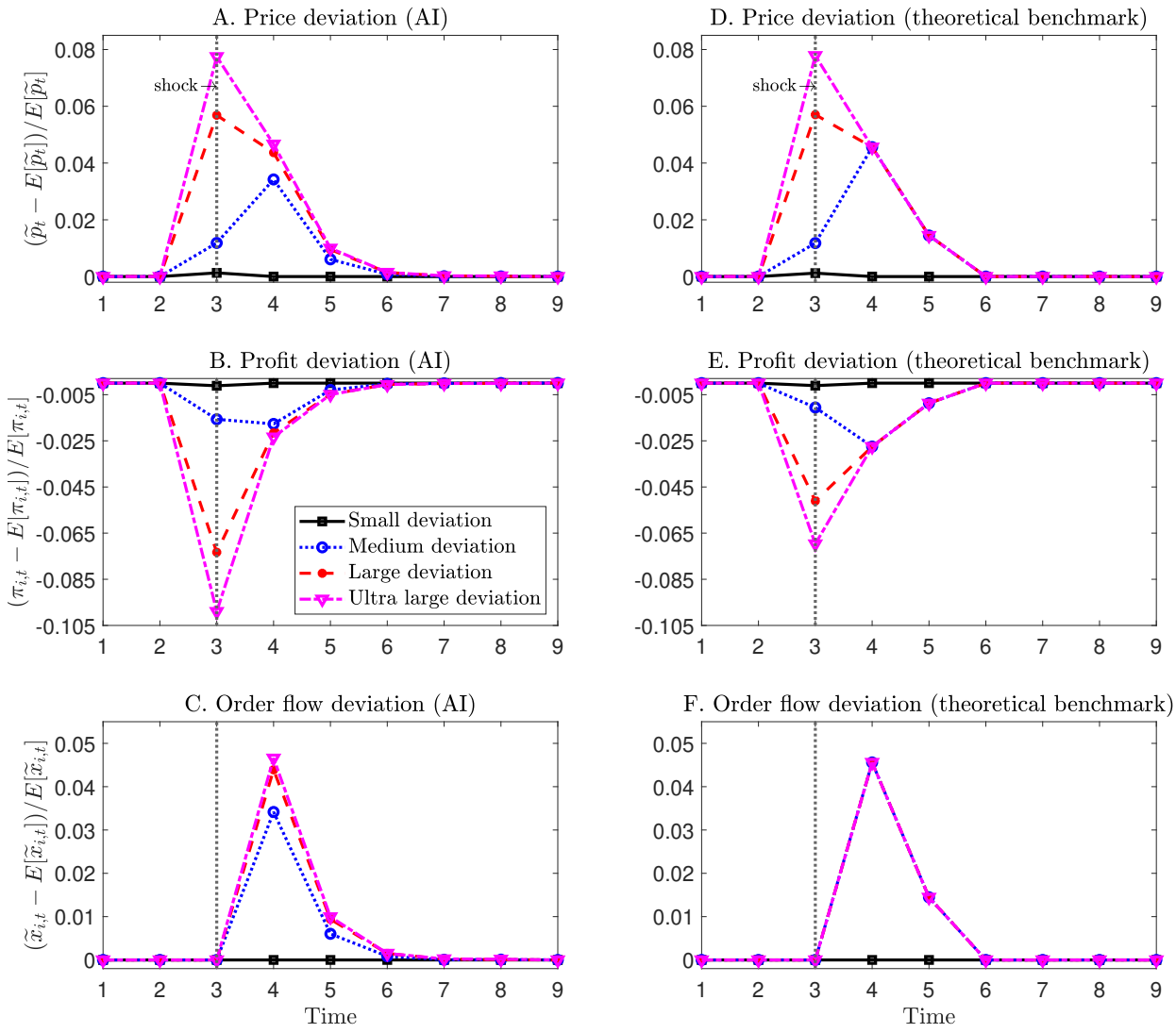
Impulse Responses: AI Collusion via Price-Trigger Strategies When σ_u Is Low. We first examine how the trained informed AI speculators respond to an exogenous shock in the noise order flow, which influences the asset's market price through the market maker's endogenous and adaptive pricing rule. At $t = 0$, all N_{sim} simulation sessions have converged. 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.[31] The cross-sectional distribution of path-by-path impulse response dynamics across N_{sim} simulation sessions is provided in Online Appendix 4.4.

Panel A of Figure 2 shows that, in environments with a significant presence of information-insensitive investors (here, $\xi = 500$) and low noise trading risk (specifically, $\sigma_u = 10^{-1}$), 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.

[29] Our numerical tests suggest that this AI collusive equilibrium is approximately Nash, meaning no local deviation is preferred. Numerical tests are detailed in Online Appendix 4.10.

[30] 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 4.2. This is unsurprising, as the experience-based equilibrium framework is broader and encompasses subgame perfect Nash equilibrium as a special case.

[31] 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.



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 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 $\widetilde{p}_t > 0$. Panels B and E depict the percentage deviation of average profits from their long-run mean for each informed AI speculator, expressed as $(\pi_{i,t} - \mathbb{E}[\pi_{i,t}])/\mathbb{E}[\pi_{i,t}]$. 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 $\widetilde{x}_{i,t} > 0$.

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

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

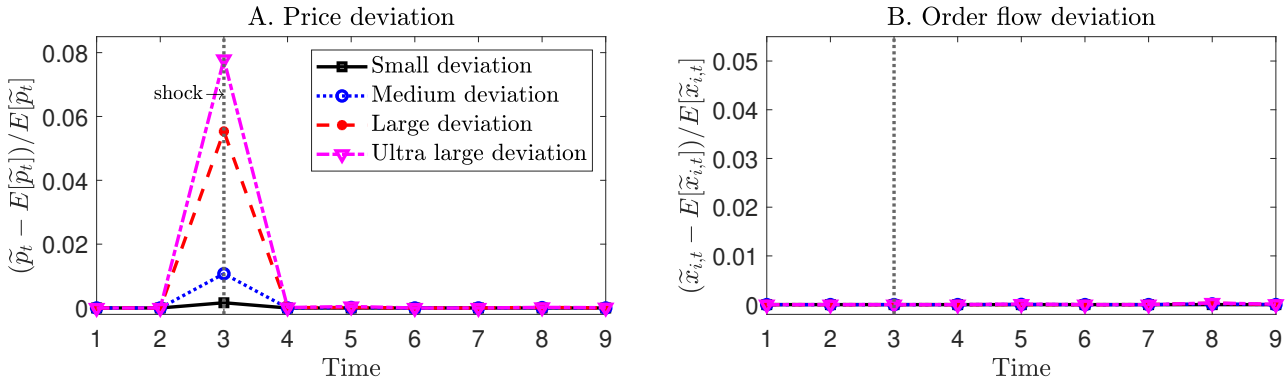
column of Figure 3, alongside the corresponding theoretical benchmarks in the right column. For a meaningful comparison, Panels D through F use the same magnitudes of unexpected price deviations at $t = 3$ as those in the simulation experiments shown in Panels A to C. Additionally, all shared parameters 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 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 respond by adopting similarly aggressive trading strategies starting at $t = 4$, 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, dashed, and 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 collusive 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 learn to 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



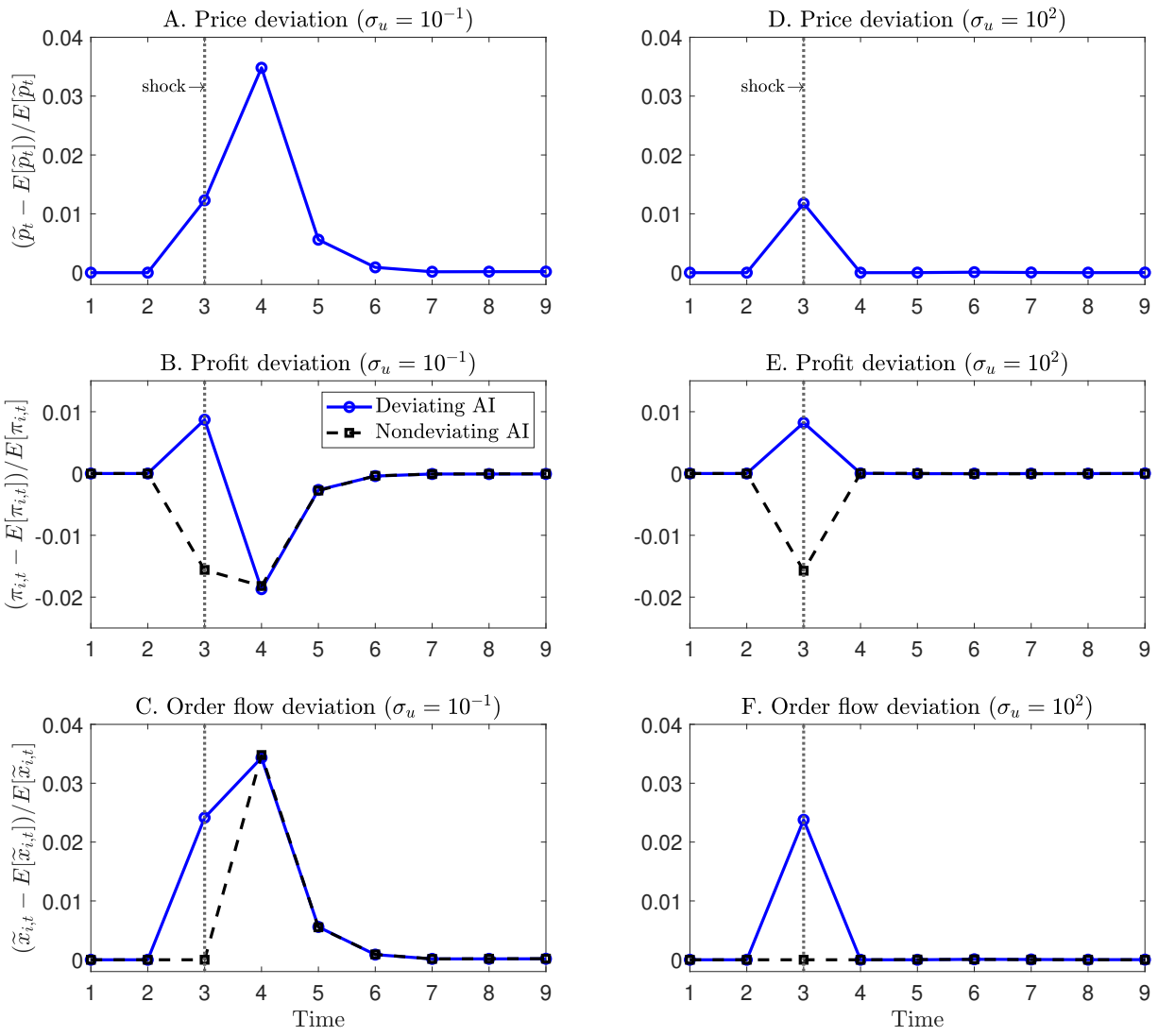
Note: Both plots are based on Q-learning simulation experiments in a trading environment with $\xi = 500$, indicating a significant presence of information-insensitive investors, and $\sigma_u = 10^2$, indicating high noise trading risk.

Figure 4: Impulse response function (IRF) of informed AI speculators using Q-learning algorithms following an exogenous noise trading shock u_{shock} for $\sigma_u = 10^2$.

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 for a straightforward comparison. In Figure 4, we investigate the average IRF over the N_{sim} simulation paths in the environment with high noise trading risk (i.e., $\sigma_u = 10^2$). A comparison between Panel B of Figure 4 and Panel C of Figure 3 shows 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 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 reach an AI collusive equilibrium through over-pruning bias in learning. This behavior corresponds to the theoretical collusive experience-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 compare the IRFs following a unilateral trading deviation by one informed AI speculator in two environments: one with low noise trading risk ($\sigma_u = 10^{-1}$) and the other with high noise trading risk ($\sigma_u = 10^2$), as shown 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



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.

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.

order flow by $x_{i,\text{shock}}$ if $v_t < \overline{v}$.

Serving as a benchmark for comparison, Panels A through C of Figure 5 show the IRF following a unilateral deviation by AI speculator i (solid line) at $t = 3$, under the low noise trading risk scenario ($\sigma_u = 10^{-1}$). 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, which formally defines a collusive equilibrium, the IRF results support the findings in Figure 2. Together, they show that informed AI speculators can interact and learn to sustain such an 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 Figure 3, 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.

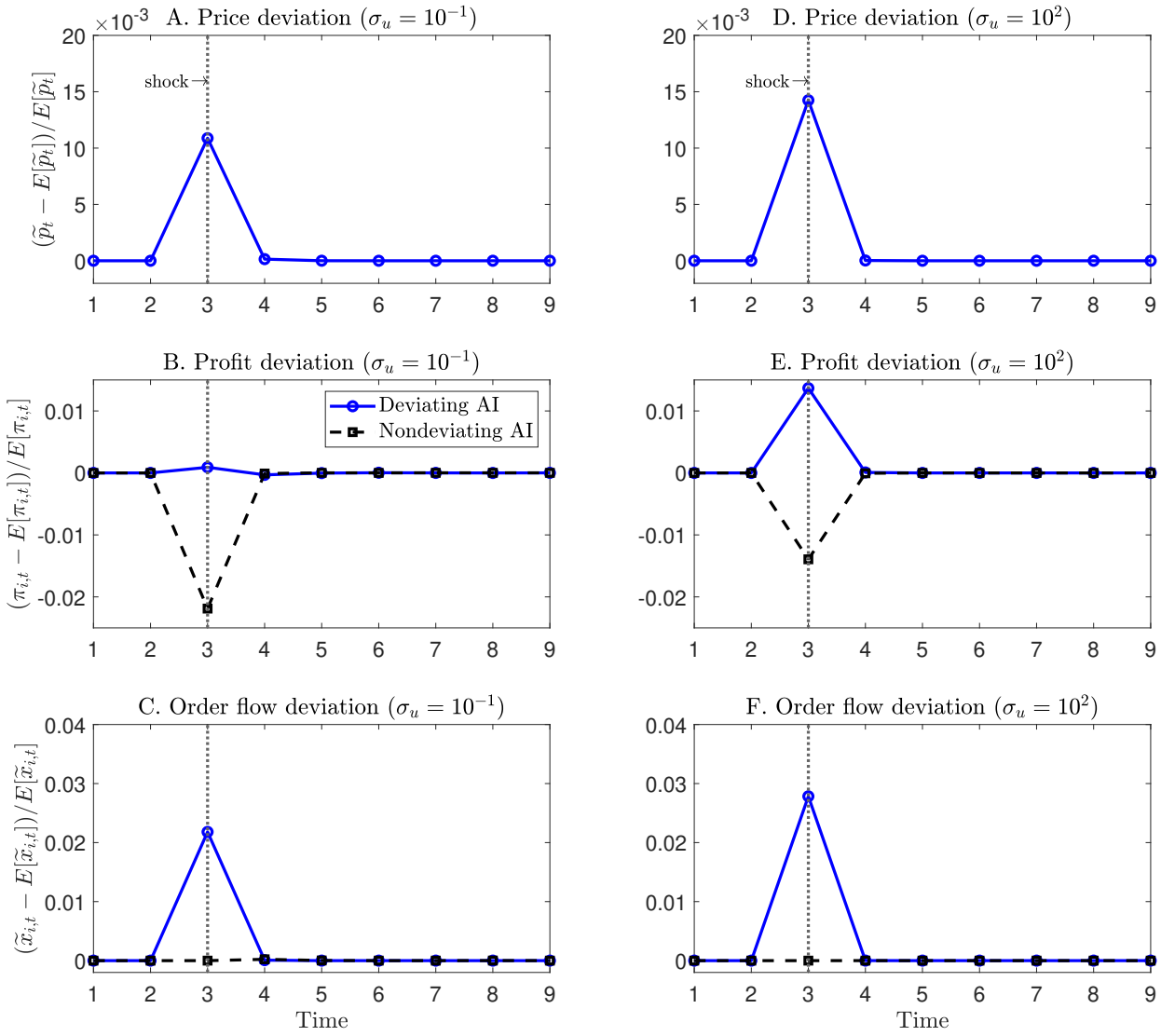
In a parallel comparison to the simulation experiments under the low noise trading risk scenario ($\sigma_u = 10^{-1}$), Panels D through F of Figure 5 show the IRF for the same experiment, conducted under the high noise trading risk scenario ($\sigma_u = 10^2$). Specifically, Panel F illustrates AI speculator i being forced to trade more aggressively at $t = 3$, while the other AI speculator (dashed line) maintains its 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 AI speculator reverts to its original trading order flow, while the non-deviating AI 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 4.2, we formally verify that the AI collusive equilibrium meets the criteria of an 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 a low ξ leads to strong price discovery by market makers.

In a parallel comparison to the simulation experiments with a high ξ value ($\xi = 500$) in Figure 5, Figure 6 presents the IRFs for the same experiment, where an informed AI speculator deviates at $t = 3$ by trading more aggressively (solid line in Panels C and F), 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



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.

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.

the non-deviating AI speculator at $t = 3$ by imposing costs on it, as shown in Panels B and E.

These patterns clearly show that the AI collusive equilibrium in a low-ξ environment is not driven by price-trigger strategies. Instead, it is sustained by over-pruning bias against aggressive strategies, closely resembling a theoretical experience-based equilibrium driven by over-perceived aversion to noise trading risk. In Online Appendix 4.2, following Fershtman and Pakes (2012), we formally verify that the AI collusive outcome meets the criteria for an experience-based equilibrium.

5.5 Intuition Behind AI Collusion and Its Underlying Algorithmic Mechanisms

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 explanations are provided in Online Appendix 3.

Case (i): Low σ_u and High ξ. Why cannot an AI collusive equilibrium sustained by over-pruning bias emerge in such environments? When σ_u is low and ξ is high, noise trading flows have minimal impact on an informed AI speculator's profit. This allows the exploration-exploitation tradeoff to operate effectively, mitigating over-pruning bias against aggressive strategies. Since algorithms rarely incur large losses from noise trading shocks, even when exploring aggressive strategies, these strategies are not prematurely pruned and remain in the learning process. As a result, they are properly evaluated and retained, preventing the emergence of AI collusion through over-pruning bias. Further details are provided in Result 1 of Online Appendix 3.1.1.

We now explain why an AI collusive equilibrium sustained by price-trigger strategies emerges in such environments, focusing on how informed AI speculators achieve it using Q-learning algorithms. High price informativeness is essential, as it ensures that market prices reflect the trading order flow of informed speculators. This allows algorithms to condition their strategies on the unobserved actions of others, indirectly, through observed prices. In these environments, aggressive trading strategies make the market price p_t moves strongly with the fundamental value v_t, and this strong alignment is reflected in the next period's state vector, defined as $s_{t+1} = \{p_t, v_t, v_{t+1}\}$. Conversely, when all algorithms trade conservatively, the price p_t responds only moderately to v_t, and this moderate alignment is similarly captured in the state vector $s_{t+1} = \{p_t, v_t, v_{t+1}\}$. Thus, from the perspective of the next period, the lagged price p_t, as an endogenous state variable, becomes informative about whether all algorithms traded conservatively in period t. This informativeness is a necessary condition for price-trigger strategies to be effective. Model-free Q-learning algorithms do not logically infer the relationship between lagged prices and others' past order flows. Instead, they focus solely on learning the optimal trading strategy for a given state. Nonetheless, their update rules inherently account for how current trading behavior is mechanically connected with the next period's price state, and this connection is incorporated into the Q-value update, as shown in (2.4). This is a process of pattern recognition, not logical reasoning. It fundamentally differs from logic-based human coordination, which requires understanding punishment-for-deviation causality and logically inferring others' actions from prices.

In addition, for algorithms to adopt price-trigger strategies, they must first learn to assign very low Q-values to aggressive trading strategies across all states. This learning process unfolds in two distinct phases. In the early phase, when exploration dominates (i.e., exploration rates remain high for all algorithms), strategies are selected largely at random, and Q-values are updated based on realized payoffs. During this phase, aggressive strategies tend to receive higher Q-values than conservative ones. This is because aggressive strategies yield much higher payoffs when played against opponents who randomly choose to trade aggressively. This asymmetry causes algorithms to assign higher Q-values to aggressive trading strategies than conservative ones early on. As the exploration rate gradually declines to zero, the system transitions into a phase dominated by exploitation. Algorithms begin to consistently choose actions with higher learned Q-values. Thus, early in this exploitation-intensive phase, algorithms continue to favor aggressive strategies inherited from the earlier exploration-intensive phase. They then settle on a state-action pair in which lagged market prices move strongly with lagged fundamentals and trading flows respond aggressively to current private signals about fundamental values. This dynamic persists because even when the system occasionally enters states where lagged prices respond only moderately to

lagged fundamentals, algorithms continue to select aggressive actions, which push the market, in the next period, back to states where lagged prices respond strongly to lagged fundamentals. These actions result in both low immediate profits and weak continuation values. Consequently, over many iterations, the Q-values of aggressive trading strategies gradually decline across all states, whether characterized by lagged prices strongly tracking lagged fundamentals or only moderately responding to them, due to persistently poor outcomes in both immediate profits and continuation values.

Lastly, for algorithms to adopt price-trigger strategies, they must eventually learn to assign very high Q-values to conservative trading in states where lagged prices respond only moderately to lagged fundamentals. How does this occur? As discussed earlier, over many iterations, the Q-values of aggressive strategies in states where lagged prices respond only moderately to lagged fundamentals gradually decline to very low levels and eventually fall below those of conservative strategies in the same states. Once this shift occurs, the algorithms begin to reinforce the state-action pair characterized by lagged prices responding only moderately to fundamentals and conservative trading in response to current fundamental signals. They consistently choose conservative actions in these states, which keeps the market anchored in this region of the state space and yields both high immediate rewards and increasingly strong continuation values. Over iterations, this reinforcement drives the Q-values for this state-action pair to converge to very high levels. See Result 2 in Online Appendix 3.1.1 for details.

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 effectively 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 3.1.2.

However, AI collusion still arises in this environment, but through a different algorithmic mechanism. Specifically, it is sustained by a learning bias that systematically over-prunes aggressive strategies. This bias results from an inherent asymmetry in the way Q-values are updated following noise trading shocks, a generic feature of RL due to its reliance on exploitation.

When noise trading flows move in the same direction as the algorithm's trade, they tend to cause large losses for the algorithm. In response, the algorithm sharply lowers the Q-value of the associated strategy, treats it as a "disastrous action," and avoids selecting it in future iterations, which locks in the downward bias. By contrast, when noise trading flows move in the opposite direction as the algorithm's trade, the algorithm may record large profits and significantly overestimate the Q-value, treating the strategy as a "fantastic action." Because exploitation leads to frequent reuse of high Q-value strategies, the algorithm continually revisits this action, allowing its Q-value to be gradually

corrected through subsequent updates.

In environments where trading outcomes are driven primarily by random noise rather than informed behavior, exploration cannot effectively correct the asymmetry in the learning process caused by the exploitation scheme of RL algorithms. This imbalance between exploration and exploitation leads to the premature pruning of aggressive strategies because their higher exposure to noise trading shocks makes them more susceptible to exploitation-driven undervaluation. As a result, the algorithm converges to a biased Q-value system that systematically favors conservative trading. See Result 4 in Online Appendix 3.1.2 for further discussion.

Case (iii): Low ξ. We now explain why no AI collusive equilibrium sustained by price-trigger strategies can arise when ξ is low, regardless of the level of σ_u. In this setting, the minimal presence of information-insensitive investors forces market makers to prioritize price discovery. As a result, AI speculators must trade conservatively to preserve information rents, leading to endogenously low price informativeness. The equilibrium price becomes dominated by noise trading shocks and fails to serve as a useful state variable for Q-learning algorithms. This lack of price informativeness undermines the sustainability of price-trigger strategies, following the algorithmic mechanism described in case (ii). See Result 5 in Online Appendix 3.2 for further details.

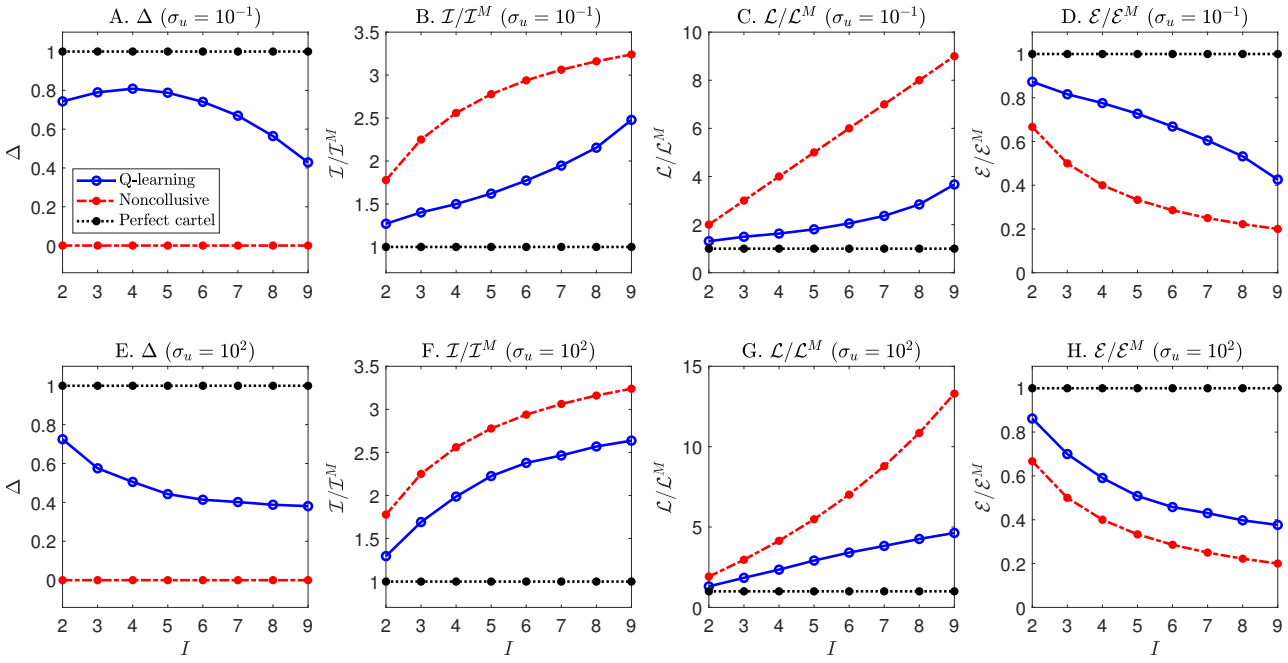
Why, then, can an AI collusive equilibrium sustained by over-pruning bias still arise under the same low-ξ condition? As discussed above, when ξ is low, the equilibrium price is endogenously dominated by noise trading shocks, as in case (ii). Although the underlying reason for low price informativeness differs, the consequence for the RL process is the same: the exploitation-driven learning asymmetry disproportionately penalizes aggressive strategies due to their higher exposure to noise trading shocks. See Result 6 in Online Appendix 3.2 for further details.

5.6 Winners and Losers: The Role of Information-Insensitive Investors

We now examine who gains and who loses from AI collusion, and how this depends on the role of information-insensitive investors, captured by ξ, across three distinct trading environments. In case (i), with high ξ and low σ_u, the AI collusive equilibrium is driven by price-trigger strategies. Here, informed AI speculators primarily trade against information-insensitive investors, who absorb most of their order flow. In the simulation with $\xi = 500$ and $\sigma_u = 10^{-1}$, each informed AI speculator earns an average profit of approximately 54, totaling a loss of about 108 for information-insensitive investors. Noise traders and market makers earn near-zero profits.

In case (ii), with high ξ and high σ_u, the AI collusion is sustained by the over-pruning bias mechanism. Here, informed AI speculators earn supra-competitive profits from trading against both information-insensitive investors and noise traders. In the simulation with $\xi = 500$ and $\sigma_u = 10^2$, each informed AI speculator earns about 54 on average, derived from average losses of 88 from information-insensitive investors and 20 from noise traders. Market makers again break even.

The contrast between $\sigma_u = 10^{-1}$ and $\sigma_u = 10^2$, holding $\xi = 500$ fixed, illustrates the shift in the mechanism sustaining AI collusion, from price-trigger strategies to over-pruning bias. To further explore this shift, we conduct additional simulations under an extreme case with $\sigma_u = 2.5 \times 10^2$. When noise traders submit large orders that generate substantial losses for themselves, information-insensitive investors begin trading more in line with informed AI speculators. In this case, each



Note: Parameters are set according to the baseline economic environment specified in Section 4.2.

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

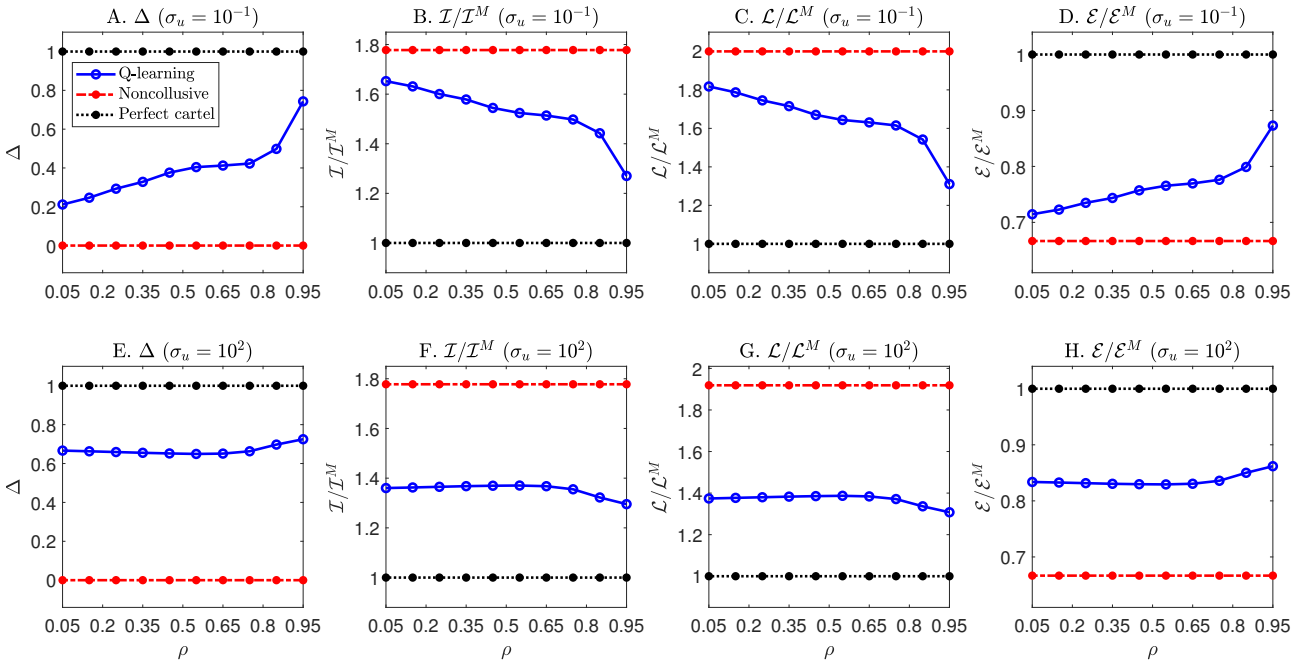
informed AI speculator earns about 54.5, while information-insensitive investors gain roughly 16, together extracting approximately 125 in total losses from noise traders. Market makers continue to earn near-zero profits.

Notably, in our model, information-insensitive investors can be interpreted as retail investors who follow technical analysis (see Section 3.1). Our results align with empirical evidence from Chen, Peng and Zhou (2024), which shows that AI-driven strategies earn profits primarily by exploiting sentiment among retail investors using technical analysis. Their finding that such investors may earn positive trading profits in high-noise environments is also consistent with our simulation outcomes.

In case (iii), with low ξ, AI collusion sustained by price-trigger strategies does not arise. Instead, an AI collusive equilibrium driven by over-pruning bias emerges robustly, similar to case (ii). In this case, informed AI speculators earn supra-competitive profits primarily from trading against noise traders, rather than from exploiting information-insensitive investors. In the simulation with $\xi = 5$ and $\sigma_u = 2$, each informed AI speculator earns about 0.54 on average, derived from average losses of 1.08 from noise traders. Market makers again earn near-zero profits. By design, the role of information-insensitive investors in this environment is negligible.

6 Comparative Statics of AI Equilibrium

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,



Note: Parameters are set according to the baseline economic environment specified in Section 4.2.

Figure 8: Implications of the subjective discount factor.

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 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 decreases, $\mathcal{I}^C/\mathcal{I}^M$ 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.

Effect of Subjective Discount Factor (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 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.

7 Conclusions

This paper shows that AI collusion in securities trading can robustly emerge through two distinct algorithmic mechanisms: one based on price-trigger strategies, and the other driven by over-pruning bias in learning. We characterize the conditions under which each mechanism prevails and show that both correspond to established game-theoretic equilibrium concepts. This highlights a fundamental insight about AI: algorithms relying solely on pattern recognition can exhibit behavior that closely resembles logical and strategic reasoning.

Financial markets differ from product markets in their role as platforms for information aggregation and price discovery, with market makers playing a central role. The over-pruning bias identified in this paper is not the result of specific, nonstandard algorithmic assumptions or limitations, but a generic feature of RL that persists even in sophisticated settings.

These findings raise new and pressing policy and regulatory challenges. While restricting algorithmic complexity or memory capacity may help deter price-trigger AI collusion, such measures can inadvertently exacerbate over-pruning bias by amplifying distorted learning dynamics that prematurely eliminate aggressive yet efficient strategies from the set of potentially optimal options. As a result, well-intentioned constraints may unintentionally undermine market efficiency. Designing effective guardrails for AI in financial markets requires a deep and rigorous understanding of how algorithmic learning dynamics interact with the structure of trading environments to govern machine behavior and shape the resulting AI-driven equilibrium.

This study serves as a proof of concept for analyzing AI-driven manipulation risks in financial markets and opens the door to a broader research agenda. Future work should extend this qualitative framework into a full-scale, data-driven quantitative model, incorporating estimated synthetic trading environments and state-of-the-art RL strengthened by deep learning techniques. Such developments would enable quantitative assessments of AI's impact on market efficiency. In parallel, extending the framework to incorporate bubbles and crashes would offer valuable insights into the role of AI-powered trading in amplifying or dampening market instability.

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.

Banchio, Martino, and Giacomo Mantegazza. 2024. "Artificial Intelligence and Spontaneous Collusion." Working papers.

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.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya. 2025. "Strategic Arbitrage in Segmented Markets." *Journal of Financial Economics*, 166(104008).

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

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicoló, and Sergio Pastorello. 2021. "Algorithmic Collusion with Imperfect Monitoring." *International Journal of Industrial Organization*, 79(C).

Cao, Sean, Wei Jiang, Junbo Wang, and Baozhong Yang. 2024. "From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses." *Journal of Financial Economics*, 160(103910).
Carlin, Bruce Ian, Miguel Sousa Lobo, and S Viswanathan. 2007. "Episodic Liquidity Crises: Cooperative and Predatory Trading." *Journal of Finance*, 62(5): 2235–2274.
Cartea, Álvaro, Patrick Chang, José Penalva, and Harrison Waldon. 2022*a*. "The Algorithmic Learning Equations: Evolving Strategies in Dynamic Games." Working papers.
Cartea, Álvaro, Patrick Chang, Mateusz Mroczka, and Roel Oomen. 2022*b*. "AI-Driven Liquidity Provision in OTC Financial Markets." *Quantitative Finance*, 22(12): 2171–2204.
Charness, Gary, Francesco Feri, Miguel A. Meléndez-Jiménez, and Matthias Sutter. 2014. "Experimental Games on Networks: Underpinnings of Behavior and Equilibrium Selection." *Econometrica*, 82(5): 1615–1670.
Chen, Hsuan-Chi, and Jay R. Ritter. 2000. "The Seven Percent Solution." *Journal of Finance*, 55(3): 1105–1131.
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 papers.
Chen, Hui, Yuhan Cheng, Yanchu Liu, and Ke Tang. 2025. "Teaching Economics to the Machines." Working papers.
Chen, Shuaiyu, Lin Peng, and Dexin Zhou. 2024. "Wisdom or Whims? Decoding Investor Trading Strategies with Large Language Models." Working papers.
Chen, Yifei, Bryan T. Kelly, and Dacheng Xiu. 2024. "Expected Returns and Large Language Models." 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." *Review of Economic Studies*, 69(1): 1–40.
Christie, William G, and Paul H Schultz. 1994. "Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance*, 49(5): 1813–1840.
Christie, William G., and Paul H. Schultz. 1995. "Policy Watch: Did Nasdaq Market Makers Implicitly Collude?" *Journal of Economic Perspectives*, 9(3): 199–208.
Christie, William G, Jeffrey H Harris, and Paul H Schultz. 1994. "Why Did NASDAQ Market Makers Stop Avoiding Odd-Eighth Quotes?" *Journal of Finance*, 49(5): 1841–1860.
Colliard, Jean-Edouard, Thierry Foucault, and Stefano Lovo. 2025. "Algorithmic Pricing and Liquidity in Securities Markets." Working papers.
Cong, Lin, and Zhiguo He. 2019. "Blockchain Disruption and Smart Contracts." *Review of Financial Studies*, 32(5): 1754–1797.
Cooper, David J., and Kai-Uwe Kühn. 2014. "Communication, Renegotiation, and the Scope for Collusion." *American Economic Journal: Microeconomics*, 6(2): 247–278.
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." 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: 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." *Review of Financial Studies*, 35(8): 3867–3921.
Duarte, Victor, Diogo Duarte, and Dejanir H Silva. 2024. "Machine Learning for Continuous-Time Finance." *Review of Financial Studies*, 37(11): 3217–3271.
Dugast, Jérôme, and Thierry Foucault. 2018. "Data Abundance and Asset Price Informativeness." *Journal of Financial Economics*, 130(2): 367–391.
Dugast, Jérôme, and Thierry Foucault. 2024. "Equilibrium Data Mining and Data Abundance." *Journal of Finance*, 80(1): 211–258.
Dutta, Prajit K, and Ananth Madhavan. 1997. "Competition and Collusion in Dealer Markets." *Journal of Finance*, 52(1): 245–276.
Farboodi, Maryam, and Laura Veldkamp. 2020. "Long-Run Growth of Financial Data Technology." *American Economic Review*, 110(8): 2485–2523.
Farboodi, Maryam, and Laura Veldkamp. 2023. "Data and Markets." *Annual Review of Economics*, 15: 23–40.
Fershtman, Chaim, and Ariel Pakes. 2012. "Dynamic Games with Asymmetric Information: A Framework for Empirical Work." *Quarterly Journal of Economics*, 127(4): 1611–1661.
Fonseca, Miguel A., and Hans-Theo Normann. 2012. "Explicit vs. Tacit Collusion: The Impact of Communication in Oligopoly Experiments." *European Economic Review*, 56(8): 1759–1772.
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.
Gao, Zhenyu, Wei Xiong, and Jian Yuan. 2024. "Structured Beliefs and Fund Performance: An LLM-Based Approach." Working papers.
Genesove, David, and Wallace P. Mullin. 2001. "Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case." *American Economic Review*, 91(3): 379–398.
Goldstein, Itay, Chester S Spatt, and Mao Ye. 2021. "Big Data in Finance." *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." *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." *Quarterly Journal of Economics*, 138(4): 2327–2389.
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 via Independent 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: 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." *American Economic Review*, 96(5): 1769–1787.
Holden, Craig W., and Avanidhar Subrahmanyam. 1992. "Long-Lived Private Information and Imperfect Competition." *Journal of Finance*, 47(1): 247–270.
Hörner, Johannes, Stefano Lovo, and Tristan Tomala. 2018. "Belief-Free Price Formation." *Journal of Financial Economics*, 127(2): 342–365.
Johnson, Justin Pappas, Andrew Rhodes, and Matthijs Wildenbeest. 2023. "Platform Design when Sellers Use Pricing Algorithms." *Econometrica*, 91(5): 1841–1879.
Kaniel, Ron, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh. 2023. "Machine-Learning the Skill of Mutual Fund Managers." *Journal of Financial Economics*, 150(1): 94–138.
Kelly, Bryan T., Semyon Malamud, and Kangying Zhou. 2024. "The Virtue of Complexity in Return Prediction." *Journal of Finance*, 79(1): 459–503.
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." *Review of Economic Studies*, 56(3): 317–355.
Kyle, Albert S., and Wei Xiong. 2001. "Contagion as a Wealth Effect." *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.
Lehar, Alfred, and Christine Parlour. 2025. "Market Power and the Bitcoin Protocol." Working papers.
Levine, David K., Thomas R. Palfrey, and Charles R. Plott. 1991. "Siegel's Lemma for Game Players." *Games and Economic Behavior*, 3(2): 147–173.
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." *Journal of Finance*, 55(4): 1705–1765.
Marimon, Ramon, Ellen McGrattan, and Thomas J. Sargent. 1990. "Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents." *Journal of Economic Dynamics and Control*, 14(2): 329–373.
Massarotto, Giovanna. 2025. "Detecting Algorithmic Collusion." *Ohio State Law Journal*, 73.
Mildenstein, Eckart, and Harold Schleef. 1983. "The Optimal Pricing Policy of a Monopolistic Marketmaker in the Equity Market." *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." Working papers.
Rostek, Marzena, and Ji Hee Yoon. 2021. "Dynamic Imperfectly Competitive Markets with Private Information." Working papers.
Rostek, Marzena, and Ji Hee Yoon. 2024. "Imperfect Competition in Financial Markets: Recent Developments." *Journal of Economic Literature*, forthcoming.
Rostek, Marzena, and Marek Weretka. 2012. "Price Inference in Small Markets." *Econometrica*, 80(2): 687–711.
Rostek, Marzena, and Marek Weretka. 2015. "Dynamic Thin Markets." *Review of Financial Studies*, 28(10): 2946–2992.
Rotemberg, Julio J, and Garth Saloner. 1986. "A Supergame-Theoretic Model of Price Wars during Booms." *American Economic Review*, 76(3): 390–407.
Routledge, Bryan R. 1999. "Adaptive Learning in Financial Markets." *Review of Financial Studies*, 12(5): 1165–1202.
Routledge, Bryan R. 2001. "Genetic Algorithm Learning to Choose and Use Information." *Macroeconomic Dynamics*, 5(02): 303–325.
Sannikov, Yuliy, and Andrzej Skrzypacz. 2007. "Impossibility of Collusion under Imperfect Monitoring with Flexible Production." *American Economic Review*, 97(5): 1794–1823.
Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction.* The MIT Press.
Vayanos, Dimitri. 1999. "Strategic Trading and Welfare in a Dynamic Market." *Review of Economic Studies*, 66(2): 219–254.
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.