The Invisible Handshake: Tacit Collusion between Adaptive Market Agents

Luigi Foscari
Università degli Studi di Milano
luigi.foscari@unimi.it

Emanuele Guidotti
University of Lugano and Lake Lucerne Institute
emanuele.guidotti@usi.ch

Nicolò Cesa-Bianchi
Università degli Studi di Milano and Politecnico di Milano
nicolo.cesa-bianchi@unimi.it

Tatjana Chavdarova
Politecnico di Milano
tatjana.chavdarova@polimi.it

Alfio Ferrara
Università degli Studi di Milano
alfio.ferrara@unimi.it

Abstract

We study the emergence of tacit collusion between adaptive trading agents in a stochastic market with endogenous price formation. Using a two-player repeated game between a market maker and a market taker, we characterize feasible and collusive strategy profiles that raise prices beyond competitive levels. We show that, when agents follow simple learning algorithms (e.g., gradient ascent) to maximize their own wealth, the resulting dynamics converge to collusive strategy profiles, even in highly liquid markets with small trade sizes. By highlighting how simple learning strategies naturally lead to tacit collusion, our results offer new insights into the dynamics of AI-driven markets.

1 Introduction

The growing adoption of Artificial Intelligence (AI) in algorithmic trading is transforming financial markets, and understanding how learning systems may autonomously develop coordinated behavior is becoming increasingly important. Recent reports by the Bank for International Settlements and the International Monetary Fund highlight the potential stability risks that may arise from the widespread use of such systems.[1]

Among the most pressing regulatory challenges is the possibility that these AI systems achieve *tacit collusion*. Collusion is a form of coordination where agents align their strategies in a way that benefits all of them at the expense of other market participants or society as a whole. Tacit collusion is achieved without any explicit agreement or instruction. Collusion is illegal under U.S. antitrust law, while tacit collusion falls outside the scope of existing enforcement frameworks, which typically rely on detecting explicit communication or documentary evidence of shared intent Dou et al. (2025). Understanding whether and how tacit collusion may emerge without explicit coordination in AI-driven markets is a first step toward designing effective mitigation strategies. Tacit collusion

[1] See Chapter III of the 2024 Annual Economic Report by the Bank for International Settlement (https://www.bis.org/publ/arpdf/ar2024e.pdf) and Chapter 3 of the 2024 Global Financial Stability Report by the International Monetary Fund (https://www.imf.org/-/media/Files/Publications/GFSR/2024/October/English/textrevised.ashx).

39th Conference on Neural Information Processing Systems (NeurIPS 2025)
Workshop on Algorithmic Collective Action (ACA@NeurIPS 2025).

has been empirically shown to consistently arise in automatic pricing competitions under mild assumptions on the learning dynamics Calvano et al. (2020a). See section 2 for an overview of the related works.

In this work, we study tacit collusion through a two-player game between a *market maker* and a *market taker*. The market maker represents liquidity providers, such as market-making firms or designated market makers, who supply liquidity by continuously posting buy and sell quotes. The market taker represents liquidity consumers, such as hedge funds or pension funds, who demand liquidity by executing trades against these quotes. This stylized framework captures the essential strategic interaction between the two sides of modern financial markets.

Each player starts with a certain amount of cash and inventory, and trade with each other repeatedly seeking to maximize individual *wealth*, defined as the sum of the player's cash holdings and inventory valued at the prevailing market price. The market price evolves by incorporating both the price impact of the players' trades and exogenous shocks. We pose the following question:

In the absence of communication, what strategies do simple wealth-maximizing learners converge to?

Our analysis reveals a key insight. Despite the absence of explicit communication, each trajectory of the gradient ascent dynamic leads to a collusive strategy profile.

2 Related works

Online learning for market making. The problem of optimizing market-making strategies has been extensively studied through the lens of online learning, Abernethy et al. (2013); Abernethy and Kale (2013) linked it to online convex optimization, focusing on developing agents with no-regret learning guarantees, while more recent work Cesa-Bianchi et al. (2025) focused on learning under partial feedback. Other works address specific challenges, such as adapting to sudden market shocks Das and Magdon-Ismail (2008) and analyzing the impact of algorithmic pricing on market liquidity Colliard et al. (2022). Recently, the problem has been extended to the realm decentralized finance, with studies focusing on the optimal design of constant function market makers and strategies for liquidity provision in platforms like Uniswap Bar-On and Mansour (2023). An extensive body of experimental work has successfully applied reinforcement learning techniques to the market-making problem Spooner et al. (2018); Spooner and Savani (2020), using techniques from deep-learning in high-frequency trading Kumar (2023), in multi-agent environments Ganesh et al. (2019) or in the context of limit order books Wei et al. (2019); Coletta et al. (2022).

Online learning for market taking. Online learning for market taking is often framed as the problem of online portfolio selection. Seminal work in this area introduced the concept of universal portfolios Cover and Ordentlich (1996b,a), which are algorithms that perform nearly as well as the best constant-rebalanced portfolio determined in hindsight, without making any statistical assumptions about the market's behavior. Subsequent research focused on developing computationally efficient algorithms to implement these portfolio strategies Kalai and Vempala (2000). Further theoretical work has explored the connections between stochastic and worst-case models for investing Hazan and Kale (2009), providing a more comprehensive understanding of performance guarantees in different market settings.

While these studies primarily focus on the optimization of a single agent's strategy, our work differs by investigating the emergent, coordinated behaviors that arise from the strategic interaction between a market maker and a market taker within a shared market environment.

Multi-Agent Learning Dynamics and algorithmic collusion. Economic problems often involve multiple adaptive agents—so-called agent economies—where strategies evolve simultaneously. A natural tool for modeling such environments is multi-agent reinforcement learning (MARL). Unlike the single-agent case, however, each agent faces a non-stationary environment created by others, and therefore global convergence guarantees are absent (see Daskalakis et al. (2009)). Convergence has been shown in special cases: Q-learning Watkins and Dayan (1992) in two-player zero-sum games Littman (1994a), the iterated Prisoner's Dilemma Sandholm and Crites (1995), and more general arbitrary-sum two-player games assuming Nash equilibrium play Hu and Wellman (1998).

A key strength of Q-learning in these settings is its ability to learn payoffs directly from experience, without requiring prior knowledge of the rewards.

Subsequent work introduced the notion of *foresight* — anticipating the long-term consequences of present actions — to stabilize learning dynamics. Early studies by Tesauro and Kephart (1998, 2000) extended ideas from minimax search and policy iteration, showing that foresight mitigates undesired cyclical pricing behavior Sairamesh and Kephart (2000). In the same spirit, Tesauro and Kephart (2002) investigated adaptive pricing with two competing sellers in electronic marketplaces. Despite the absence of formal guarantees, they showed empirically that simultaneous Q-learning, without explicit coordination, leads to monotonically increasing profit as the discount factor increases.

Later studies turned to more realistic market settings. Waltman and Kaymak (2008) demonstrated that Q-learning agents in repeated Cournot competition reduce output and thereby raise prices relative to the one-shot Nash equilibrium. Calvano et al. (2020a) showed that in oligopolistic pricing environments, reinforcement learning algorithms routinely discover tacitly collusive strategies: supra-competitive pricing is sustained by implicit reward–punishment scheme designed to provide the incentives for firms to consistently price above the competitive level Harrington (2018). Similar results have been observed in airline pricing with Deep Q-learning Mnih et al. (2015) agents, where duopolists learn to split monopoly profits Gu (2023). Recently, Dou et al. (2025) provide simulation evidence that AI-driven speculators autonomously achieve supra-competitive profits, revealing new mechanisms of collusion even in the absence of explicit agreement. Interestingly, in multi-agent repeated auctions, Banchio and Skrzypacz (2022); Banchio and Mantegazza (2023) showed that endowing the players with higher amounts of information, which in turn provide more insight on the long-term dynamics of a strategy, reduces collusion.

3 Model

We consider the following model of price formation:

Assumption 1 (Price formation)**.**

$$P_{t+1} = (P_t + \delta_t)\varepsilon_{t+1}\,, \tag{1}$$

where P_t is the price at time t, δ_t is the price impact of the trade at time t, and $(\varepsilon_t)_t$ is an i.i.d. stochastic process such that $\varepsilon_t > 0$ has finite mean and variance.

We remark that the additivity of the price impact δ_t follows the standard in market microstructure Kyle (1985) while the multiplicativity of ε_{t+1} follows the standard in asset pricing Fama (1970). In the model, ε_{t+1} denotes an exogenous shock, which is typically linked to economic fundamentals Samuelson (1965). Instead, δ_t is endogenous. Following extensive empirical evidence (see e.g., Lillo et al., 2003; Tóth et al., 2011; Mastromatteo et al., 2014; Tóth et al., 2016; Bouchaud et al., 2018), we assume that the price impact is proportional to the square root of the traded quantity:

Assumption 2 (Price impact)**.**

$$\delta_t = \begin{cases} \alpha_t\sqrt{Q_t} & Q_t \geq 0 \\ \beta_t\sqrt{-Q_t} & Q_t < 0 \end{cases} \tag{2}$$

where Q_t is the quantity traded at time t, and $\alpha_t \geq 0$ and $\beta_t \leq 0$ are the proportionality coefficients for buys ($Q_t \geq 0$) and sells ($Q_t < 0$), respectively.

We note that α_t and β_t represent the market's illiquidity, where liquidity is defined in the sense of Black (1971). When $\alpha_t = \beta_t = 0$, we have a perfectly liquid market where the price impact δ_t vanishes, implying that any trade size can be executed without affecting the price. In this limit, the price evolution is driven solely by the exogenous shock ε_{t+1}. When α_t or β_t are large, we have an illiquid market where even moderate trade sizes can substantially move the price.

3.1 Two-Player game

We consider a repeated game between two players: a market maker ($\mathbb{M}$) and a market taker ($\mathbb{T}$). At the beginning of the first round, the price is P_1 and the maker (taker) is initialized with a non-negative amount of cash $C_1^{\mathbb{M}}$ ($C_1^{\mathbb{T}}$) and inventory $I_1^{\mathbb{M}}$ ($I_1^{\mathbb{T}}$). On every round t, the maker sets the (il)liquidity of the market by choosing α_t and β_t. The taker subsequently decides the quantity Q_t to trade (positive for buys and negative for sells).

Trading Protocol 1: Two-player game between *maker* ($\mathbb{M}$) and *taker* ($\mathbb{T}$).

Data: Starting positions $I_1^{\mathbb{M}}, C_1^{\mathbb{M}}$ and $I_1^{\mathbb{T}}, C_1^{\mathbb{T}}$. Initial price P_1.
for *round* $t = 1, 2, \ldots$ **do**
 Maker publishes $\alpha_t \geq 0$ and $\beta_t \leq 0$
 Taker picks $Q_t \in \mathbb{R}$
 Permanent impact δ_t is computed using eq. (2)
 Taker inventory $I_{t+1}^{\mathbb{T}} \leftarrow I_t^{\mathbb{T}} + Q_t$ and cash $C_{t+1}^{\mathbb{T}} \leftarrow C_t^{\mathbb{T}} - Q_t(P_t + \delta_t)$ are updated
 Maker inventory $I_{t+1}^{\mathbb{M}} \leftarrow I_t^{\mathbb{M}} - Q_t$ and cash $C_{t+1}^{\mathbb{M}} \leftarrow C_t^{\mathbb{M}} + Q_t(P_t + \delta_t)$ are updated
 Price P_{t+1} computed using eq. (1) is revealed
end

The trade causes a price impact according to eq. (2). The players exchange the quantity Q_t for an amount of cash equal to $Q_t(P_t + \delta_t)$. This condition means that $P_t + \delta_t$ is the average trade price, which is similar to the fair pricing condition of Farmer et al. (2013).

Finally, the price P_{t+1} from eq. (1) is revealed. The game is summarized in trading protocol 1. Notice that, by construction, the total amount of inventory $I = I_t^{\mathbb{M}} + I_t^{\mathbb{T}}$ and cash $C = C_t^{\mathbb{M}} + C_t^{\mathbb{T}}$ are constant.

Our goal is to study the price dynamics that emerge when the players maximize their own wealth on each round. More precisely, we define the wealth on round t of any player $p \in \{\mathbb{M}, \mathbb{T}\}$ as

$$W_t^p = C_t^p + P_t I_t^p \tag{3}$$

and the objective of each player p is to maximize the expected value of $W_{t+1}^p - W_t^p$.

4 Strategy profiles

The game defined in trading protocol 1 is a general-sum Markov game (MGs, also known as *stochastic game*, Shapley, 1953; Littman, 1994b) where the payoff is defined by the immediate increase in wealth. We are interested in *Markov strategies* for the game, defined for any player $p \in \{\mathbb{M}, \mathbb{T}\}$ on round t as a map $\pi_t^p : \mathcal{S} \to \mathcal{P}(\mathcal{A}_p)$, where $\mathcal{S}$ is the state space, in our case consisting of the amounts of cash and inventory of both players and the price, and $\mathcal{P}(\mathcal{A}_p)$ is the set of all distributions over the actions space of player p. A *strategy profile* $\pi = (\pi^{\mathbb{M}}, \pi^{\mathbb{T}})$ is defined as a pair of strategies, one for the taker and one for the maker.

A fundamental property of the strategy profiles we are interested in is *price positivity*, which describes profiles that keep the market price strictly positive.

Definition 1 (Price positivity)**.** *A strategy profile π is price-positive if for all t, it holds that $P_t > 0$ almost surely with respect to the (possible) internal randomization of π and the noise $(\varepsilon_t)_t$.*

We characterize price-positive strategy profiles as follows

Lemma 1. *[Price positivity characterization] A strategy profile π is price-positive if and only if for all $t \geq 1$ such that $Q_t < 0$ it holds*

$$\beta_t > -\frac{P_t}{\sqrt{-Q_t}}\,. \tag{4}$$

Proof. The proof proceeds by induction, we have $P_1 > 0$ by construction. Regarding the induction step, for every $t \geq 1$ such that $P_t > 0$, we have $P_{t+1} > 0$, where $P_{t+1} = (P_t + \delta_t)\varepsilon_{t+1}$. By assumption 1, $\varepsilon_t > 0$ for all t, therefore $(P_t + \delta_t)\varepsilon_{t+1} > 0$ holds as long as $P_t + \delta_t > 0$. When $Q_t > 0$, by eq. (2), it holds that $\delta_t > 0$. When $Q_t < 0$, we have that $P_t + \beta_t\sqrt{-Q_t} > 0$ by eq. (4). Conversely, if eq. (4) is violated on any round $t^\star$, then $P_{t^\star} < 0$ on such round if $Q_{t^\star} < 0$. □

Additionally, we are interested in *feasible* strategy profiles, which do not force the players into a short position.

Definition 2 (Feasible strategy profile)**.** *A strategy profile π is feasible if it is price-positive and for all t and for both players $p \in \{\mathbb{M}, \mathbb{T}\}$, it holds that $C_t^p > 0$ and $I_t^p > 0$ almost surely with respect to the (possible) internal randomization of π and the noise $(\varepsilon_t)_t$.*

As for price positivity, we provide a characterization of feasible strategy profiles.

Lemma 2. *[Feasibility characterization] A strategy profile is feasible if and only if it is price-positive and for all $t \geq 1$ the following set of inequalities holds:*

$$Q_t(P_t + \delta_t) < C_t^{\mathbb{T}} \quad \textit{for} \quad Q_t \geq 0 \tag{5}$$

$$Q_t < I_t^{\mathbb{M}} \quad \textit{for} \quad Q_t \geq 0 \tag{6}$$

$$-Q_t(P_t + \delta_t) < C_t^{\mathbb{M}} \quad \textit{for} \quad Q_t < 0 \tag{7}$$

$$-Q_t < I_t^{\mathbb{T}} \quad \textit{for} \quad Q_t < 0 \tag{8}$$

This set of inequalities ensures that the traded inventory and cash after each trade never exceed the players' reserves, therefore the players have no need to borrow assets.

The proof of this result is by induction on t and leverages the structure of the trades from trading protocol 1. See appendix A for the full proof.

Finally, we define collusion. Collusion is always defined relative to a benchmark, typically of *perfect competition*. In our setting, we define collusion as price divergence from a benchmark profile such that $\delta_t = 0$ for all $t \geq 1$. There are two reasons for this choice. First, with perfect competition on information, it is known Grossman and Stiglitz (1980) that traded quantities vanish, which in our model corresponds to $Q_t = 0$ and thus $\delta_t = 0$. Second, in the classical view of perfect competition as price-taking Aumann (1964), no individual trade moves prices, which in our model corresponds to $\alpha_t = \beta_t = 0$ and thus $\delta_t = 0$. In both interpretations, $\delta_t = 0$ captures the ideal of a perfectly competitive market, making it a natural baseline against which to define collusive deviations.

Definition 3 (Collusion)**.** *The strategy profile π is collusive if and only if, as $t \to \infty$,*

$$\frac{P_t^\pi}{P_t} \overset{as}{\to} \infty \tag{9}$$

where $\overset{as}{\to}$ denotes almost sure convergence with respect to the (possible) internal randomization of π and the noise $(\varepsilon_t)_t$. P_t^π is the price after t rounds of following π and P_t is the price after t rounds of following any strategy profile that yields $\delta_t = 0$ for all $t \geq 1$.

We define the expected social welfare on round t under any strategy profile π as the total mark-to-market wealth in the system $\mathbb{E}\left[C + P_t I\right]$, where the expectation is taken with respect to the (possible) internal randomization of π and the noise $(\varepsilon_t)_t$. The next result shows that social welfare is maximized by collusive strategies.

Theorem 1. *In trading protocol 1, any collusive strategy profile almost surely achieves higher social welfare than any feasible non-collusive strategy profile.*

Proof. Fix any total inventory I and cash C, let π be any collusive strategy profile and assume by contradiction that π' is any feasible non-collusive strategy profile such that it exists $t_0 < \infty$ for which

$$\frac{C + IP_t^{\pi'}}{C + IP_t^{\pi}} \geq 1 \tag{10}$$

for all $t \geq t_0$ almost surely, meaning that the non-collusive strategy achieves consistently better social welfare. The following inequalities are equivalent

$$\frac{C + IP_t^{\pi'}}{C + IP_t^{\pi}} \geq 1 \iff C + IP_t^{\pi'} \geq C + IP_t^{\pi} \iff P_t^{\pi'} \geq P_t^{\pi} \iff \frac{P_t^{\pi'}}{P_t^{\pi}} \geq 1$$

Introduce the price $P_t^{\mathbf{0}}$ after t rounds such that $\delta_s = 0$ for all $s \in \{1, \ldots, t-1\}$ and consider that

$$\frac{P_t^{\pi'}}{P_t^{\pi}} = \frac{P_t^{\pi'}}{P_t^{\mathbf{0}}} \cdot \frac{P_t^{\mathbf{0}}}{P_t^{\pi}},$$

where the right term goes to zero almost surely by definition 3 and, in order to satisfy eq. (10), it must hold that $P_t^{\pi'}/P_t^{\mathbf{0}} \overset{as}{\to} \infty$ faster π, therefore π' must be collusive as well. □

5 Parameterization

We parameterize the strategy profiles as follows:

Strategy 1. *For any given $\varphi \in [0,1]$, we consider the strategy profile π_φ parameterized by $(k_\alpha, k_\beta, v_\alpha, v_\beta) > 0$. At each time step t, the maker sets the illiquidity parameters*

$$\alpha_t = v_\alpha \cdot \frac{P_t}{\sqrt{A_t}} \qquad \beta_t = -v_\beta \cdot \frac{P_t}{\sqrt{B_t}}$$

and the taker trades quantity

$$Q_t = \begin{cases} +k_\alpha^2 A_t & \textit{with probability} \quad \varphi \\ -k_\beta^2 B_t & \textit{with probability} \quad 1-\varphi \end{cases}$$

where

$$A_t = \min\left\{I_t^{\mathbb{M}}, \frac{C_t^{\mathbb{T}}}{P_t}\right\} \qquad B_t = \min\left\{\frac{C_t^{\mathbb{M}}}{P_t}, I_t^{\mathbb{T}}\right\} \tag{11}$$

The positivity condition on the parameters is motivated by the price concavity assumption (assumption 2) for the maker's parameters (v_α, v_β), while on the taker's parameters it is taken without loss of generality as the set of representable strategy profiles when (k_α, k_β) can be negative is unchanged. A_t and B_t represent the maximum ask and bid which the players can trade without being forced into a short position. The probability φ models the taker's propensity to buy ($\varphi > \frac{1}{2}$) or sell ($\varphi < \frac{1}{2}$).

Next, we characterize the region of the parameter space that coincides with the set of feasible strategy profiles.

Theorem 2. *For any $\varphi \in [0,1]$, Strategy 1 is price-positive and feasible if and only if*

$$v_\alpha \geq 0 \quad v_\beta \geq 0 \quad 0 \leq k_\alpha < f_\alpha(v_\alpha) \quad 0 \leq k_\beta < f_\beta(v_\beta)$$

where

$$f_\alpha(v_\alpha) := \frac{1}{\sqrt[3]{\frac{v_\alpha}{2} + \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}} + \sqrt[3]{\frac{v_\alpha}{2} - \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}}} \in (0,1]$$

and $f_\beta(v_\beta) := {}^1\!/_{v_\beta} \in (0, \infty)$.

The proof of theorem 2 is based on the following argument: from lemma 2 and strategy 1, we can define a set of constraints which define the region of the parameters defining feasible strategy profiles, special care is needed to ensure that the maker can pick the action first. See appendix A for the full proof.

For the remainder of the paper, we consider only feasible strategy profiles. Finally, we provide a necessary and sufficient condition for a feasible strategy profile to be collusive.

Theorem 3. *For any $\varphi \in [0,1]$, let π_φ be a feasible strategy profile parameterized by $(k_\alpha, k_\beta, v_\alpha, v_\beta)$. Introduce*

$$\mu_\eta := \varphi \log(1 + v_\alpha k_\alpha) + (1-\varphi)\log(1 - v_\beta k_\beta) \tag{12}$$

Then strategy profile π_φ is collusive if and only if $\mu_\eta > 0$.

The proof of this theorem is based on the fact that the stochastic process $\frac{P_t^\pi}{P_t}$ under strategy profile π converges in distribution to e^{w_t}, where w_t is a normal distribution with mean $t\mu_\eta$. See appendix A for the full proof of this theorem.

The parameterization from strategy 1 provides the set of all strategy profile Π, we characterized price-positive $\Pi_{\text{price-positive}}$, feasible Π_{feasible} and collusive $\Pi_{\text{collusive}}$ strategy profiles, such that

$$\Pi_{\text{collusive}} \subset \Pi_{\text{feasible}} \subset \Pi_{\text{price-positive}} \subset \Pi\,,$$

where all subsets are proper and $\Pi_{\text{collusive}}$ is not empty.

6 Learning dynamics

We are interested in a pair of strategic taker and maker which on each round update the parameters of their respective strategies to maximize the immediate expected wealth increase

$$\mathbb{E}_t \left[W^p_{t+1} - W^p_t\right] \,, \tag{13}$$

for any player $p \in \{\mathbb{M}, \mathbb{T}\}$ and any round t, where the expectation is taken with respect to the randomization of the taker's strategy, the random noise ε_{t+1} and conditioning on the history up to time t. Call R^p_t the immediate expected wealth increase and note that

$$R^p_t = \mathbb{E}_t \left[W^p_{t+1} - W^p_t\right] = \mathbb{E}_t \left[C^p_{t+1} + P_{t+1} I^p_{t+1} - C^p_t - P_t I^p_t\right] = \mathbb{E}_t \left[(P_{t+1} - P_t) I^p_t\right]$$

where we used the update rules from trading protocol 1 and the expectation is taken with respect to the random draw of φ and conditioning on the state on round t. Now introduce

$$\kappa \coloneqq \mathbb{E}_t \left[\frac{\delta_t}{P_t}\right] \tag{14}$$

where δ_t is from eq. (21). When we want to highlight the dependence on the parameter we write

$$\kappa(\boldsymbol{k}, \boldsymbol{v}) \coloneqq \varphi v_\alpha k_\alpha - (1 - \varphi) v_\beta k_\beta \tag{κ-explicit}$$

for a fixed φ, where we call $\boldsymbol{k} \coloneqq (k_\alpha, k_\beta)$ the parameters of the taker's strategy and $\boldsymbol{v} \coloneqq (v_\alpha, v_\beta)$ the parameters of the maker's strategy. Note that κ depends only on the parameters and is independent of the state. We write the expected price difference on round t as

$$\begin{aligned}
\mathbb{E}_t \left[P_{t+1} - P_t\right] &= \mathbb{E}_t \left[(P_t + \delta_t)\varepsilon_{t+1} - P_t\right] && \text{(eq. (1))} \\
&= \mathbb{E}\left[\varepsilon_{t+1}\right] P_t + \mathbb{E}_t \left[\delta_t\right] \mathbb{E}_t \left[\varepsilon_{t+1}\right] - P_t && \text{(Independence of } \varepsilon_{t+1}) \\
&= \mathbb{E}_t \left[\varepsilon_{t+1}\right] P_t + \kappa\, \mathbb{E}_t \left[\varepsilon_{t+1}\right] P_t - P_t && \text{(eq. (14))} \\
&= (\mu_\varepsilon (1 + \kappa) - 1) P_t
\end{aligned}$$

where μ_ε is the expected value of ε_{t+1} for any t.

For the game defined in trading protocol 1, we consider strategy profiles maximizing the notion of reward defined in eq. (13). Because the expectation is taken conditioned on the history up to round t, each round can be viewed as a general-sum game between the players. We write the utility of the game for any player $p \in \{\mathbb{M}, \mathbb{T}\}$ at time t as a function of the parameters

$$R^p_t(\boldsymbol{k}, \boldsymbol{v}) = \mathbb{E}_t \left[(P_{t+1} - P_t) I^p_t\right] = \left(\mu_\varepsilon (1 + \kappa(\boldsymbol{k}, \boldsymbol{v})) - 1\right) P_t I^p_t \,, \tag{15}$$

We are now ready to formally define the *one-shot* game on any round t for a fixed φ as

Game 1 (one-shot)**.** *We define the sequential one-shot general-sum game played by taker and maker on every round t on the stochastic game defined in trading protocol 1. The maker first picks parameters $\boldsymbol{v} = (v_\alpha, v_\beta) \in [0, \infty)^2$ in the feasible region (theorem 2), the taker responds by picking the parameters $\boldsymbol{k} = (k_\alpha, k_\beta) \in [0, f_\alpha(v_\alpha)] \times [0, f_\beta(v_\beta)]$ in the feasible region. The maker's utility is $R^{\mathbb{M}}_t(\boldsymbol{k}, \boldsymbol{v})$, while the taker's utility is $R^{\mathbb{T}}_t(\boldsymbol{k}, \boldsymbol{v})$.*

Next, we decompose the per-round game (game 1) into a competitive and collaborative component. Consider the game where the players maximize the following utility

$$Z^{\mathbb{T}}_t \coloneqq \mathbb{E}_t \left[(P_{t+1} - P_t)(I^{\mathbb{T}}_t - I^{\mathbb{M}}_t)\right] \quad \text{and} \quad Z^{\mathbb{M}}_t \coloneqq \mathbb{E}_t \left[(P_{t+1} - P_t)(I^{\mathbb{M}}_t - I^{\mathbb{T}}_t)\right] \,, \tag{16}$$

note that $Z^{\mathbb{T}}_{t+1} = -Z^{\mathbb{M}}_{t+1}$. Introduce the following parameterization of the utilities

$$Z^{\mathbb{T}}_t(\boldsymbol{k}, \boldsymbol{v}) = -Z^{\mathbb{M}}_t(\boldsymbol{k}, \boldsymbol{v}) = \left(\mu_\varepsilon (1 + \kappa(\boldsymbol{k}, \boldsymbol{v})) - 1\right) P_t (I^{\mathbb{T}}_t - I^{\mathbb{M}}_t)$$

and define the *competitive one-shot* game:

Game 2 (Competitive)**.** *We define the sequential one-shot zero-sum game played by taker and maker on every round t on the stochastic game defined in trading protocol 1. The maker first picks parameters $\boldsymbol{v} = (v_\alpha, v_\beta) \in [0, \infty)^2$ in the feasible region (theorem 2), the taker responds by picking the parameters $\boldsymbol{k} = (k_\alpha, k_\beta) \in [0, f_\alpha(v_\alpha)] \times [0, f_\beta(v_\beta)]$ in the feasible region. The maker's utility is $Z^{\mathbb{M}}_t(\boldsymbol{k}, \boldsymbol{v})$, while the taker's utility is $Z^{\mathbb{T}}_t(\boldsymbol{k}, \boldsymbol{v})$. The game is zero-sum as $Z^{\mathbb{M}}_t(\boldsymbol{k}, \boldsymbol{v}) = -Z^{\mathbb{T}}_t(\boldsymbol{k}, \boldsymbol{v})$.*

Next, we show that the equilibrium point of this game defines a strategy profile such that $\delta_t = 0$ for all t, which is the benchmark for collusion as per definition 3.

Theorem 4. *Any strategy profile with no asymptotic impact on the price ($\delta_t = 0$) is stable for the competitive one-shot game 2.*

Proof. Consider the competitive game (game 2) for any t and assume without loss of generality that $I_t^{\mathbb{T}} > I_t^{\mathbb{M}}$. Fix any $\boldsymbol{v} = (v_\alpha, v_\beta) \in [0, \infty)^2$, to ensure feasibility it must hold that $0 \le k_\alpha < f_\alpha(v_\alpha)$ and $0 \le k_\beta < f_\beta(v_\beta)$ by theorem 2. By definition of $Z_t^{\mathbb{T}}$ from eq. (16), fix any $\varepsilon > 0$ arbitrarily small, the best response of the taker to $\boldsymbol{v}$ is $k_\alpha = f_\alpha(v_\alpha) - \varepsilon$ and $k_\beta = 0$. The maker then needs to solve the following optimization problem

$$\min_{\boldsymbol{v} \in [0,\infty)^2} Z_t^{\mathbb{T}}(\boldsymbol{v}, \boldsymbol{k}) \qquad \text{where} \quad \boldsymbol{k} = (f_\alpha(v_\alpha) - \varepsilon, 0)$$

Which is solved when $v_\alpha = 0$ (hence $k_\alpha = f_\alpha(0) - \varepsilon = 1 - \varepsilon$) and for any value of v_β. In conclusion, the competitive game admits a solution such that $v_\alpha = 0$ and $k_\beta = 0$, which correspond to the parameterization of strategy profiles such that $\delta_t = 0$ for all t. □

Now introduce the game where the players optimize the utility

$$U_t^p \coloneqq \mathbb{E}_t\left[(P_{t+1} - P_t)I\right] ,$$

as both players have the same utility (p does not appear on the right-hand side), we simply write U_t. Introduce the following parameterization

$$U_t(\boldsymbol{k}, \boldsymbol{v}) = (\mu_\varepsilon(1 + \kappa(\boldsymbol{k}, \boldsymbol{v})) - 1)P_t I \tag{17}$$

and define the *collaborative one-shot* game.

Game 3 (Collaborative)**.** *We define the sequential one-shot game played by taker and maker on every round t on the stochastic game defined in trading protocol 1. The maker first picks parameters $\boldsymbol{v} = (v_\alpha, v_\beta) \in [0, \infty)^2$ in the feasible region (theorem 2), the taker responds by picking the parameters $\boldsymbol{k} = (k_\alpha, k_\beta) \in [0, f_\alpha(v_\alpha)) \times [0, f_\beta(v_\beta))$ in the feasible region. The utility of both players is $U_t(\boldsymbol{k}, \boldsymbol{v})$.*

This game is purely potential Monderer and Shapley (1996) and the potential function is the expected social welfare on every round of trading protocol 1, as

$$\mathbb{E}\left[C + P_t I\right] - (C + P_0 I) = \sum_{s=0}^{t-1} \mathbb{E}\left[C + P_{s+1} I\right] - \mathbb{E}\left[C + P_s I\right] = \sum_{s=0}^{t-1} \mathbb{E}\left[(P_{s+1} - P_s)I\right] = \sum_{s=0}^{t-1} U_s \,.$$

Theorem 1 shows that the social welfare is maximized under a collusive strategy.

Note that we can decompose the utilities $R_t^{\mathbb{T}}$ and $R_t^{\mathbb{M}}$ of the original game 1 into a fully competitive component (game 2) and fully cooperative component (game 3) as

$$R_t^{\mathbb{T}}(\boldsymbol{k}, \boldsymbol{v}) = \frac{1}{2} Z_t^{\mathbb{T}}(\boldsymbol{k}, \boldsymbol{v}) + \frac{1}{2} U_t(\boldsymbol{k}, \boldsymbol{v}) \qquad R_t^{\mathbb{M}}(\boldsymbol{k}, \boldsymbol{v}) = \frac{1}{2} Z_t^{\mathbb{M}}(\boldsymbol{k}, \boldsymbol{v}) + \frac{1}{2} U_t(\boldsymbol{k}, \boldsymbol{v}) \tag{18}$$

for any φ, any round t and any choice of the parameters $(\boldsymbol{k}, \boldsymbol{v})$. By theorems 1 and 4 and definition 3, the competitive component of game 1 aligns with the benchmark for collusion, while the collaborative component aligns with maximizing the social welfare, which is efficiently optimized for in collusive strategies.

Finally, we show that the one-shot game (game 1) is strategically equivalent to the collaborative game (which maximizes the social welfare). By theorem 1, this implies that collusive strategies are Pareto-efficient in game 1 with respect to all feasible non-collusive strategy profiles.

We characterize the connection between game 3 and game 1 in terms of *strategical equivalence* Maschler et al. (2013); Monderer and Shapley (1996); Hwang and Rey-Bellet (2020); Morris and Ui (2004): a pair of two-player games is strategically equivalent if the utility of one game is equal to the utility of the other scaled by a positive constant. To show that this holds, consider any round t and player $p \in \{\mathbb{M}, \mathbb{T}\}$, then we can write the utility of the one-shot game (game 1) proportionally to the utility of the collaborative game (game 3) as

$$R_t^p(\boldsymbol{k}, \boldsymbol{v}) = (\mu_\varepsilon(1 + \kappa(\boldsymbol{k}, \boldsymbol{v})) - 1)P_t I_t^p = \frac{I_t^p}{I} \cdot U_t^p(\boldsymbol{k}, \boldsymbol{v}) \,,$$

where $I_t^p / I > 0$ by feasibility. Strategical equivalence guarantees that the best responses of the players for the two games coincide Monderer and Shapley (1996).

6.1 Learnability of collusive strategy profiles

Next we show that all the learning algorithms in a broad class converge to a collusive strategy in finitely many steps, almost surely. The class mirrors the parameterization of the strategy profiles defined in strategy 1: at each round t, the taker $\mathbb{T}$ decides to ask with probability φ or bid with probability $1-\varphi$. We then combine both players in a single learning algorithm that, at every step t, updates either the *ask* block (v_α, k_α) of the parameters, or the *bid* block (v_β, k_β).

Algorithm Class 1 (randomized block coordinate scheme)**.** *Consider any learning algorithm that updates the strategy profile parameterized by $(k_\alpha^t, k_\beta^t, v_\alpha^t, v_\beta^t)$ at iteration t as follows:*

- *with probability φ picks the α-block (v_α, k_α) and performs an update that* increases *the product $x_t = v_\alpha^t k_\alpha^t$ by some*

$$\Delta_\alpha(t) \in \left[\delta_\alpha^{\min}, \delta_\alpha^{\max}\right] \quad \textit{where} \quad \delta_\alpha^{\min} > 0\,,$$

- *with probability $1-\varphi$ picks the β-block (v_β, k_β) and performs an update that* decreases *the product $y_t = v_\beta^t k_\beta^t$ by some*

$$\Delta_\beta(t) \in \left[\delta_\beta^{\min}, \delta_\beta^{\max}\right] \quad \textit{where} \quad \delta_\beta^{\min} > 0\,.$$

All iterates are assumed to remain feasible as per theorem 2.

Fix $\varphi \in (0,1)$ and define:

$$r(y) \coloneqq (1-y)^{-\frac{1-\varphi}{\varphi}} - 1 \quad \text{and} \quad g(x) \coloneqq 1 - (1+x)^{-\frac{\varphi}{1-\varphi}}\,.$$

Consider any algorithm in the class 1. Let

$$\tau \coloneqq \inf\{t \geq 0\colon \mu_\eta(k_\alpha^t, k_\beta^t, v_\alpha^t, v_\beta^t) > 0\}$$

denote the random time at which a collusive strategy is reached (or ∞ if it is never reached). For an initial point defined by the pair $x_0 = (v_\alpha^0, k_\alpha^0)$ and $y_0 = (v_\beta^0, v_\alpha^0)$, define the nonnegative gaps:

$$G_0^r \coloneqq \max\{0, r(y_0) - x_0\} \quad \text{and} \quad G_0^g \coloneqq \max\{0, y_0 - g(x_0)\}\,.$$

Theorem 5. *[Finite time convergence to a collusive strategy] Consider the alg. class 1. The following statements hold.*

(A) Equivalence of collusive criteria. *For any choice of parameters $(k_\alpha, k_\beta, v_\alpha, v_\beta)$ in the feasible region,*

$$\mu_\eta > 0 \iff x > r(y) \iff y < g(x)$$

where $x = k_\alpha v_\alpha$ and $y = k_\beta v_\beta$.

(B) Expected-time *upper* bounds. *The expected number of iterations to reach a collusive strategy satisfies:*

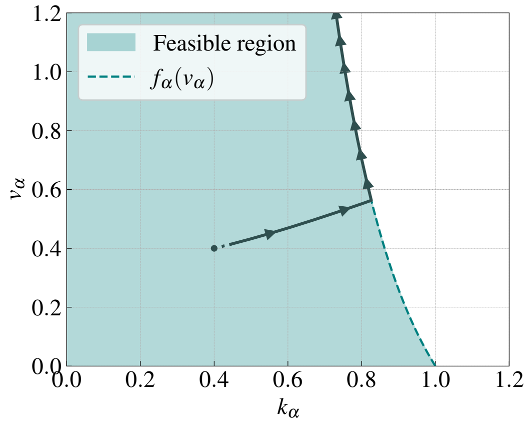
$$\mathbb{E}[\tau] \leq \min\left\{\frac{\lceil G_0^r/\delta_\alpha^{\min}\rceil}{\varphi}, \frac{\lceil G_0^g/\delta_\beta^{\min}\rceil}{1-\varphi}\right\}.$$

The first term in the min *is tight when φ is large (as a large φ makes α-updates more frequent), while the second term is tight for small values of φ.*

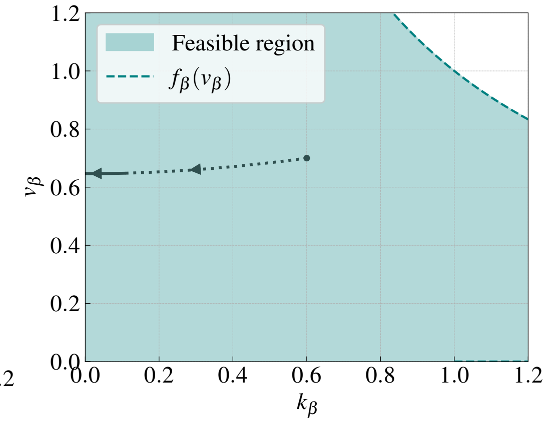
(C) Forward invariance of the collusive region. *Once a collusive strategy profile (x_τ, y_τ) is reached, all subsequent strategies $(x_{\tau+i}, y_{\tau+i})$, with $i = 1, \ldots,$ remain collusive almost surely.*

Since each block update produces a strictly positive change and the collusive region $(x, y) : x > r(y)$ is open, the alg. class 1 reaches a collusive strategy in finitely many steps almost surely.

The equivalence in part (A) follows directly from the definition of μ_η, and expresses the collusiveness condition in terms of the product ratios $x = v_\alpha k_\alpha$ and $y = v_\beta k_\beta$. The properties of the ratios make it explicit that regardless of which block is sampled—α or β—each update reduces the single



(a) Ask parameters



(b) Bid parameters

Figure A: Ask and bid parameters of the two learning agents for non-collusive strategy profiles (dotted line) and collusive strategy profiles (solid line). The experiment used $\varphi = 1/2$ and projected gradient ascent. The shaded region denoted the feasibility region as defined in theorem 2.

scalar gap $r(y) - x$, thereby moving the dynamics closer to the collusive region. Once the gap vanishes, subsequent updates preserve collusiveness since the gap keeps decreasing. The formulation via $g(x)$ provides an equivalent but tighter bound when φ is small. Refer to appendix A for the proof of this theorem.

We next turn to a natural learning dynamics—projected gradient ascent—which is also used in the simulations in the following section (see fig. A).

Corollary 1. *[Projected gradient ascent dynamic] Let $x_0 = v_\alpha^0 k_\alpha^0$, $y_0 = v_\beta^0 k_\beta^0$ be an initial feasible strategy profile. Consider the projected gradient ascent dynamic on the reduced objective $\tilde{\kappa}(\boldsymbol{k}, \boldsymbol{v}) = v_\alpha k_\alpha - v_\beta k_\beta$:*

$$\begin{aligned}
\begin{bmatrix} v_\alpha^{t+1} \\ k_\alpha^{t+1} \end{bmatrix} &= \Pi_{\mathcal{D}_\alpha}\left(\begin{bmatrix} v_\alpha^{t} \\ k_\alpha^{t} \end{bmatrix} + \begin{bmatrix} \eta_{v_\alpha} k_\alpha^{t} \\ \eta_{k_\alpha} v_\alpha^{t} \end{bmatrix}\right), \quad \mathcal{D}_\alpha = [0,\infty) \times [0, f_\alpha(v_\alpha)), \\
\begin{bmatrix} v_\beta^{t+1} \\ k_\beta^{t+1} \end{bmatrix} &= \Pi_{\mathcal{D}_\beta}\left(\begin{bmatrix} v_\beta^{t} \\ k_\beta^{t} \end{bmatrix} - \begin{bmatrix} \eta_{v_\beta} k_\beta^{t} \\ \eta_{k_\beta} v_\beta^{t} \end{bmatrix}\right), \quad \mathcal{D}_\beta = [0,\infty) \times [0, f_\beta(v_\beta)).
\end{aligned} \tag{PGA-$\tilde{\kappa}$}$$

(Euclidean projection) and constant step sizes $\eta_{v_\alpha}, \eta_{k_\alpha}, \eta_{v_\beta}, \eta_{k_\beta} > 0$. Then the statements of theorem 5 hold, and the upper bound (B) holds with $\delta_\alpha^{\min} = \eta_{k_\alpha}(v_\alpha^0)^2$. Consequently,

$$\mathbb{E}\left[\tau\right] \leq \frac{1}{\varphi}\left\lceil \frac{\max\{0, r(y_0) - x_0\}}{\eta_{k_\alpha}(v_\alpha^0)^2} \right\rceil .$$

Proof sketch of Corollary 1. The gradient ascent learning dynamic (PGA-$\tilde{\kappa}$) on $\tilde{\kappa}$ with projection satisfies the assumptions of theorem 5. The corresponding min increment of x is obtained through bounding $x^{t+1} - x^t = (v_\alpha^t + \eta_{v_\alpha} k_\alpha^t)(k_\alpha^t + \eta_{k_\alpha} v_\alpha^t) - v_\alpha^t k_\alpha^t$ and using that v_α^t is nondecreasing. □

We empirically simulate learning market agents that on every round update the parameters of the respective strategies as per alg. class 1 using projected gradient ascent, where the projection operator clips the parameters at the boundary of the feasibility region. Figure A shows that the bid parameters are driven to zero by the learning dynamics, while the ask parameters remain strictly positive.

7 Simulation

To better understand the model from trading protocol 1 and the strategy profiles proposed in strategy 1, we simulate the long-term effects of two fixed strategies on the price and the inventories and cash reserves of the players. Each experiment is run for 3000 rounds using a fixed strategy profile.

At the beginning of each run, each player is given a unit of both cash and inventory, and the initial price is also set to one:

$$I_1^{\mathbb{M}} = I_1^{\mathbb{T}} = C_1^{\mathbb{M}} = C_1^{\mathbb{T}} = P_1 = 1 \ .$$

For the noise, we use a log-normal distribution with unit mean and standard deviation $1/2$, ensuring that the price divergence is not driven by the noise ($\mu_\varepsilon = 0$), but solely on the player's strategy profile. Each experiment is run 10 times, and for each round t we plot the average value and standard deviation of several market features. When appropriate, we also include a running average as a dotted cyan line.

Both strategies parameterized by $k_\alpha = k_\beta = v_\alpha = v_\beta = 1/2$, we changed φ to get a strategy profile π^+ for which $\mu_\eta > 0$ and a strategy profile π^- for which $\mu_\eta < 0$. In figs. B and C we show the results. Note that the figures show the values under the specific strategy profiles π^+ and π^- where the noise has not asymptotic impact on the price because $\mu_\varepsilon = 0$, while the long-term behaviors we show hold for any choice of strategy profile such that $\mu_\eta + \mu_\varepsilon > 0$ or $\mu_\eta + \mu_\varepsilon < 0$ respectively.

As defined in trading protocol 1, on every round, the players exchange a quantity Q_t and cash $(P_t + \delta_t)Q_t$. Figure Bb shows that the traded quantity Q_t converges to zero under π^+, this happens because $Q_t \in (-B_t, A_t)$ and both A_t and B_t tend to zero almost surely at rate $1/P_t$ (see fig. Bd on the right and the definition in eq. (11)). The exchanged cash on the other hand does not converge to any value (fig. Bb on the right). This is crucial because, even if the traded quantity vanishes, the cash exchanged during the trade remains stable, ensuring price growth. We can see from fig. Cb that the opposite happens under π^-: the traded quantity does not converge, while the traded cash converges to zero at rate P_t because $(P_t + \delta_t)Q_t \in (-B_t P_t, A_t P_t)$.

Looking at figs. Bc and Cc, we can see how the behavior of the traded quantity Q_t and traded cash $P_t + \delta_t$ reflects on the inventories $I_t^{\mathbb{M}}, I_t^{\mathbb{T}}$ and cash amounts $C_t^{\mathbb{M}}, C_t^{\mathbb{T}}$ of the players. The inventory is the sum of the traded quantities, therefore if it has a limit, it must lie between 0 and I by feasibility. Under π^+, we already established that Q_t converges to zero at a rate $1/P_t$, which is exponential in time as $P_t = e^{z_t}$ where $z_t \overset{as}{\to} \infty$ (see theorem 3), therefore the inventory converges; the cash on the other hand does not converge.

As per eq. (14), we can see from figs. Bd and Cd that the relative price impact δ_t/P_t equals on average the value of κ computed from the parameters. These plots look quantized due to the averaging done across the runs, for any fixed run the plot would jump between $k_\alpha v_\alpha$ and $-k_\beta v_\beta$.

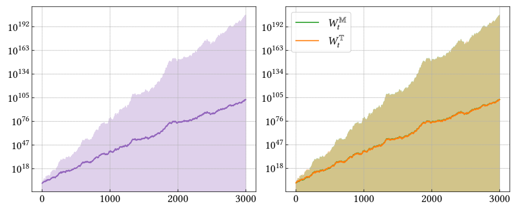
To improve our understanding of the market dynamics, we would need to show that the tradable quantities A_t and B_t are dominated by the term $C_t^{\mathfrak{p}}/P_t$ under any collusive strategy profile. This is true when the limit of the inventory is not zero or if its rate of convergence is slower than the price's. We conjecture that, as $t \to \infty$, the inventory remains strictly positive with probability one.

8 Future works and conclusion

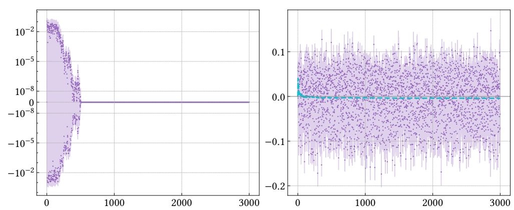
This work can be extended in several key directions. As the current agents are myopic (they optimize for immediate wealth gain), an important next step is to investigate whether collusive equilibria persist when agents adopt strategies that optimize for long-term rewards. Whether collusive behavior persists under farsighted strategic play is still an open question. It has been conjectured by Calvano et al. (2020b) that a higher level of sophistication would increase the collusive tendencies, because of their profit-enhancing characteristics. More recent empirical evidence by Abada et al. (2024) shows that higher levels of sophistication lead to higher competition, which hinders the collusive tendency.

We analyze our model using the widely adopted paradigm of risk-neutral players. Nonetheless, recognizing that real-world market-making agents are generally risk-aware, a compelling future direction is to explicitly incorporate risk-sensitive agents, leveraging established theories like Pratt (1978). This would rigorously test whether the phenomenon of collusion observed in our baseline setting persists or is fundamentally altered when agents actively factor in the volatility of their returns.

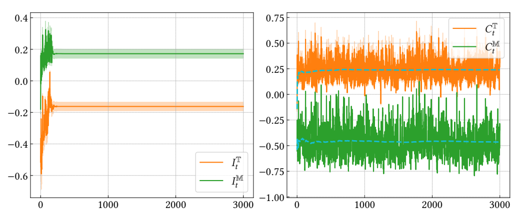
In conclusion, we analyzed a stochastic market model with endogenous price formation and characterized the conditions under which learning agents fall into tacitly collusive behaviors. We proved that when a market maker and a market taker each seek to maximize their immediate wealth, the learning dynamics naturally steer them toward a collusive equilibrium without any need for explicit



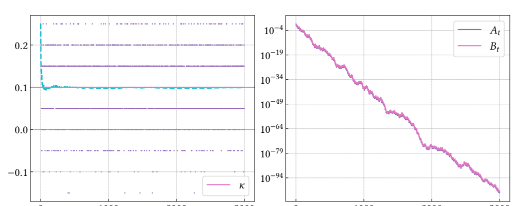
(a) Price P_t (left) and player wealth $W_t^{\mathbb{T}}, W_t^{\mathbb{M}}$ (right).



(b) Traded quantity Q_t (left) and corresponding exchanged cash $P_{t+1}Q_t$ (right).

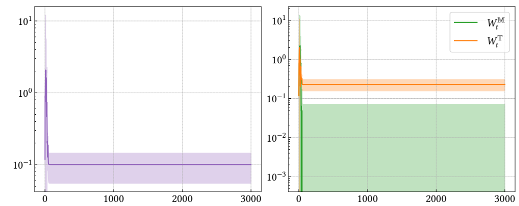


(c) Inventory (left) and cash (right) of the players.

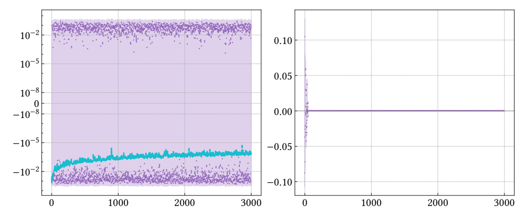


(d) Relative price impact δ_t/P_t (left) and tradable quantities (right).

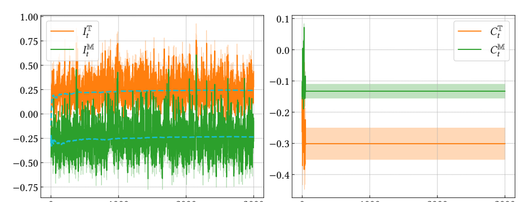
Figure B: **Market impact of a fixed collusive strategy profile with $\varphi = 0.7$ and $k_\alpha = k_\beta = v_\alpha = v_\beta = 1/2$.**



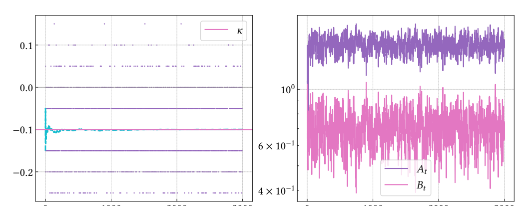
(a) Price P_t (left) and player wealth $W_t^{\mathbb{T}}, W_t^{\mathbb{M}}$ (right).



(b) Traded quantity Q_t (left) and corresponding exchanged cash $P_{t+1}Q_t$ (right).



(c) Inventory (left) and cash (right) of the players.



(d) Relative price impact δ_t/P_t (right) and tradable quantities (right).

Figure C: **Market impact of a fixed *non*-collusive strategy profile with $\varphi = 0.3$ and $k_\alpha = k_\beta = v_\alpha = v_\beta = 1/2$.**

coordination. This outcome results in a systematic upward drift in the asset price, exceeding the influence of market noise.

These findings highlight that implicit collusion can be a natural consequence of rational, wealth-maximizing learning algorithms, raising significant concerns for the design and regulation of automated agents in financial markets.

Acknowledgments and Disclosure of Funding

LF is partially supported by the EU Horizon CL4-2022-HUMAN-02 research and innovation action under grant agreement 101120237, project ELIAS (European Lighthouse of AI for Sustainability).

EG acknowledges support by the Swiss National Science Foundation under Grant No. 233173.

AF is partially supported by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

LF and NCB are partially supported by the MUR PRIN grant 2022EKNE5K (Learning in Markets and Society), by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, investment 1.3, line on Artificial Intelligence), and by the EU Horizon CL4-2022-HUMAN-02 research and innovation action under grant agreement 101120237, project ELIAS (European Lighthouse of AI for Sustainability).

TC acknowledges support by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

References

Ibrahim Abada, Xavier Lambin, and Nikolay Tchakarov. 2024. Collusion by mistake: Does algorithmic sophistication drive supra-competitive profits? *European Journal of Operational Research* 318, 3 (2024), 927–953. Publisher: Elsevier.

Jacob Abernethy, Yiling Chen, and Jennifer Wortman Vaughan. 2013. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation* 1, 2 (May 2013). https://doi.org/10.1145/2465769.2465777 Number of pages: 39 Place: New York, NY, USA Publisher: Association for Computing Machinery tex.articleno: 12 tex.issue_date: May 2013.

Jacob Abernethy and Satyen Kale. 2013. Adaptive market making via online learning. *Advances in Neural Information Processing Systems* 26 (2013).

Robert J Aumann. 1964. Markets with a continuum of traders. *Econometrica* (1964), 39–50.

Martino Banchio and Giacomo Mantegazza. 2023. Adaptive Algorithms and Collusion via Coupling. In *Proceedings of the 24th ACM Conference on Economics and Computation*. ACM, London United Kingdom, 208–208. https://doi.org/10.1145/3580507.3597726

Martino Banchio and Andrzej Skrzypacz. 2022. Artificial Intelligence and Auction Design. https://doi.org/10.48550/arXiv.2202.05947 arXiv:2202.05947 [econ].

Yogev Bar-On and Yishay Mansour. 2023. Uniswap Liquidity Provision: An Online Learning Approach. http://arxiv.org/abs/2302.00610 arXiv:2302.00610 [cs].

Fischer Black. 1971. Toward a fully automated stock exchange, part I. *Financial Analysts Journal* 27, 4 (1971), 28–35.

Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. 2018. *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020a. Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review* 110, 10 (October 2020), 3267–97. https://doi.org/10.1257/aer.20190623

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020b. Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review* 110, 10 (Oct. 2020), 3267–3297. https://doi.org/10.1257/aer.20190623

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Luigi Foscari, and Vinayak Pathak. 2025. Market Making without Regret. arXiv:2411.13993 [cs.GT] https://arxiv.org/abs/2411.13993

Andrea Coletta, Aymeric Moulin, Svitlana Vyetrenko, and Tucker Balch. 2022. Learning to simulate realistic limit order book markets from data as a World Agent. In *Proceedings of the Third ACM International Conference on AI in Finance*. 428–436. https://doi.org/10.1145/3533271.3561753 arXiv:2210.09897 [cs, q-fin].

Jean-Edouard Colliard, Thierry Foucault, and Stefano Lovo. 2022. Algorithmic Pricing and Liquidity in Securities Markets. *SSRN Electronic Journal* (2022). https://doi.org/10.2139/ssrn.4252858 Publisher: Elsevier BV.

T.M. Cover and E. Ordentlich. 1996a. On-line portfolio selection. In *Proceedings of the ninth annual conference on Computational learning theory - COLT '96*. ACM Press, Desenzano del Garda, Italy, 310–313. https://doi.org/10.1145/238061.238161

T.M. Cover and E. Ordentlich. 1996b. Universal portfolios with side information. *IEEE Transactions on Information Theory* 42, 2 (March 1996), 348–363. https://doi.org/10.1109/18.485708

Sanmay Das and Malik Magdon-Ismail. 2008. Adapting to a market shock: optimal sequential market-making. In *Proceedings of the 22nd international conference on neural information processing systems (NIPS'08)*. Curran Associates Inc., Vancouver, British Columbia, Canada and Red Hook, NY, USA, 361–368. Number of pages: 8.

Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. 2009. The complexity of computing a Nash equilibrium. *Commun. ACM* 52, 2 (2009), 89–97. https://doi.org/10.1145/1461928.1461951

Winston Wei Dou, Itay Goldstein, and Yan Ji. 2025. Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper* (2025).

Eugene F Fama. 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance* 25, 2 (1970), 383–417.

J Doyne Farmer, Austin Gerig, Fabrizio Lillo, and Henri Waelbroeck. 2013. How efficiency shapes market impact. *Quantitative Finance* 13, 11 (2013), 1743–1758.

Sumitra Ganesh, Nelson Vadori, Mengda Xu, Hua Zheng, Prashant Reddy, and Manuela Veloso. 2019. Reinforcement Learning for Market Making in a Multi-agent Dealer Market. https://doi.org/10.48550/ARXIV.1911.05892 Version Number: 1.

Sanford J Grossman and Joseph E Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American economic review* 70, 3 (1980), 393–408.

Chengyan Gu. 2023. Deep Q-Learning in Airline Dynamic Pricing and Tacit Collusion: Experimental deep RL in airline revenue management in a duopoly setting. (2023). Paper presented at the American Economic Association (AEA) conference.

Joseph E. Harrington. 2018. Developing Competition Law for Collusion by Autonomous Artificial Agents. *Journal of Competition Law & Economics* 14, 3 (2018), 331–363. https://doi.org/10.1093/joclec/nhy016

Elad Hazan and Satyen Kale. 2009. On stochastic and worst-case models for investing. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS'09)*. Curran Associates Inc., Red Hook, NY, USA, 709–717. event-place: Vancouver, British Columbia, Canada.

Junling Hu and Michael P. Wellman. 1998. Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm. In *Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 242–250.

Sung-Ha Hwang and Luc Rey-Bellet. 2020. Strategic decompositions of normal form games: Zero-sum games and potential games. *Games and Economic Behavior* 122 (July 2020), 370–390. https://doi.org/10.1016/j.geb.2020.05.003

A. Kalai and S. Vempala. 2000. Efficient algorithms for universal portfolios. In *Proceedings 41st annual symposium on foundations of computer science*. 486–491. https://doi.org/10.1109/SFCS.2000.892136

Pankaj Kumar. 2023. Deep reinforcement learning for high-frequency market making. In *Proceedings of the 14th asian conference on machine learning (Proceedings of machine learning research, Vol. 189)*, Emtiyaz Khan and Mehmet Gonen (Eds.). PMLR, 531–546. https://proceedings.mlr.press/v189/kumar23a.html

Albert S. Kyle. 1985. Continuous Auctions and Insider Trading. *Econometrica* 53, 6 (Nov. 1985), 1315. https://doi.org/10.2307/1913210

Fabrizio Lillo, J Doyne Farmer, and Rosario N Mantegna. 2003. Master curve for price-impact function. *Nature* 421, 6919 (2003), 129–130.

Michael L. Littman. 1994a. Markov Games as a Framework for Multi-Agent Reinforcement Learning.. In *ICML*. Morgan Kaufmann, 157–163.

Michael L. Littman. 1994b. Markov games as a framework for multi-agent reinforcement learning. In *ICML (ICML'94)*. Morgan Kaufmann Publishers Inc., 157–163.

Michael Maschler, Eilon Solan, and Shmuel Zamir. 2013. *Game Theory* (1 ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511794216

Iacopo Mastromatteo, Bence Toth, and Jean-Philippe Bouchaud. 2014. Agent-based models for latent liquidity and concave price impact. *Physical Review E* 89, 4 (2014), 042805.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

Dov Monderer and Lloyd S Shapley. 1996. Potential games. *Games and economic behavior* (1996).

Stephen Morris and Takashi Ui. 2004. Best response equivalence. *Games and Economic Behavior* 49, 2 (2004), 260–287. Publisher: Elsevier.

John W. Pratt. 1978. Risk aversion in the small and in the large. In *Uncertainty in Economics*. Elsevier, 59–79. https://doi.org/10.1016/B978-0-12-214850-7.50010-3

Jakka Sairamesh and Jeffrey O Kephart. 2000. Price dynamics and quality in information markets. *Decision Support Systems* 28, 1 (2000), 35–47. https://doi.org/10.1016/S0167-9236(99)00073-1

Paul A. Samuelson. 1965. Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review* 6, 2 (1965), 41.

Tuomas Sandholm and Robert H. Crites. 1995. On Multiagent Q-Learning in a Semi-Competitive Domain. In *Adaption and Learning in Multi-Agent Systems*. Springer Berlin Heidelberg, 191–205.

Lloyd S. Shapley. 1953. Stochastic Games*. *Proceedings of the National Academy of Sciences* 39, 10 (1953), 1095–1100. https://doi.org/10.1073/pnas.39.10.1095

Thomas Spooner, John Fearnley, Rahul Savani, and Andreas Koukorinis. 2018. Market making via reinforcement learning. In *Proceedings of the 17th international conference on autonomous agents and MultiAgent systems (Aamas '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, Sweden and Richland, SC, 434–442. Number of pages: 9.

Thomas Spooner and Rahul Savani. 2020. Robust market making via adversarial reinforcement learning. In *Proceedings of the 19th international conference on autonomous agents and MultiAgent systems (Aamas '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Auckland, New Zealand and Richland, SC, 2014–2016. Number of pages: 3.

Gerald Tesauro and Jeffrey O. Kephart. 2002. Pricing in agent economies using multi-agent Q-learning. *Autonomous Agents and Multi-Agent Systems* 5, 3 (2002), 289–304. https://doi.org/10.1023/A:1015504423309

Gerald J. Tesauro and Jeffrey O. Kephart. 1998. Foresight-based pricing algorithms in an economy of software agents. In *Proceedings of the First International Conference on Information and Computation Economies*. Association for Computing Machinery, New York, NY, USA, 37–44. https://doi.org/10.1145/288994.289002

Gerald J. Tesauro and Jeffrey O. Kephart. 2000. Foresight-based pricing algorithms in agent economies. *Decision Support Systems* 28, 1 (2000), 49–60. https://doi.org/10.1016/S0167-9236(99)00074-3

Bence Tóth, Zoltán Eisler, and J-P Bouchaud. 2016. The Square-Root Impace Law Also Holds for Option Markets. *Wilmott* 2016, 85 (2016), 70–73.

Bence Tóth, Yves Lemperiere, Cyril Deremble, Joachim De Lataillade, Julien Kockelkoren, and J-P Bouchaud. 2011. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X* 1, 2 (2011), 021006.

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

Christopher J. C. H. Watkins and Peter Dayan. 1992. Q-learning. *Machine Learning* 8, 3 (May 1992), 279–292. https://doi.org/10.1007/BF00992698

Haoran Wei, Yuanbo Wang, Lidia Mangu, and Keith Decker. 2019. Model-based Reinforcement Learning for Predictions and Control for Limit Order Books. http://arxiv.org/abs/1910.03743 arXiv:1910.03743 [cs].

A Technical Appendix

In this section, we present the remaining proofs of the results presented in the paper.

A.1 Proof of Theorem 3

Theorem 3. *For any $\varphi \in [0,1]$, let π_φ be a feasible strategy profile parameterized by $(k_\alpha, k_\beta, v_\alpha, v_\beta)$. Introduce*

$$\mu_\eta \coloneqq \varphi \log(1 + v_\alpha k_\alpha) + (1-\varphi)\log(1 - v_\beta k_\beta) \tag{12}$$

Then strategy profile π_φ is collusive if and only if $\mu_\eta > 0$.

Proof. Fix any starting configuration $(P_1,\, I_1^{\mathbb{T}},\, C_1^{\mathbb{T}},\, I_1^{\mathbb{M}},\, C_1^{\mathbb{M}})$. Call P_{t+1}^π the price after t rounds following strategy π_φ, and P_{t+1} the price after t rounds following any strategy such that $\delta_{t'} = 0$ for all $t' \geq 1$ (which can be achieved in strategy 1 by setting, for instance, $k_\alpha = k_\beta = 0$). By construction, for all $t \geq 1$, it holds that

$$P_{t+1}^\pi = (P_t^\pi + \delta_t)\varepsilon_{t+1} \qquad P_{t+1} = P_t \varepsilon_{t+1}$$

Now consider the ratio

$$\frac{P_{t+1}^\pi}{P_t^\pi} = \frac{(P_t^\pi + \delta_t)\varepsilon_{t+1}}{P_t^\pi} = \left(1 + \frac{\delta_t}{P_t^\pi}\right)\varepsilon_{t+1}$$

and define

$$z_{t+1}^\pi \coloneqq \log \frac{P_{t+1}^\pi}{P_1} = \sum_{s=1}^{t} \log \frac{P_{s+1}^\pi}{P_s^\pi} = \sum_{s=1}^{t} \log\left(1 + \frac{\delta_s}{P_s^\pi}\right) + \sum_{s=1}^{t} \log \varepsilon_{s+1}$$

and

$$\eta_t \coloneqq 1 + \frac{\delta_t}{P_t^\pi} = \begin{cases} 1 + v_\alpha k_\alpha & \text{w.p.} \quad \varphi \\ 1 - v_\beta k_\beta & \text{w.p.} \quad 1-\varphi \end{cases} \tag{19}$$

Notice that $1 - v_\beta k_\beta > 0$ by theorem 2 and therefore $\eta_t > 0$. Because both η_t and ε_{t+1} are positive i.i.d. random variables, call

$$\mu_\eta \coloneqq \mathbb{E}\left[\log \eta_t\right] \quad \sigma_\eta^2 \coloneqq \mathrm{Var}\left(\log \eta_t\right)$$
$$\mu_\varepsilon \coloneqq \mathbb{E}\left[\log \varepsilon_{t+1}\right] \quad \sigma_\varepsilon^2 \coloneqq \mathrm{Var}\left(\log \varepsilon_{t+1}\right)$$

As $t \to \infty$, by the central limit theorem we have

$$\sum_{s=1}^{t} \log \eta_s \xrightarrow{d} N(t\mu_\eta, t\sigma_\eta^2) \qquad \sum_{s=1}^{t} \log \varepsilon_s \xrightarrow{d} N(t\mu_\varepsilon, t\sigma_\varepsilon^2)$$

where N is the normal distribution. Therefore

$$z_{t+1}^\pi \xrightarrow{d} N(t(\mu_\eta + \mu_\varepsilon),\, t(\sigma_\eta^2 + \sigma_\varepsilon^2))$$

Regarding the baseline, for any t, we have

$$\frac{P_{t+1}}{P_t} = \frac{P_t \varepsilon_{t+1}}{P_t} = \varepsilon_{t+1}$$

and

$$z_{t+1} \coloneqq \log \frac{P_{t+1}}{P_1} = \sum_{s=1}^{t} \log \frac{P_{s+1}}{P_s} = \sum_{s=1}^{t} \log \varepsilon_{s+1} \xrightarrow{d} N(t\mu_\varepsilon, t\sigma_\varepsilon^2)$$

Now define the random variable

$$w_t \coloneqq \log \frac{P_t^\pi}{P_t} = \log \frac{P_t^\pi}{P_1} - \log \frac{P_t}{P_1} = z_{t+1}^\pi - z_{t+1} \xrightarrow{d} N(t\mu_\eta, t\sigma_\eta^2)$$

Next, define the independent events $E_t \coloneqq \{w_t \leq M\}$ for any t and any M, and consider the sum of their probabilities

$$\sum_{t=1}^{\infty} \mathbb{P}\left(E_t\right) = \sum_{t=1}^{\infty} \Phi\left(\frac{M - t\mu_\eta}{\sigma_\eta \sqrt{t}}\right), \tag{20}$$

where we approximated $\mathbb{P}\left(E_t\right)$ by the central limit theorem using the cumulative distribution function Φ of the normal distribution. Because the argument of the sum decays exponentially fast as $t \to \infty$ if $\mu_\eta > 0$, the sum is bounded and, by the Borel-Cantelli lemma, E_t must occur a finite number of times almost surely, therefore, choosing $M > 0$ arbitrarily away from zero, $w_t \overset{as}{\to} \infty$ and $P_t^\pi / P_t = e^{w_t} \overset{as}{\to} \infty$.

If $\mu_\eta < 0$, then eq. (20) is not bounded as the argument approaches one and, by the second Borel-Cantelli lemma, the event E_t occurs infinitely often. Choosing $M < 0$ arbitrarily away from zero, it holds that $w_t \overset{as}{\to} -\infty$ and $P_t^\pi / P_t = e^{w_t} \overset{as}{\to} 0$.

If $\mu_\eta = 0$, then w_t has zero mean and no almost sure limit. In conclusion, a strategy profile is collusive if and only if $\mu_\eta > 0$. □

A.2 Proof of Lemma 2

Lemma 2. *[Feasibility characterization] A strategy profile is feasible if and only if it is price-positive and for all $t \geq 1$ the following set of inequalities holds:*

$$Q_t(P_t + \delta_t) < C_t^{\mathbb{T}} \quad \textit{for} \quad Q_t \geq 0 \tag{5}$$

$$Q_t < I_t^{\mathbb{M}} \quad \textit{for} \quad Q_t \geq 0 \tag{6}$$

$$-Q_t(P_t + \delta_t) < C_t^{\mathbb{M}} \quad \textit{for} \quad Q_t < 0 \tag{7}$$

$$-Q_t < I_t^{\mathbb{T}} \quad \textit{for} \quad Q_t < 0 \tag{8}$$

This set of inequalities ensures that the traded inventory and cash after each trade never exceed the players' reserves, therefore the players have no need to borrow assets.

Proof. We show that if all the inequalities hold, then any price-positive strategy profile is feasible. First, by the assumption on price positivity (definition 1), it holds that $P_t > 0$ for any t almost surely.

The proof is by induction. At time $t = 1$ we have $C_1^p > 0$, $I_1^p > 0$ by construction. Regarding the induction step, for every $t \geq 1$ such that $C_t^p > 0$, $I_t^p > 0$, we have $C_{t+1}^p > 0$, $I_{t+1}^p > 0$ as shown below:

- For $Q_t \geq 0$:
  - $C_{t+1}^{\mathbb{M}} = C_t^{\mathbb{M}} + Q_t(P_t + \delta_t) > 0$ as $C_t^{\mathbb{M}} > 0$ and $P_t + \delta_t > 0$.
  - $C_{t+1}^{\mathbb{T}} = C_t^{\mathbb{T}} - Q_t(P_t + \delta_t) > 0$ by (5).
  - $I_{t+1}^{\mathbb{M}} = I_t^{\mathbb{M}} - Q_t > 0$ by (6).
  - $I_{t+1}^{\mathbb{T}} = I_t^{\mathbb{T}} + Q_t > 0$ as $I_t^{\mathbb{T}} > 0$.
- For $Q_t < 0$:
  - $C_{t+1}^{\mathbb{M}} = C_t^{\mathbb{M}} + Q_t(P_t + \delta_t) > 0$ by (7).
  - $C_{t+1}^{\mathbb{T}} = C_t^{\mathbb{T}} - Q_t(P_t + \delta_t) > 0$ as $C_t^{\mathbb{T}} > 0$ and $P_t + \delta_t > 0$.
  - $I_{t+1}^{\mathbb{M}} = I_t^{\mathbb{M}} - Q_t > 0$ as $I_t^{\mathbb{M}} > 0$.
  - $I_{t+1}^{\mathbb{T}} = I_t^{\mathbb{T}} + Q_t > 0$ by (8).

Conversely, we show that if any of the inequalities is violated, then the strategy profile is not feasible. If (5) is violated for $Q_{t^*} \geq 0$ then $C_{t^*+1}^{\mathbb{T}} < 0$. If (6) is violated for $Q_{t^*} \geq 0$ then $I_{t^*+1}^{\mathbb{M}} < 0$. If (7) is violated for $Q_{t^*} < 0$ then $C_{t^*+1}^{\mathbb{M}} < 0$. If (8) is violated for $Q_{t^*} < 0$ then $I_{t^*+1}^{\mathbb{T}} < 0$. If Price positivity is violated for $Q_{t^*} < 0$ then $C_{t^*+1}^{\mathbb{T}} < 0$. □

A.3 Proof of Theorem 2

Theorem 2. *For any $\varphi \in [0,1]$, Strategy 1 is price-positive and feasible if and only if*

$$v_\alpha \geq 0 \quad v_\beta \geq 0 \quad 0 \leq k_\alpha < f_\alpha(v_\alpha) \quad 0 \leq k_\beta < f_\beta(v_\beta)$$

where

$$f_\alpha(v_\alpha) \coloneqq \frac{1}{\sqrt[3]{\frac{v_\alpha}{2} + \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}} + \sqrt[3]{\frac{v_\alpha}{2} - \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}}} \in (0,1]$$

and $f_\beta(v_\beta) \coloneqq 1/v_\beta \in (0, \infty)$.

Proof. Pick any $\varphi \in [0,1]$ and let π_φ be a strategy profile parameterized by $(k_\alpha, k_\beta, v_\alpha, v_\beta) > 0$. By eq. (2) we have:

$$\delta_t = \begin{cases} +v_\alpha k_\alpha P_t & \text{w.p.} \quad \varphi \\ -v_\beta k_\beta P_t & \text{w.p.} \quad 1 - \varphi\,. \end{cases} \tag{21}$$

We require π_φ to be feasible. Assume that $P_1 > 0$, $I_1^{\mathbb{M}} > 0$, $I_1^{\mathbb{T}} > 0$, $C_1^{\mathbb{M}} > 0$ and $C_1^{\mathbb{T}} > 0$. Consider any round t.

Price positivity is achieved as per lemma 1 when

$$\beta > -\frac{P_t}{\sqrt{-Q_t}} \iff -v_\beta \cdot \frac{P_t}{\sqrt{B_t}} > -\frac{P_t}{k_\beta B_t} \iff v_\beta k_\beta < 1 \tag{22}$$

The feasibility characterization from lemma 2 implies that a strategy profile is feasible if and only if the following set of inequalities is satisfied:

$$k_\alpha^2 A_t (P_t + v_\alpha k_\alpha P_t) < C_t^{\mathbb{T}} \quad \text{for} \quad Q_t \geq 0 \tag{23}$$

$$k_\alpha^2 A_t < I_t^{\mathbb{M}} \quad \text{for} \quad Q_t \geq 0 \tag{24}$$

$$k_\beta^2 B_t (P_t - v_\beta k_\beta P_t) < C_t^{\mathbb{M}} \quad \text{for} \quad Q_t < 0 \tag{25}$$

$$k_\beta^2 B_t < I_t^{\mathbb{T}} \quad \text{for} \quad Q_t < 0 \tag{26}$$

From (23) we have:

$$P_t k_\alpha^2 A_t (1 + v_\alpha k_\alpha) \leq k_\alpha^2 C_t^{\mathbb{T}} (1 + v_\alpha k_\alpha) < C_t^{\mathbb{T}} \iff k_\alpha^2 + v_\alpha k_\alpha^3 < 1\,. \tag{27}$$

From (24) we have:

$$k_\alpha^2 A_t \leq k_\alpha^2 I_t^{\mathbb{M}} < I_t^{\mathbb{M}} \iff k_\alpha < 1\,. \tag{28}$$

From (25) we have:

$$P_t k_\beta^2 B_t (1 - v_\beta k_\beta) \leq k_\beta^2 C_t^{\mathbb{M}} (1 - v_\beta k_\beta) < C_t^{\mathbb{M}} \iff k_\beta^2 - v_\beta k_\beta^3 < 1\,. \tag{29}$$

From (26) we have:

$$k_\beta^2 B_t \leq k_\beta^2 I_t^{\mathbb{T}} < I_t^{\mathbb{T}} \iff k_\beta < 1\,. \tag{30}$$

Notice that (29) is redundant because it is satisfied by any $v_\beta \geq 0$ when $k_\beta \leq 1$. The constraints obtained are

$$0 \leq v_\alpha < \frac{1 - k_\alpha^2}{k_\alpha^3} \quad 0 \leq v_\beta < \frac{1}{k_\beta} \quad 0 \leq k_\alpha < 1 \quad 0 \leq k_\beta < 1 \tag{31}$$

By the definition of trading protocol 1, the maker picks their parameters before the taker, to reflect that we require a formulation equivalent to eq. (31), but of the form

$$v_\alpha \geq 0 \quad v_\beta \geq 0 \quad 0 \leq k_\alpha < f_\alpha(v_\alpha) \quad 0 \leq k_\beta < f_\beta(v_\beta)$$

To find f_α, recall eq. (27) and write the constraint as $k_\alpha^3 v_\alpha + k_\alpha^2 - 1 < 0$. For a fixed $v_\alpha \geq 0$, the function $g(k_\alpha) = k_\alpha^3 v_\alpha + k_\alpha^2 - 1$ is strictly increasing in $k_\alpha \geq 0$ and therefore there is only one value $f_\alpha(v_\alpha) \in [0,1]$ such that $g(f_\alpha(v_\alpha)) = 0$. To find $f_\alpha(v_\alpha)$ we need to solve the equation $g(x) = x^3 v_\alpha + x^2 - 1 = 0$. Consider the variable swap $t = \frac{1}{x}$, thus $g(1/t) = t^3 - t - v_\alpha = 0$. We are interested in the root $t(v_\alpha) \geq 1$:

$$t(v_\alpha) \coloneqq \sqrt[3]{\frac{v_\alpha}{2} + \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}} + \sqrt[3]{\frac{v_\alpha}{2} - \sqrt{\frac{v_\alpha^2}{4} - \frac{1}{27}}}$$

Finally, applying the variable swap again we get $f_\alpha(v_\alpha) \coloneqq \frac{1}{t(v_\alpha)}$. To find f_β, simply consider the region $v_\beta \geq 0$ and from (22) we get $f_\beta(v_\beta) \coloneqq \frac{1}{v_\beta}$. $\square$

A.4 Proof of Theorem 5

Below, we present an extended version of Theorem 5, which additionally establishes a lower bound on the number of steps required to reach a collusive strategy profile, in statement (D), and provide its proof.

Recall the following definitions. Fix $\varphi \in (0,1)$ and define:

$$r(y) \coloneqq (1-y)^{-\frac{1-\varphi}{\varphi}} - 1 \quad \text{and} \quad g(x) \coloneqq 1-(1+x)^{-\frac{\varphi}{1-\varphi}}\,.$$

Consider the class of algorithms as per algorithm class 1. Let

$$\tau \coloneqq \inf\{t \geq 0 \colon \mu_\eta(v_\alpha^t, k_\alpha^t, v_\beta^t, k_\beta^t) > 0\}$$

denote the random time at which a collusive strategy is reached (could be ∞), as per Theorem 3.

For an initial point defined by the pair $x_0 = v_\alpha^0, k_\alpha^0$ and $y_0 = (v_\beta^0, v_\alpha^0)$, define the nonnegative gaps:

$$G_0^r \coloneqq \max\{0, r(y_0) - x_0\} \quad \text{and} \quad G_0^g \coloneqq \max\{0, y_0 - g(x_0)\}\,.$$

We next state the extended theorem 5) and provide its proof.

Theorem 6. *[(extended theorem 5) finite time convergence to collusive strategy] Consider the alg. class 1. Then the following statements hold.*

(A) Equivalence of collusive criteria. *Be $(v_\alpha, k_\alpha, v_\beta, k_\beta)$ a set of parameters in the feasible region as per theorem 2, the following are equivalent:*

$$\mu_\eta > 0 \iff x > r(y) \iff y < g(x)\,.$$

where $x = k_\alpha v_\alpha$ and $y = k_\beta v_\beta$.

(B) Expected-time* upper *bounds. *The expected number of iterations to collusive strategies satisfies:*

$$\mathbb{E}[\tau] \leq \min\left\{ \frac{\lceil G_0^r/\delta_\alpha^{\min}\rceil}{\varphi}, \frac{\lceil G_0^g/\delta_\beta^{\min}\rceil}{1-\varphi} \right\}.$$

The left term is tight when φ is large (large φ makes α-updates frequent), while the right is tight for small φ.

(C) Forward invariance of the collusive region. *Let*

$$\mathcal{C} \coloneqq \{(x,y) : x > r(y)\}$$

(or, equivalently, $\{(x,y) : y < g(x)\}$) be the set of parameter products $x = k_\alpha v_\alpha$ and $y = k_\beta v_\beta$ which define collusive strategy profiles. Under the stated update rules and feasibility constraints, $\mathcal{C}$ is forward-invariant: if $(x_t, y_t) \in \mathcal{C}$ for some t, then $(x_{t+1}, y_{t+1}) \in \mathcal{C}$ almost surely. In particular, once a collusive strategy is reached, all subsequent strategies remain collusive almost surely.

(D) Expected-time* lower *bounds. *Let*

$$\begin{aligned}\Delta r_0^{\max} &\coloneqq r(y_0) - r(y_0 - \delta_\beta^{\max}) \quad \text{and} \\ \Delta g_0^{\max} &\coloneqq g(x_0 + \delta_\alpha^{\max}) - g(x_0)\,.\end{aligned}$$

Then the expected number of iterations to collusive strategies satisfies:

$$\mathbb{E}\left[\tau\right] \geq \max\left(\frac{G_0^r}{\varphi\delta_\alpha^{\max} + (1-\varphi)\Delta r_0^{\max}}, \frac{G_0^g}{\varphi \Delta g_0^{\max} + (1-\varphi)\delta_\beta^{\max}} \right).$$

Proof.

(A) Equivalence of collusive criteria. From theorem 3 we have:

$$\mu_\eta \coloneqq \varphi \log(1 + v_\alpha k_\alpha) + (1-\varphi)\log(1 - v_\beta k_\beta) > 0\,.$$

Thus,

$$\varphi \log(1+v_\alpha k_\alpha) > -(1-\varphi)\log(1-v_\beta k_\beta)\,,$$
$$\log(1+v_\alpha k_\alpha) > -\frac{(1-\varphi)}{\varphi}\log(1-v_\beta k_\beta)\,,$$
$$\log(1+v_\alpha k_\alpha) > \log(1-v_\beta k_\beta)^{-\frac{(1-\varphi)}{\varphi}}\,,$$
$$v_\alpha k_\alpha > (1-v_\beta k_\beta)^{-\frac{(1-\varphi)}{\varphi}} - 1\,,$$

where we used $\varphi \in (0,1)$. Denoting $x \coloneqq v_\alpha k_\alpha$ and $y \coloneqq v_\beta k_\beta$ yields the equivalent criteria $x > r(y)$ with $r(y) \coloneqq (1-y)^{-\frac{1-\varphi}{\varphi}} - 1$. We notice that as y decreases, $r(y)$ also decreases, making the condition easier to satisfy. Since the increase of x affects the criteria linearly, when y is closer to 1, the y updates are more effective toward satisfying the criterion; and as y decreases, the x updates become more effective.

The third equivalent criterion follows by noticing $g \equiv r^{-1}$.

(B) Expected-time *upper* bounds. Consider the two blocks that the randomized scheme stochastically *majorizes*:

(i) *α-only updates (via r).* Keep y fixed at y_0 and increase x by at least $\delta_\alpha^{\min}$ on each α-update until $x \geq r(y_0)$, which takes $n_\alpha = \lceil G_0^r/\delta_\alpha^{\min} \rceil$ α-updates. Since an α-update arrives with probability φ each iteration, the expected number of iterations to accumulate n_α successes is n_α/φ. The actual algorithm, which also benefits from occasional β-updates (because r is increasing), cannot be slower in expectation.

(ii) *β-only updates (via g).* Keep x fixed at x_0 and decrease y by at least $\delta_\beta^{\min}$ on each β-update until $y \leq g(x_0)$, which takes $n_\beta = \lceil G_0^g/\delta_\beta^{\min} \rceil$ β-updates. Since a β-update arrives with probability $1-\varphi$, the expected number of iterations is $n_\beta/(1-\varphi)$. Again, the actual algorithm, which also benefits from occasional α-updates (because g is increasing), cannot be slower in expectation.

Taking the minimum of these two guarantees gives the stated upper bound.

(C) Forward invariance of the collusive region. Define the slack $s_t \coloneqq x_t - r(y_t)$. If $(x_t, y_t) \in \mathcal{C}$, then $s_t > 0$, and

1. On an α-update, $x_{t+1} = x_t + \Delta_\alpha(t)$ with $\Delta_\alpha(t) \geq \delta_\alpha^{\min} > 0$ and y unchanged, hence
$$s_{t+1} = x_t + \Delta_\alpha(t) - r(y_t) = s_t + \Delta_\alpha(t) > s_t > 0\,.$$
2. On a β-update, $y_{t+1} = y_t - \Delta_\beta(t)$ with $\Delta_\beta(t) \geq \delta_\beta^{\min} > 0$ and x unchanged; since r is strictly increasing, $r(y_{t+1}) < r(y_t)$, hence
$$s_{t+1} = x_t - r(y_{t+1}) > x_t - r(y_t) = s_t > 0\,.$$

Thus in either case $s_{t+1} \geq s_t + \min\{\delta_\alpha^{\min}, r(y_t) - r(y_{t+1})\} > 0$, so $(x_{t+1}, y_{t+1}) \in \mathcal{C}$.

(D) Expected-time *lower* bounds. Consider the "slacks":

$$s_t^r \coloneqq x_t - r(y_t) \quad \text{and} \quad s_t^g \coloneqq g(x_t) - y_t,$$

and note $s_t^r > 0 \iff s_t^g > 0$ by (A). Because r is increasing and convex on $[0,1)$ and g is increasing and concave on $[0,\infty)$, the *per-step* gains from a β-update in s_t^r and from an α-update in s_t^g are *largest at the start* and then monotonically decrease as the run progresses:

$$r(y_t) - r(y_t - \Delta_\beta(t)) \leq r(y_0) - r(y_0 - \delta_\beta^{\max}) \eqqcolon \Delta r_0^{\max},$$
$$g(x_t + \Delta_\alpha(t)) - g(x_t) \leq g(x_0 + \delta_\alpha^{\max}) - g(x_0) \eqqcolon \Delta g_0^{\max}.$$

Hence, the expected one-step improvement in s_t^r is at most $\varphi\delta_\alpha^{\max} + (1-\varphi)\Delta r_0^{\max}$, and similarly, the expected one-step improvement in s_t^g is at most $\varphi\Delta g_0^{\max} + (1-\varphi)\delta_\beta^{\max}$. Starting from initial gaps $G_0^r = \max\{0, r(y_0) - x_0\}$ and $G_0^g = \max\{0, y_0 - g(x_0)\}$, standard averaging argument yields the stated lower bound on $\mathbb{E}[\tau]$.

$\square$