Received: 27 May 2022 | Accepted: 9 May 2023

DOI: 10.1111/mafi.12401

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

SPECIAL ISSUE ARTICLE

Dynamics of market making algorithms in dealer markets: Learning and tacit collusion

Rama Cont | **Wei Xiong**

Mathematical Institute, University of Oxford, Oxford, UK

Correspondence
Wei Xiong, Mathematical Institute, University of Oxford, Oxford, UK.
Email: wei.xiong@maths.ox.ac.uk

[Correction added on 31st August 2023, after first online publication: This article was incorrectly published as Original Article and has been corrected to Special Issue Article in this version.]

Funding information
EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation, Grant/Award Number: EP/S023925/1

Abstract

The widespread use of market-making algorithms in electronic over-the-counter markets may give rise to unexpected effects resulting from the autonomous learning dynamics of these algorithms. In particular the possibility of "tacit collusion" among market makers has increasingly received regulatory scrutiny. We model the interaction of market makers in a dealer market as a stochastic differential game of intensity control with partial information and study the resulting dynamics of bid-ask spreads. Competition among dealers is modeled as a Nash equilibrium, while collusion is described in terms of Pareto optima. Using a decentralized multi-agent deep reinforcement learning algorithm to model how competing market makers learn to adjust their quotes, we show that the interaction of market making algorithms via market prices, without any sharing of information, may give rise to tacit collusion, with spread levels strictly above the competitive equilibrium level.

KEYWORDS
differential games, decentralized learning, intensity control, learning, Market microstructure, market making, multi-agent actor-critic algorithm, Nash equilibrium, reinforcement tacit collusion

1 | INTRODUCTION

The widespread use of algorithmic trading and market-marking in financial markets has led to a market landscape dominated by autonomous algorithms capable of learning from market data. While this evolution had undoubtedly increased the operational efficiency of markets in many ways, it has also given rise to new sources of risk and unexpected consequences, which have focused the attention of market participants and regulators. An important question is thus to understand the market dynamics resulting from interactions of such market-making algorithms. In particular, there have been regulatory concerns whether the interactions of learning algorithms could result in undesirable outcomes, even though the algorithms are not intended to do so by design.

In an insightful study, Calvano et al. (2020) have shown, in the setting of repeated auctions in a goods market, that competition among pricing algorithms with learning may lead to prices set at levels different from competitive benchmarks, without any explicit exchange of information between market makers, a situation known as "tacit collusion".

Tacit collusion is a well-known issue in oligopolistic markets Ivaldi et al. (2003) and Tirole (1988), and may emerge in repeated auctions without explicit information sharing (Han, 2021; Skrzypacz & Hopenhayn, 2004). The possibility that tacit collusion may emerge from the interaction of autonomous algorithms learning from market data in a decentralized fashion, a situation sometimes referred to as "algorithmic collusion" (Assad et al., 2021; Han, 2021), has attracted the concerns of market regulators (Competition & Markets Authority, 2021). It is thus of interest to investigate the present work we investigate whether the tacit collusion exhibited by Calvano et al. may also arise in the setting of financial markets with market makers competing for two-sided (buy and sell) order flow. In a recent work (Xiong & Cont, 2021), we examined conditions under which tacit collusion may arise in a discrete-time dealer market with multiple market makers. In the present work we revisit these issues in more detail in the framework of a market with continuous-time trading with competing market makers who learn from market data, extending the results of Calvano et al. (2020) and Xiong and Cont (2021) to a continuous-time setting which better captures some important features of intraday trading in financial markets.

Contribution. We model the interaction of market makers in a dealer market as a stochastic differential game of intensity control with partial information and study the resulting dynamics of bid-ask spreads resulting from competition among market makers and their learning dynamics.

We first study two benchmark cases: a competitive market, modeled as a Nash equilibrium of the game, and collusion among dealers, modeled as a Pareto optimum of the game. We give conditions for the existence of a Nash equilibrium, which we characterize in terms of a system of coupled Hamilton–Jacobi equations, and exhibit an algorithm based on fictitious play for computing Nash equilibria.

These benchmark cases correspond to the hypothetical situations where the dynamics of order flow is known to market makers. In practice, market makers interact with client order flow and learn to adjust their quotes in order to maximize their profit. We model this learning process using a decentralized multi-agent deep reinforcement learning algorithm (Hambly et al., 2023) using a policy gradient method (Fazel et al., 2018) to update market makers strategies, parameterized via neural networks. Our simulation results show that the interaction of market making algorithms through market prices, without any sharing of information, may give rise to tacit collusion, as evidenced by quoted spread levels significantly higher than in competitive (Nash) equilibrium.

This emergence of "tacit collusion" through learning has interesting implications for market design and market regulation, which call to be explored. Our model, coupled with the use of

multi-agent deep reinforcement learning, provides a conceptual framework for studying "tacit collusion" showing that the latter is a useful tool for exploring these issues.

Related literature. Our modeling framework builds on the recent literature on continuous-time models for optimal market making in dealer markets: following pioneering work of Ho and Stoll (1980, 1983) and (Avellaneda & Stoikov, 2008), the probem of optimal market making has been formulated as a stochastic control problem where market makers quote ask/bid prices dynamically to maximize their expected profit adjusted for inventory risk over a finite or infinite time horizon (Avellaneda & Stoikov, 2008; Barzykin et al., 2023 ; Bergault & Guéant, 2021; Cartea et al., 2014; Guéant, 2017; Guéant & Manziuk, 2019). In contrast to the earlier literature in which market makers are assumed to know the market dynamics, the recent literature has explored the more realistic case where market makers learn through trial and error, using reinforcement learning (Guéant & Manziuk, 2019, 2020; Bergault & Guéant, 2021; Barzykin et al., 2023). In most of these models, the market maker faces a random environment represented by an order flow represented as a point process, so a natural mathematical modeling framework for the problem is that of *intensity control* of point processes (Bremaud, 1981). Competition of market makers may be modeled in this setting as a stochastic differential game (Cont et al., 2021; Guo & Xu, 2019; Luo & Zheng, 2021). Cont et al. (2021) model inter-bank lending as a stochastic differential game of singular control and study Pareto optimal strategies. Competition among market makers is studied by Luo and Zheng (2021).

The emergence of tacit collusion from learning has also been studied in various contexts. Waltman and Kaymak (2008) show how competing producers using a Q-learning algorithm learn to raise prices above Nash equilibrium price by reducing production in a Cournot competition model. Calvano et al. (2020) show that Q-learning in a Bertrand competition model may result in prices strictly above competitive levels associated with Nash equilibrium. Abada and Lambin (2023) apply multi-agent Q-learning to Cournot competition in electricity markets, and conclude that the collusion may result from imperfect exploration. Asker et al. (2021) compare pricing outcomes from asynchronous versus synchronous learning algorithms. Hettich (2021) shows that deep Q-learning (DQN) algorithms lead to collusive strategies significantly faster. Han (2021) investigates the effects of experience replay in learning algorithm, which leads to prices closer to equilibrium level.

This literature primarily focuses on a one-sided goods market where "producers" fix prices, and may not directly be applicable to a financial market with two-sided order flow. Xiong and Cont (2021) study the emergence of tacit collusion among market makers via deep reinforcement learning in a discrete-time repeated game with competition between multiple market makers. More recently, Álvaro Cartea et al. (2022) study the impact of discreteness of "tick size" on algorithmic collusion. Álvaro Cartea et al. (2022) study tacit collusion in a model with competing liquidity providers whose behavior is modeled using a multi-armed bandit algorithm. Ganesh et al. (2019) and Ardon et al. (2021) build up multi-agent dealer market simulators using reinforcement learning agents and show that reinforcement learning agents replicate some "stylized facts" of dealer market.

Outline. Section 2 describes our model setting for a continuous-time dealer market with competition among market makers, formulated as a stochastic differential game of intensity control. Section 3 describes competition among dealers in terms of a Nash equilibrium. Section 4 describes collusion among dealers and establishes the connection between collusion and Pareto optima. In Section 5 we describe a fictitious play algorithm for numerically computing Nash equilibria. In Section 6 we describe how learning dynamics of

market makers may be modeled using decentralized multi-agent deep reinforcement learning and its implementation using a Decentralized Multi-agent Deep Deterministic Policy Gradient (Decentralized MADDPG) algorithm and presents simulation evidence for tacit collusion.

2 | A CONTINUOUS-TIME DEALER MARKET WITH MULTIPLE MARKET MAKERS

We propose a continuous-time dynamic model of a dealer market with competing market-makers, formulated as a stochastic differential game of *intensity control* Bremaud (1981) with partial information. In the following, we will be interchangeably using the terms "market maker", "dealer" and "agent".

We consider a market with a single asset and N market makers. The market price of the asset is modeled by a Brownian motion

$$S_t = S_0 + \sigma W_t \tag{1}$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying usual conditions, representing the flow of market information.

Order flow arrives in the form of Request for Quotes (RFQ). Clients send RFQs to all market makers, who respond by proposing bid/ask quotes around the market price. Market maker i quotes an ask price $S_t^{a,i}$ and a bid price $S_t^{b,i}$ where

$$S_t^{a,i} = S_t - \delta_t^{a,i} \qquad S_t^{b,i} = S_t + \delta_t^{b,i}.$$

We refer to $\delta_t^{a,i}$ and $\delta_t^{b,i}$ as the centered ask and bid quotes.

We model the buy/sell order flows as *càdlàg* marked point processes $N^a(dt)$ and $N^b(dt)$ with constant arrival intensity λ^a and λ^b and constant order size Δ. [1]

The market makers' quotes depend on her inventory, which we denote by q_t^i. As orders arrive in multiples of Δ, the inventory q_t^i takes discrete values, multiples of Δ and is typically subject to limits. We thus assume that the inventory of market maker i to take values in $\mathcal{Q}_i = \{-H_i, -H_i + \Delta, \dots, H_i - \Delta, H_i\}$. Hence there are $1 + 2H_i/\Delta$ possible values for q_t^i. Note that we impose inventory limits not only for considering practical market making scenarios, but also for mathematical reason of boundedness (Guéant, 2017; Guéant et al., 2013). The inventory limits are essential for proving boundedness of objective function in Proposition 3.2, which will be used in proof of Theorem 3.6. It would be mathematically challenging to obtain existence of Nash equilibrium if inventory were unbounded.

The *quoting strategy* for market maker i may then be represented as a map $\delta^i : \mathcal{Q}_i \mapsto \mathbb{R}$ which we can represent as a vector $\boldsymbol{\delta}^i = (\vec{\delta}^{a,i}, \vec{\delta}^{b,i})$ with components indexed by inventory levels in $\mathcal{Q}_i$:

$$\vec{\delta}^{a,i} = (\delta_{q_i}^{a,i})_{q_i\in\mathcal{Q}_i}, \quad \vec{\delta}^{b,i} = (\delta_{q_i}^{a,i})_{q_i\in\mathcal{Q}_i}.$$

We use bold symbols to represent collection of vectors, where each vector is a quoting strategy for one market maker.

- The centered ask and bid quotes $\vec{\delta}^{a,i}, \vec{\delta}^{b,i}$ by market maker i are vectors in $\mathbb{R}^{2\frac{H_i}{\Delta}+1}$ with each coordinate corresponding to quote at a specific inventory level $q_i \in \{-H_i, -H_i+\Delta, \dots, H_i-\Delta, H_i\}$. For convenience we index the coordinates of these vectors by inventory levels directly. For example $\delta^{a,i}_{q_i}, \delta^{b,i}_{q_i}$ are, respectively, $(\frac{H_i+q_i}{\Delta}+1)-th$ coordinates of $\vec{\delta}^{a,i}, \vec{\delta}^{b,i}$, where they represent centered ask/bid quotes by market maker i when her inventory level is q_i.
- $\boldsymbol{\delta}^i = (\vec{\delta}^{a,i}, \vec{\delta}^{b,i})$ denote the quoting strategy of market maker i. Hence $\boldsymbol{\delta}^i \in \mathbb{R}^{2\frac{H_i}{\Delta}+1} \times \mathbb{R}^{2\frac{H_i}{\Delta}+1}$.
- $\boldsymbol{\delta}^{a,-i}, \boldsymbol{\delta}^{b,-i}$ are collections of ask and bid quoting strategies of all market makers *excluding* i. They include $N-1$ vectors each being $\mathbb{R}^{2\frac{H_j}{\Delta}+1}$-valued corresponding to the quoting strategy of market maker $j \neq i$. These vectors are sorted by market maker's index. Namely $\boldsymbol{\delta}^{a,-i} = (\vec{\delta}^{a,j})_{j=1,\dots,i-1,i+1,\dots,N} \in \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, for $\boldsymbol{\delta}^{b,-i}$ vice versa. We write $\boldsymbol{\delta}^{-i} = (\boldsymbol{\delta}^{a,-i}, \boldsymbol{\delta}^{b,-i})$ to denote collections of both ask and bid quoting strategies of market maker i's competitors. For simplicity of notations we use $\vec{\delta}^{-i}$ to denote either $\boldsymbol{\delta}^{a,-i}$ or $\boldsymbol{\delta}^{b,-i}$ as variables in functions f^i_a and f^i_b below.

The probability that the RFQ gets executed against a market maker depends on their quotes and those of other competing market makers. We model this probability through a pair of functions f^i_a and f^i_b representing the dependence of execution probabilities for market maker i on market makers quotes: [2]

$$f^i_a(\delta, \vec{\delta}^{-i}) : \mathbb{R} \times \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \mathbb{R}^+, \qquad f^i_b(\delta, \vec{\delta}^{-i}) : \mathbb{R} \times \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \mathbb{R}^+$$

satisfying

$$\sum_{i=1}^N f^i_a(\delta, \vec{\delta}^{-i}) \leq 1, \qquad \sum_{i=1}^N f^i_b(\delta, \vec{\delta}^{-i}) \leq 1.$$

The inequality corresponds to the possibility that an RFQ is not executed by any market maker, due to cancelation, with probability $(1 - \sum_{i=1}^N f^i_a)$ (resp. $(1 - \sum_{i=1}^N f^i_b)$). We will specify assumptions on execution probabilities in more detail below.

Market maker i only quotes prices when her inventory does not exceed the limits $\pm H_i$.The (ask/bid) order flow executed by market maker i may then be represented as pair of point process $N^{a,i}(dt)$ and $N^{b,i}(dt)$ with intensity

$$\nu^{a,i}_t = \lambda^a f^i_a(\delta^{a,i}_{q^i_t}, \boldsymbol{\delta}^{a,-i}) \mathbb{I}(q^i_t > -H_i)\ \nu^{b,i}_t = \lambda^b f^i_b(\delta^{b,i}_{q^i_t}, \boldsymbol{\delta}^{b,-i}) \mathbb{I}(q^i_t < H_i) \tag{2}$$

We consider stationary feedback quoting strategies $\vec{\delta}^{a,i}, \vec{\delta}^{b,i}$: $\vec{\delta}^{a,i}, \vec{\delta}^{b,i}$ represented as $\mathbb{R}^{2\frac{H_i}{\Delta}+1}$-valued vectors, in which each coordinate $\delta^{a,i}_{q_i}$ corresponds to the centered quote at inventory level $q_i \in \mathcal{Q}_i$. Thus at time t market maker i will quote

$$\delta^{a,i}_t = \delta^{a,i}_{q^i_t}, \quad \delta^{b,i}_t = \delta^{b,i}_{q^i_t}$$

The rare circumstance where the quotes are negative is called a "crossed market". In this case $\delta^{a,i}$ and $\delta^{b,i}$ could be negative, and ask price is lower than bid price. We assume that the centered quotes under crossed market is constrained by a limit, that is $\delta^{a,i} \geq -\delta_\infty, \delta^{b,i} \geq -\delta_\infty$ where $\delta_\infty > 0$ is a given constant.[3] Therefore, the admissible strategy space of each market maker i is

$$\mathcal{A}_i = \left\{ \boldsymbol{\delta}_t^i = (\delta^{a,i}_{q_t^i}, \delta^{b,i}_{q_t^i}) \middle| \delta^{a,i}, \delta^{b,i} \in \mathbb{R}^{2\frac{H_i}{\Delta}+1}; \delta^{a,i}_{q_i} \geq -\delta_\infty, \delta^{b,i}_{q_i} \geq -\delta_\infty, \forall q_i \in \{-H_i, \dots, H_i\} \right\} \tag{3}$$

We let I_δ denote the interval $[-\delta_\infty, \infty)$. $\mathcal{A}_i$ contains stochastic processes $\boldsymbol{\delta}_t^i = (\delta^{a,i}_{q_t^i}, \delta^{b,i}_{q_t^i})$ which are feedback strategies depending on the inventory q_t^i, while $(I_\delta)^{2\frac{H_i}{\Delta}+1} \times (I_\delta)^{2\frac{H_i}{\Delta}+1}$ is the space of possible values of a market makers' quoting strategies.

We use the following notations for partial derivatives: for $i \in \{1, \dots, N\}$ and $j \neq i$ we denote

$$\begin{aligned}
\partial_1 f_a^i &= \frac{\partial f_a^i}{\partial \delta}(\delta, \vec{\delta}^{-i}), \quad & \partial_1 f_b^i &= \frac{\partial f_b^i}{\partial \delta}(\delta, \vec{\delta}^{-i}) \\
\partial^2_{11} f_a^i &= \frac{\partial^2 f_a^i}{\partial \delta^2}(\delta, \vec{\delta}^{-i}), \quad & \partial^2_{11} f_b^i &= \frac{\partial^2 f_b^i}{\partial \delta^2}(\delta, \vec{\delta}^{-i}) \\
\partial_{j,q_j} f_a^i &= \frac{\partial f_a^i}{\partial \delta^j_{q_j}}(\delta, \vec{\delta}^{-i}), \quad & \partial_{j,q_j} f_b^i &= \frac{\partial f_b^i}{\partial \delta^j_{q_j}}(\delta, \vec{\delta}^{-i}) \\
\partial_{j,q_j}\partial_1 f_a^i &= \frac{\partial^2 f_a^i}{\partial \delta^j_{q_j}\partial \delta}(\delta, \vec{\delta}^{-i}), \quad & \partial_{j,q_j}\partial_1 f_b^i &= \frac{\partial^2 f_b^i}{\partial \delta^j_{q_j}\partial \delta}(\delta, \vec{\delta}^{-i})
\end{aligned} \tag{4}$$

Note that the symbol ∂_{j,q_j} represents first-order derivative with respect to coordinate $\delta^j_{q_j}$ in $\vec{\delta}^{-i}$.

We make the following assumptions on f_a^i and f_b^i.

Assumption 2.1. f_a^i, f_b^i are twice continuously differentiable on $\mathbb{R} \times \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ and satisfy

$$f_a^i(\delta, \vec{\delta}^{-i}) > 0, f_b^i(\delta, \vec{\delta}^{-i}) > 0, \quad \sum_{i=1}^N f_a^i(\delta, \vec{\delta}^{-i}) \leq 1, \quad \sum_{i=1}^N f_b^i(\delta, \vec{\delta}^{-i}) \leq 1$$

There exists a function $\Lambda(\delta) \in C^2(\mathbb{R})$, such that for $m \in \{a, b\}$, $0 < f_m^i(\delta, \vec{\delta}^{-i}) < \Lambda(\delta)$, and

$$\lim_{\delta\to\infty} \Lambda(\delta)\delta = 0, \Lambda'(\delta) < 0, \Lambda(\delta)\Lambda''(\delta) \leq 2(\Lambda'(\delta))^2$$

Remark 2.2. $\Lambda(\delta)$ in Assumption 2.1 can be understood as the execution probability in the single market maker case as in Guéant (2017). The assumption corresponds to the intuition that competition lowers the execution rate for each market maker.

Assumption 2.3. The execution probabilities $\{f^i_m\}_{m\in\{a,b\}}$ satisfy: $\forall(\delta,\vec{\delta}^{-i})\in\mathbb{R}\times\prod_{j\neq i}\mathbb{R}^{2\frac{H_j}{\Delta}+1}, \forall j\neq i, \forall q_j\in\mathcal{Q}_j$,

$$\begin{aligned}
&\partial_1 f^i_m<0, \partial_{j,q_j} f^i_m\geq 0, \qquad \frac{\partial^2_{11} f^i_m\cdot f^i_m}{(\partial_1 f^i_m)^2}<2\\
&2(\partial_1 f^i_m)^2-\partial^2_{11} f^i_m\cdot f^i_m-\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\left|\partial_1 f^i_m\cdot\partial_{j,q_j} f^i_m-f^i_m\cdot\partial_{j,q_j}\partial_1 f^i_m\right|>0\\
&\lim_{\delta\to+\infty}\frac{f^i_m(\delta,\vec{\delta}^{-i})}{\partial_1 f^i_m(\delta,\vec{\delta}^{-i})}<\infty, \qquad \forall\vec{\delta}^{-i}\in\prod_{j\neq i}\mathbb{R}^{2\frac{H_j}{\Delta}+1}
\end{aligned}\tag{5}$$

Remark 2.4. Assumptions 2.1 and 2.3 are needed for proving existence of Nash equilibrium in Section 3. In Assumption 2.3 $\partial_1 f^i_m<0, \partial_{j,q_j} f^i_m\geq 0$ corresponds to monotonicity of execution rate as functions of market maker's and competitors' quotes. Similar to execution probabilities in Guéant (2017) we need the function $\delta\to\delta\cdot f^i_m(\delta,\vec{\delta}^{-i})$ to reach a unique maximum on $[-\delta_\infty,\infty)$ hence we assume $\frac{\partial^2_{11} f^i_m\cdot f^i_m}{(\partial_1 f^i_m)^2}<2$. The last two lines in (5) are the strongest conditions specifying the growth regularity of execution rate functions. These conditions are motivated by Luo and Zheng (2021) but generalized to N-player circumstance. This assumption is used in Proposition A.11 to prove implicit function property of centered quotes $\boldsymbol{\delta}$ as function of value vector $\boldsymbol{p}$. Intuitively, the second line of (5) specifies the first variable δ, which is the market maker's own quote, dominates the growth of execution rate $f^i_m(\delta,\vec{\delta}^{-i})$ compared to competitors' quotes. the last line in (5) specifies that $\partial_1 f^i_m(\delta,\vec{\delta}^{-i})$ does not vanish too fast as function of δ. An example is given in Remark 2.5 satisfying these assumptions.

Remark 2.5. Examples of functions satisfying Assumptions 2.1 and 2.3 are

$$f^i_a(\delta,\vec{\delta}^{-i})=\frac{1}{N}\frac{1}{1+e^{a\delta+b}}\frac{e^{\frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}(a^j_{q_j}\delta^j_{q_j}+b^j_{q_j})}}{1+e^{a\delta+\frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}(a^j_{q_j}\delta^j_{q_j}+b^j_{q_j})}}\tag{6}$$

where $K_i=\sum_{j\neq i, j\in\{1,\dots,N\}}(2\frac{H_i}{\Delta}+1)$ is the number of inventory levels of market maker i's competitors, and $a, b, a^j_{q_j}, a^j_{q_j}$ are parameters.

Market maker i's inventory q^i_t evolves according to

$$dq^i_t=\Delta(N^{b,i}(dt)-N^{a,i}(dt))\tag{7}$$

and her cash holdings X^i_t evolve according to

$$\begin{aligned}
dX^i_t&=\Delta(S_t+\delta^{a,i}_t)N^{a,i}(dt)-\Delta(S_t-\delta^{b,i}_t)N^{b,i}(dt)\\
&=\Delta\left(\delta^{a,i}_t N^{a,i}(dt)+\delta^{b,i}_t N^{b,i}(dt)\right)-S_t dq^i_t
\end{aligned}\tag{8}$$

The objective of each market maker is to maximize expected discounted profit. We consider the problem over an infinite horizon in order to study stationary feedback quoting strategies, which are more readily accessible to reinforcement learning. Denoting $\boldsymbol{\delta}^i = (\vec{\delta}^{a,i}, \vec{\delta}^{b,i}), \boldsymbol{\delta}^{-i} = (\boldsymbol{\delta}^{a,-i}, \boldsymbol{\delta}^{b,-i})$, market maker i has objective function

$$\tilde{J}_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i, x_0^i, s) = x_0^i + q_0^i s + \mathbb{E}\left[\int_0^\infty e^{-rt} d(X_t^i + q_t^i S_t) - \int_0^\infty e^{-rt}\psi_i(q_t^i)dt\right] \tag{9}$$

where $\psi_i : \mathbb{R} \to \mathbb{R}_+$ is the running cost for holding inventory.

Lemma 2.6. *The objective function in (9) can be written as:*

$$\tilde{J}_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i, x_0^i, s) = x_0^i + q_0^i s + J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i) \tag{10}$$

where

$$\begin{aligned} J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i) = \mathbb{E}^{q_0^i}\Bigg[\int_0^\infty e^{-rt}\Big(\lambda^a \Delta\delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})\mathbb{I}(q_t^i > -H_i) + \lambda^b \Delta\delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}_t^{b,-i})\mathbb{I}(q_t^i < H_i)\Big)dt \\ - \int_0^\infty e^{-rt}\psi_i(q_t^i)dt\Bigg] \end{aligned} \tag{11}$$

Proof. Since q_t^i takes values from finite set $\mathcal{Q}_i$, $\psi_i(q_t^i)$ is uniformly bounded for $t \geq 0$ and the expectation for the running cost term $\mathbb{E}[\int_0^\infty e^{-rt}\psi_i(q_t^i)dt] < \infty$ is well defined.

From Assumption 2.1 we have $f_a^i(\delta, \cdot) \leq \Lambda(\delta), f_b^i(\delta, \cdot) \leq \Lambda(\delta), \forall \delta \in \mathbb{R}$, where $\Lambda(\delta)$ is a C^2 function. Given quoting strategies $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$, since $\vec{\delta}^{a,i} = (\delta_q^{a,i})_{q\in\mathcal{Q}_i}, \vec{\delta}^{b,i} = (\delta_q^{b,i})_{q\in\mathcal{Q}_i}$ are vectors in space $\mathbb{R}^{2\frac{H_i}{\Delta}+1}$, define $A := \sup_{q\in\mathcal{Q}_i} |\delta_q^{a,i}|, B := \sup_{q\in\mathcal{Q}_i} |\delta_q^{b,i}|$ and denote q_t^i market maker i's inventory process under quoting strategies $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$. Recall that $\delta_t^{a,i} = \delta_{q_t^i}^{a,i}, \delta_t^{b,i} = \delta_{q_t^i}^{b,i}$, we also write $\boldsymbol{\delta}_t^{a,-i}, \boldsymbol{\delta}_t^{b,-i}$ to denote the quoting strategy process of other maker makers as functions of their corresponding inventory levels. We obtain

$$\begin{aligned} \int_0^\infty e^{-rt}\Big(\lambda^a \delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})\mathbb{I}(q_t^i > -H_i)\Big)dt \leq A\Lambda(-\delta_\infty)\int_0^\infty e^{-rt}dt < \infty \\ \int_0^\infty e^{-rt}\Big(\lambda^b \delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}_t^{b,-i})\mathbb{I}(q_t^i < H_i)\Big)dt \leq B\Lambda(-\delta_\infty)\int_0^\infty e^{-rt}dt < \infty \end{aligned} \tag{12}$$

Therefore we have

$$\begin{aligned} \mathbb{E}\left[\int_0^\infty e^{-rt}\delta_t^{a,i}N^{a,i}(dt)\right] &= \mathbb{E}\left[\int_0^\infty e^{-rt}\Big(\lambda^a \delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})\mathbb{I}(q_t^i > -H_i)\Big)dt\right] \\ \mathbb{E}\left[\int_0^\infty e^{-rt}\delta_t^{b,i}N^{b,i}(dt)\right] &= \mathbb{E}\left[\int_0^\infty e^{-rt}\Big(\lambda^b \delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}_t^{b,-i})\mathbb{I}(q_t^i < H_i)\Big)dt\right] \end{aligned}$$

We can apply Itô's lemma to $X_t^i + q_t^i S_t$ in objective function, to rewrite (9) as:

$$\begin{aligned}\tilde{J}_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i, x_0^i, s) &= \mathbb{E}^{q_0^i}\Bigg[\int_0^\infty e^{-rt}\Big(\lambda^a \Delta \delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})\mathbb{I}(q_t^i > -H_i) + \lambda^b \Delta \delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}_t^{b,-i})\mathbb{I}(q_t^i < H_i)\Big)dt \\ &\quad - \int_0^\infty e^{-rt}\psi_i(q_t^i)dt)\Bigg] + x_0^i + q_0^i s \\ &= x_0^i + q_0^i s + J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_0^i)\end{aligned} \tag{13}$$

□

The expectation in (13) is taken with respect to the law of the process q^i whose evolution is given by (7), with initial condition q_0^i. For simplicity of notations, we will use $\mathbb{E}$ instead of $\mathbb{E}^{q_0^i}$.

Remark 2.7. In the case where (7) admits a stationary solution (which will be typically the case in many realistic situations), this term does not depend on q_0^i and we will denote this situation by $J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$

Remark 2.8. The discount rate r in (11) is interpreted as interest rate to discount future rewards. Guéant and Manziuk (2020) prove asymptotic ergodic property of value function as $r \to 0$ for stochastic optimal control on graphs. See also Guéant and Manziuk (2019) in which authors set small r to approximate the ergodic constant instead of adhering to interest rate value.

3 | COMPETITION AMONG MARKET MAKERS: NASH EQUILIBRIUM

A situation of competition among market makers may be modeled through the concept of Nash equilibrium. In this section we discuss the existence of Nash equilibrium and its characterization in terms of a Hamilton–Jacobi–Bellman system in the setting of our model.

We first show the existence of Nash equilibrium under Assumptions 2.1 and 2.3. We then characterize Nash equilibrium quoting strategies in terms of an HJB equation (28) in Proposition 3.4.

There are N competing market makers whose quotes jointly affect the execution of market order flow.

Definition 3.1 (Nash equilibrium). A Nash equilibrium for system (11) is a tuple of quoting strategies $\vec{\boldsymbol{\delta}}^* = ((\boldsymbol{\delta}^1)^*, \dots, (\boldsymbol{\delta}^N)^*) \in \prod_{j=1}^N ((I_\delta)^{2\frac{H_j}{\Delta}+1} \times (I_\delta)^{2\frac{H_j}{\Delta}+1})$, such that for any $\boldsymbol{q}_0 \in \mathbb{R}^N$ and any $i \in \{1, \dots, N\}$,

$$J_i((\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*; q_0^i) \geq J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*; q_0^i), \qquad \forall \boldsymbol{\delta}^i \in (I_\delta)^{2\frac{H_i}{\Delta}+1} \times (I_\delta)^{2\frac{H_i}{\Delta}+1} \tag{14}$$

We say that $\vec{\boldsymbol{\delta}}^*$ is a stationary Nash equilibrium if under $\vec{\boldsymbol{\delta}}^*$ the inventories (7) admit a stationary solution (see Remark 2.7).

We denote

$$V_i(q_i) = J_i((\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*; q_i) \tag{15}$$

the value function of player i under the Nash equilibrium quoting strategy, with initial condition $q_0^i = q_i$.

We first state a proposition on uniformly boundedness of objective function $J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)$. This result will be useful for proving existence of Nash equilibrium.

Proposition 3.2. *Under Assumption 2.1 and 2.3, there exists a constant $J_{max} > 0$ such that for any strategy $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$*

$$\forall i \in \{1, \dots, N\}, \forall q_i \in \mathcal{Q}_i, \qquad |J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)| \le J_{max} \tag{16}$$

Proof. From the definition of the objective function, we have

$$\begin{aligned} |J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)| \le \mathbb{E}[&\left| \int_0^\infty e^{-rt} \Big(\lambda^a \Delta \delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i}) \mathbb{I}(q_t^i > -H_i) \Big) dt \right| \\ &+ \left| \int_0^\infty e^{-rt} \Big(\lambda^b \Delta \delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}_t^{b,-i}) \mathbb{I}(q_t^i < H_i) \Big) dt \right| + \left| \int_0^\infty e^{-rt} \psi_i(q_i) dt \right|] \end{aligned} \tag{17}$$

Since q_t^i takes values from a finite set $\mathcal{Q}_i = \{-H_i, -H_i + \Delta, \dots, H_i - \Delta, H_i\}$, $\psi_i(q_i)$ is uniformly bounded by $\Psi_i := \max_{q \in \mathcal{Q}_i} \psi_i(q_i)$. Hence

$$\mathbb{E}[\left| \int_0^\infty e^{-rt} \psi_i(q_i) dt \right|] \le \Psi_i \int_0^\infty e^{-rt} dt = \frac{\Psi_i}{r}$$

From Assumption 2.1, $|\delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})| \le |\delta_t^{a,i} \Lambda(\delta_t^{a,i})|$, $\lim_{\delta \to \infty} \Lambda(\delta)\delta = 0$, and $\delta_t^{a,i} = \delta_{q_t^i}^{a,i}$ takes values on $[-\delta_\infty, \infty)$, then there exists a constant $B > 0$ such that $|\delta_t^{a,i} \Lambda(\delta_t^{a,i})| \le B$. In fact $\lim_{\delta \to \infty} \Lambda(\delta)\delta = 0$ implies there exists a positive number $M > 0$, such that $\forall \delta > M$, we have $|\Lambda(\delta)\delta| < 1$. Also $\Lambda(\delta)\delta$ is continuous function on $[-\delta_\infty, M]$, hence is bounded. We can then define $K := \max(\sup_{\delta \in [-\delta_\infty, M]} \Lambda(\delta)\delta, 1)$. Hence $|\delta_t^{a,i} f_a^i(\delta_t^{a,i}, \boldsymbol{\delta}_t^{a,-i})| \le K$ uniformly.

Similarly we have can prove for the bid side $|\delta_t^{b,i} f_b^i(\delta_t^{b,i}, \boldsymbol{\delta}^{b,-i})| \le K$. Therefore from (17) we obtain

$$|J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)| \le \frac{K(\lambda^a \Delta + \lambda^b \Delta) + \max_{i \in \{1, \dots, N\}} \Psi_i}{r} \tag{18}$$

The bound is uniform with respect to i, q_i and $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$. □

We shall denote the execution probabilities as $f_a^i(\delta_{q_i}^{a,i}, (\vec{\delta}^{a,j})_{j \ne i})$ and $f_b^i(\delta_{q_i}^{b,i}, (\vec{\delta}^{b,j})_{j \ne i})$ to emphasize that $\delta_{q_i}^{a,i}, \delta_{q_i}^{b,i}$ are the current ask and bid quotes by market maker i given her inventory level q_i. This notation will provide clarity in the optimality equations.

3.1 | Dynamic programming principle

The equilibrium value function (15) associated to Nash equilibrium satisfies a Dynamic Programming Principle:

Lemma 3.3. *(Dynamic Programming Principle) Let $q_i \in \mathcal{Q}_i$. Given any finite stopping time θ the value function V_i defined by (15)*

$$V_i(q_i) = \sup_{\delta^i \in \mathcal{A}_i} \mathbb{E}[\int_0^\theta e^{-rt}(\lambda^a \Delta\delta^{a,i}_{q_t^{i,q_i}} f_a^i(\delta^{a,i}_{q_t^{i,q_i}}, (\vec{\delta}^{a,j})^*_{j\neq i})\mathbb{I}(q_t^{i,q_i} > -H_i) + \lambda^b \Delta\delta^{b,i}_{q_t^{i,q_i}} f_b^i(\delta^{b,i}_{q_t^{i,q_i}}, (\vec{\delta}^{b,j})^*_{j\neq i})\mathbb{I}(q_t^{i,q_i} < H_i))dt + V_i(q_\theta^{i,q_i})] \tag{19}$$

where q_t^{i,q_i} is the inventory process of market maker i under joint quoting strategy $(\delta^i, (\delta^{-i})^*)$, with $q_0^{i,q_i} = q_i$.

Proof. Given the quoting strategies $(\delta^i, (\delta^{-i})^*)$, and a finite stopping time $\theta < \infty$,, from the Markovian property of q_t^{i,q_i}, we have

$$q_s^{i,q_i} = q_s^{i,q_\theta^{i,q_i}}, \forall s \geq \theta$$

From properties of conditional expectation and change of variable, we have

$$\begin{aligned}
J_i(\delta^i, (\delta^{-i})^*, q_i) &= \mathbb{E}\left[(\int_0^\theta + \int_\theta^\infty) e^{-rt} d(X_t^i + q_t^i S_t) - (\int_0^\theta + \int_\theta^\infty) e^{-rt}\psi_i(q_t^i)dt\right] \\
&= \mathbb{E}[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) \\
&\quad + \mathbb{E}_{q_\theta^{i,q_i}}\left(\int_\theta^\infty e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right)] \\
&= \mathbb{E}[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) \\
&\quad + \mathbb{E}_{q_\theta^{i,q_i}}\left(\int_0^\infty e^{-r(u+\theta)} d(X_{u+\theta}^i + q_{u+\theta}^{i,q_i} S_{u+\theta}) - e^{-r(u+\theta)}\psi_i(q_{u+\theta}^{i,q_i})du\right)] \\
&= \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} J_i(\delta^i, (\delta^{-i})^*; q_\theta^{i,q_i})\right]
\end{aligned} \tag{20}$$

By Definition 3.1 $V_i(q_i) = J_i((\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*, q_i) = \sup_{\boldsymbol{\delta}^i} J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*, q_i)$, we have

$$\begin{aligned} J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*, q_i) &\le \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i})\right] \\ &\le \sup_{\boldsymbol{\delta}^i} \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i}))\right] \end{aligned} \tag{21}$$

Taking supremum over $\boldsymbol{\delta}^i$ in left hand side of (21) we obtain:

$$V_i(q_i) \le \sup_{\boldsymbol{\delta}^i} \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i}))\right] \tag{22}$$

On the other hand, for joint quoting strategy $(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*)$ and given stopping time θ, from Definition 3.1 for any $\epsilon > 0, \omega \in \Omega$ there exists quoting strategy $\boldsymbol{\delta}^{i,\epsilon,\omega} = (\delta^{a,i,\epsilon,\omega}, \delta^{b,i,\epsilon,\omega}) \in (I_\delta)^{2\frac{H_i}{\Delta}+1} \times (I_\delta)^{2\frac{H_i}{\Delta}+1} \times (I_\delta)^{2\frac{H_i}{\Delta}+1} \times (I_\delta)^{2\frac{H_i}{\Delta}+1}$, such that $\boldsymbol{\delta}^{i,\epsilon,\omega}$ is an $\epsilon-$optimal control for value function $V_i(q^{i,q_i}_{\theta(\omega)}(\omega))$ starting at $q^{i,q_i}_{\theta(\omega)}(\omega)$:

$$V_i(q^{i,q_i}_{\theta(\omega)}(\omega)) - \epsilon \le J_i(\boldsymbol{\delta}^{i,\epsilon,\omega}, (\boldsymbol{\delta}^{-i})^*; q^{i,q_i}_{\theta(\omega)}(\omega))$$

Now define the following quoting strategy:

$$\hat{\boldsymbol{\delta}}^i{}_t(\omega) = \begin{cases} \boldsymbol{\delta}^i_{q_t^{i,q_i}(\omega)}(\omega) & t \in [0, \theta(\omega)] \\ \boldsymbol{\delta}^{i,\epsilon,\omega}_{q_t^{i,q_i}(\omega)}(\omega) & t \in [\theta(\omega), \infty] \end{cases} \tag{23}$$

Using a measurable selection argument Bertsekas and Shreve (1978), the process $\hat{\boldsymbol{\delta}}^i$ is an admissible strategy. Then by the law of iterated conditional expectation we have

$$\begin{aligned} V(q_i) \ge J_i(\hat{\boldsymbol{\delta}}^i, (\boldsymbol{\delta}^{-i})^*, q_i) &= \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} J_i(\delta^{i,\epsilon}, (\boldsymbol{\delta}^{-i})^*; q_\theta^{i,q_i})\right] \\ &\ge \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i})) - e^{-r\theta}\epsilon\right] \\ &\ge \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i}))\right] - \epsilon \end{aligned} \tag{24}$$

Since $\boldsymbol{\delta}^i$ θ and $\epsilon > 0$ taken arbitrarily we have

$$V_i(q_i) \geq \sup_{\boldsymbol{\delta}^i} \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i}))\right] \tag{25}$$

Combining (22) and (25) we have

$$V_i(q_i) = \sup_{\boldsymbol{\delta}^i} \mathbb{E}\left[\int_0^\theta \left(e^{-rt} d(X_t^i + q_t^{i,q_i} S_t) - e^{-rt}\psi_i(q_t^{i,q_i})dt\right) + e^{-r\theta} V_i(q_\theta^{i,q_i}))\right] \tag{26}$$

□

For arbitrary given quoting strategies of N market makers $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$, we define objective function associated with this quoting strategy, denoted by $V_i^\delta(q_i)$, where

$$V_i^\delta(q_i) = J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)$$

For consistency with reinforcement learning literature we name V^δ state-action value function. Note that if we denote V_i^δ by $V_i^{\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}}$ to emphasize the dependence on competitors' strategies $\boldsymbol{\delta}^{-i}$, then we have the relation $V_i(q_i) = \sup_{\boldsymbol{\delta}^i \in \mathcal{A}_i} V_i^{\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*}(q_i)$ where $\vec{\boldsymbol{\delta}}^*$ is a Nash equilibrium. The objective function V^δ associated with given quoting strategies satisfies following linear Bellman equation[4]: (Guéant & Manziuk, 2019)

$$\begin{aligned} &rV_i^\delta(q_i) + \psi_i(q_i) - \mathbb{I}(q_i > -H_i)\lambda^a \Delta f_a^i(\delta_{q_i}^{a,i}, (\vec{\delta}^{a,j})_{j\neq i})\left(\delta_{q_i}^{a,i} - \frac{V_i^\delta(q_i) - V_i^\delta(q_i - \Delta)}{\Delta}\right) \\ &- \mathbb{I}(q_i < H_i)\lambda^b \Delta f_b^i(\delta_{q_i}^{b,i}, (\vec{\delta}^{b,j})_{j\neq i})\left(\delta_{q_i}^{b,i} - \frac{V_i^\delta(q_i) - V_i^\delta(q_i + \Delta)}{\Delta}\right) = 0 \end{aligned} \tag{27}$$

The Dynamic Programming Principle thus leads to a system of Hamilton-Jacobi equations for the equilibrium value functions $V_i(q_i), i \in \{1, \ldots, N\}$:

$$\begin{aligned} &rV_i(q_i) + \psi_i(q_i) - \mathbb{I}(q_i > -H_i)\lambda^a \Delta \sup_{\delta_{q_i}^{a,i} \geq -\delta_\infty} \left[f_a^i\left(\delta_{q_i}^{a,i}, (\vec{\delta}^{a,j})_{j\neq i}\right)\left(\delta_{q_i}^{a,i} - \frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta}\right)\right] \\ &- \mathbb{I}(q_i < H_i)\lambda^b \Delta \sup_{\delta_{q_i}^{b,i} \geq -\delta_\infty} \left[f_b^i\left(\delta_{q_i}^{b,i}, (\vec{\delta}^{b,j})_{j\neq i}\right)\left(\delta_{q_i}^{b,i} - \frac{V_i(q_i) - V_i(q_i + \Delta)}{\Delta}\right)\right] = 0 \end{aligned} \tag{28}$$

where Nash equilibrium quoting strategy $\vec{\boldsymbol{\delta}}^*$ is such that the supremum in equation (28) is achieved simultaneously for $i \in \{1, \ldots, N\}$.

Proposition 3.4. *If there exists a Nash equilibrium quoting strategy*

$$\vec{\delta}^* = ((\delta^1)^*, \dots, (\delta^N)^*) \in \prod_{j=1}^{N} \left((I_\delta)^{2\frac{H_j}{\Delta}+1} \times (I_\delta)^{2\frac{H_j}{\Delta}+1} \right)$$

with corresponding equilibrium value functions $V_i(q_i), q_i \in \{-H_i, \dots, H_i\}$*, then the* V_i *satisfy the system of equations (28), where* $\forall q_i \in \{-H_i, \dots, H_i\}$

$$\begin{aligned}
(\delta^{a,i}_{q_i})^* &= \underset{\delta^{a,i}_{q_i} \geq -\delta_\infty}{\arg\max} \left[f^i_a(\delta^{a,i}_{q_i}, ((\vec{\delta}^{a,j})^*)_{j\neq i}) \left(\delta^{a,i}_{q_i} - \frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta} \right) \right] \\
(\delta^{b,i}_{q_i})^* &= \underset{\delta^{b,i}_{q_i} \geq -\delta_\infty}{\arg\max} \left[f^i_b(\delta^{b,i}_{q_i}, ((\vec{\delta}^{b,j})^*)_{j\neq i}) \left(\delta^{b,i}_{q_i} - \frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta} \right) \right]
\end{aligned} \tag{29}$$

Proof. The proof of Proposition 3.4 follows by a standard argument using Dynamic Programming Principle in Lemma 3.3. □

Next proposition is a verification theorem stating that solution to (28) is indeed a Nash equilibrium for N market maker system.

Proposition 3.5. *Under Assumptions 2.1 and 2.3, if there exists quoting strategies* $(\delta^{a,i})^*, (\delta^{b,i})^*, \forall i \in \{1, \dots, N\}$ *and functions* $V_i(q_i), q_i \in \{-H_i, \dots, H_i\}$ *such that* V_i *satisfy system of equations (28), and* $\forall q_i \in \{-H_i, \dots, H_i\}$

$$\begin{aligned}
(\delta^{a,i}_{q_i})^* &= \underset{\delta^{a,i}_{q_i} \geq -\delta_\infty}{\arg\max} \left[f^i_a(\delta^{a,i}_{q_i}, ((\vec{\delta}^{a,j})^*)_{j\neq i}) \left(\delta^{a,i}_{q_i} - \frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta} \right) \right] \\
(\delta^{b,i}_{q_i})^* &= \underset{\delta^{b,i}_{q_i} \geq -\delta_\infty}{\arg\max} \left[f^i_b(\delta^{b,i}_{q_i}, ((\vec{\delta}^{b,j})^*)_{j\neq i}) \left(\delta^{b,i}_{q_i} - \frac{V^\delta_i(q_i) - V^\delta_i(q_i - \Delta)}{\Delta} \right) \right]
\end{aligned} \tag{30}$$

then $(\delta^{a,i})^*, (\delta^{b,i})^*$ *define a Nash equilibrium (Def. 3.1) and* V_i *are the corresponding equilibrium value functions.*

Proof. To show that $(\boldsymbol{\delta}^i)^* = ((\delta^{a,i})^*, (\delta^{b,i})^*), i \in \{1, \dots, N\}$ are Nash equilibrium quoting strategies and $V_i(q_i)$ is the equilibrium value function, we need to verify $(\delta^{a,i})^*, (\delta^{b,i})^*$ and $V_i(q_i)$ satisfy Definition 3.1. Given any market maker i, we denote the joint quoting strategies of her competitors by $(\boldsymbol{\delta}^{-i})^*$. Assuming market maker i takes another quoting strategy $\boldsymbol{\delta}^i = (\delta^{a,i}, \delta^{b,i}) \in \mathbb{R}^{2\frac{H_i}{\Delta}+1} \times \mathbb{R}^{2\frac{H_i}{\Delta}+1}$ while her competitors still keep the joint quoting strategies $(\boldsymbol{\delta}^{-i})^*$, we need to show that

$$V_i(q_i) \geq J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*, q_i), \forall q_i \in \{-H_i, \dots, H_i\}$$

Let $(X^{i,\delta}_t)_{t\geq 0}$ denote the cash process and $(q^{i,\delta}_t)_{t\geq 0}$ denote the inventory process of market maker i with joint strategies $(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*)$, where $q^{i,\delta}_0 = q_i$. Since the process $q^{i,\delta}_t$ takes values from a finite

set $\{-H_i, -H_i+\Delta, \dots, H_i-\Delta, H_i\}$, the value function $V_i(q_t^{i,\delta})$ is uniformly bounded. Denote this uniform bound by $M(M>0)$. We can then apply Itô's formula on function $e^{-rt}V_i(q_t^{i,\delta})$.

For $T>0$,

$$\begin{aligned} e^{-rT}V_i(q_T^{i,\delta}) = V_i(q_i) &- \int_0^T re^{-rt}V_i(q_t^{i,\delta})dt \\ &+ \int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}-\Delta) - V_i(q_t^{i,\delta})\Big]N^{a,i}(dt) \\ &+ \int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}+\Delta) - V_i(q_t^{i,\delta})\Big]N^{b,i}(dt) \end{aligned} \tag{31}$$

From Assumption 2.1, $f_a^i(\delta,\delta^{-i}) < \Lambda(\delta)$, $f_b^i(\delta,\delta^{-i}) < \Lambda(\delta)$, and $\Lambda(\delta)$ is monotonically decreasing on $\mathbb{R}$. Moreover, centered quotes are bounded from below by $-\delta_\infty$, we have

$$\begin{aligned} &\left|\mathbb{E}\int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}-\Delta) - V_i(q_t^{i,\delta})\Big]f_a^i(\delta^{a,i}_{q_t^{i,\delta}}, (\vec{\delta}^{a,j})_{j\neq i})\mathbb{I}(q_t^{i,\delta} > -H_i)dt\right| \\ &\le 2M\cdot\mathbb{E}\left[\int_0^T e^{-rt}\Lambda(\delta^{a,i}_{q_t^{i,\delta}})dt\right] \le 2M\cdot\mathbb{E}\left[\int_0^T e^{-rt}\Lambda(-\delta_\infty)dt\right] \\ &\le 2M\Lambda(-\delta_\infty)\int_0^\infty e^{-rt}dt < \infty \end{aligned} \tag{32}$$

Similarly we also have

$$\left|\mathbb{E}\int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}+\Delta) - V_i(q_t^{i,\delta})\Big]f_b^i(\delta^{b,i}_{q_t^{i,\delta}}, (\vec{\delta}^{b,j})_{j\neq i})\mathbb{I}(q_t^{i,\delta} < H_i)dt\right| < \infty \tag{33}$$

Therefore we can take expectation on both sides of (31), and obtain

$$\begin{aligned} \mathbb{E}\Big[e^{-rT}V_i(q_T^{i,\delta})\Big] = V_i(q_i) &- \mathbb{E}\left[\int_0^T re^{-rt}V_i(q_t^{i,\delta})dt\right] \\ &+ \mathbb{E}\int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}-\Delta) - V_i(q_t^{i,\delta})\Big]f_a^i(\delta^{a,i}_{q_t^{i,\delta}}, (\vec{\delta}^{a,j})_{j\neq i})\mathbb{I}(q_t^{i,\delta} > -H_i)dt \\ &+ \mathbb{E}\int_0^T e^{-rt}\Big[V_i(q_t^{i,\delta}+\Delta) - V_i(q_t^{i,\delta})\Big]f_b^i(\delta^{b,i}_{q_t^{i,\delta}}, (\vec{\delta}^{b,j})_{j\neq i})\mathbb{I}(q_t^{i,\delta} < H_i)dt \end{aligned} \tag{34}$$

Since $V_i(q_i)$ satisfies HJB equation (28) for any $q_i \in \{-H_i, \dots, H_i\}$ with $(\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*$ being the Nash equilibrium quoting strategy, we can then replace $(\boldsymbol{\delta}^i)^*$ by $\boldsymbol{\delta}^i$ and obtain inequality

$$rV_i(q_t^{i,\delta}) + \psi_i(q_t^{i,\delta}) - \mathbb{I}(q_t^{i,\delta} > -H_i)\lambda^a\Delta\left[f_a^i(\delta^{a,i}_{q_t^{i,\delta}}, (\vec{\delta}^{a,j})^*_{j\neq i})\left(\delta^{a,i}_{q_t^{i,\delta}} - \frac{V_i(q_t^{i,\delta}) - V_i(q_t^{i,\delta} - \Delta)}{\Delta}\right)\right]$$
$$- \mathbb{I}(q_t^{i,\delta} < H_i)\lambda^b\Delta\left[f_b^i(\delta^{b,i}_{q_t^{i,\delta}}, (\vec{\delta}^{b,j})^*_{j\neq i})\left(\delta^{b,i}_{q_t^{i,\delta}} - \frac{V_i(q_t^{i,\delta}) - V_i(q_t^{i,\delta} + \Delta)}{\Delta}\right)\right] \geq 0 \tag{35}$$

Combining (34) and (35), we obtain

$$\mathbb{E}\left[e^{-rT}V_i(q_T^{i,\delta})\right] \leq V_i(q_i)$$
$$- \mathbb{E}\left[\int_0^T e^{-rt}\left(\mathbb{I}(q_t^{i,\delta} > -H_i)\lambda^a\Delta f_a^i(\delta^{a,i}_{q_t^{i,\delta}}, (\vec{\delta}^{a,j})^*_{j\neq i}) + \mathbb{I}(q_t^{i,\delta} < H_i)\lambda^b\Delta f_b^i(\delta^{b,i}_{q_t^{i,\delta}}, (\vec{\delta}^{b,j})^*_{j\neq i})\right)dt\right]$$
$$+ \mathbb{E}\left[\int_0^T e^{-rt}\psi_i(q_t^{i,\delta})\right] \tag{36}$$

From (32) and (33) and dominated convergence theorem, we can let $T \to \infty$ in both sides of (36). Since V_i is uniformly bounded, we have $\mathbb{E}[e^{-rT}V_i(q_T^{i,\delta})] \xrightarrow{T\to\infty} 0$. We then obtain

$$V_i(q_i) \geq$$
$$\mathbb{E}\left[\int_0^\infty e^{-rt}\left(\mathbb{I}(q_t^{i,\delta} > -H_i)\lambda^a\Delta f_a^i(\delta^{a,i}_{q_t^{i,\delta}}, (\vec{\delta}^{a,j})^*_{j\neq i}) + \mathbb{I}(q_t^{i,\delta} < H_i)\lambda^b\Delta f_b^i(\delta^{b,i}_{q_t^{i,\delta}}, (\vec{\delta}^{b,j})^*_{j\neq i})\right)dt\right]$$
$$- \mathbb{E}\left[\int_0^\infty e^{-rt}\psi_i(q_t^{i,\delta})\right] \tag{37}$$

The right hand side of (37) is exactly $J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*, q_i)$. Hence we have

$$V_i(q_i) \geq J_i(\boldsymbol{\delta}^i, (\boldsymbol{\delta}^{-i})^*, q_i) \tag{38}$$

$V_i(q_i)$ is the equilibrium value function in Definition 3.1.

Now if market maker i takes quoting strategy $(\boldsymbol{\delta}^i)^*$, while other market makers keep strategies $(\boldsymbol{\delta}^{-i})^*$, equality in (35) is achieved with $\boldsymbol{\delta}^i$ replaced by $(\boldsymbol{\delta}^i)^*$, as $(\boldsymbol{\delta}^i)^*$ is the maximum point in equation (28). Subsequently in (36) and (37) equality will be achieved with $\boldsymbol{\delta}^i$ replaced by $(\boldsymbol{\delta}^i)^*$. Therefore

$$V_i(q_i) \geq J_i((\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*, q_i) \tag{39}$$

Therefore $\{((\boldsymbol{\delta}^i)^*, (\boldsymbol{\delta}^{-i})^*), i \in \{1, \dots, N\}\}$ is the Nash equilibrium quoting strategy in Definition 3.1. □

3.2 | Existence of Nash equilibrium

With Proposition 3.4 and Proposition 3.5 the existence Nash equilibrium can be established by seeking the solution to the system of equations 28. We state the following existence result of Nash equilibrium. We briefly sketch the proof idea and complete the proof in Appendix A.

Theorem 3.6 (Existence of Nash equilibrium). *Under Assumptions 2.1 and 2.3 a Nash equilibrium exists.*

The proof of Theorem 3.6 is motivated by Luo and Zheng (2021). We extend the proof given by Luo and Zheng (2021) to the circumstance of multiple market makers. The first main result we prove is Proposition A.9 that shows the fixed point property of the arg max functions defined in (A.4). Subsequently we show by Proposition A.11 the uniqueness and continuity of the fixed point $\boldsymbol{\delta}_{\boldsymbol{p}} = \boldsymbol{\delta}(\boldsymbol{p})$ as function of any given value function vector $\boldsymbol{p}$, using a global implicit function theorem A.10. Finally we prove the system of non-linear equations (28) has a solution by Brouwer's fixed point theorem.

4 | COLLUSION AND PARETO OPTIMA

Collusion refers to the case where market makers coordinate their actions as a cartel in order to jointly maximize the sum of their objective functions, while sharing their inventory information Tirole (1988). An explicit collusion strategy is thus equivalent to solving a "central planner" problem whose objective is the sum of all market makers' profits.

Under collusion, market makers make decisions based on both their own inventory and inventories of other market makers in the cartel. Consequently the quoting strategy for market maker i, depends on the entire set of inventories $\boldsymbol{q} \in \prod_{j=1}^{N} \mathcal{Q}_j$ of all market makers. Note that this is an essential difference from competitive setting studied in Sections 2 and 3, where market maker i's quoting strategies are solely based on her own inventory level q_i. Her quoting strategies $\vec{\delta}^{a,i}, \vec{\delta}^{b,i}$ are represented as vectors, with coordinates indexed by $\boldsymbol{q} \in \prod_{j=1}^{N} \mathcal{Q}_j$, instead of $q_i \in \mathcal{Q}_i$. Because of the difference in dimension of quoting strategies, we need to adapt the execution probability functions f_a^i, f_b^i to be compatible with expanded dimension, as we will see in (47)–(48).

Denoting the product space by $\mathcal{Q} = \prod_{j=1}^{N} \mathcal{Q}_j$, ask and bid quotes have the form $\vec{\delta}^{a,i} = (\delta^{a,i}(\boldsymbol{q}))_{\boldsymbol{q}\in\mathcal{Q}}, \vec{\delta}^{b,i} = (\delta^{b,i}(\boldsymbol{q}))_{\boldsymbol{q}\in\mathcal{Q}}$, where $\vec{\delta}^{a,i}, \vec{\delta}^{b,i} \in (I_\delta)^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)}$. The quoting strategy of market maker i can thus be represented as $\boldsymbol{\delta}^i = (\vec{\delta}^{a,i}, \vec{\delta}^{b,i}) \in (I_\delta)^{\prod_{j=1}^{N} 2\frac{H_i}{\Delta}+1} \times (I_\delta)^{\prod_{j=1}^{N} 2\frac{H_i}{\Delta}+1}$. We denote $\mathcal{S} = \prod_{j=1}^{N}((I_\delta)^{\prod_{j=1}^{N} 2\frac{H_j}{\Delta}+1} \times (I_\delta)^{\prod_{j=1}^{N} 2\frac{H_j}{\Delta}+1})$ the space of (joint) two-sided quoting strategies for all market makers.

Definition 4.1 (Collusion). A set of quoting strategies $\vec{\delta}^c = ((\delta^1)^c, \dots, (\delta^N)^c) \in \mathcal{S}$ represents *collusion* if it maximizes the sum of all market makers' objective functions for any $\boldsymbol{q} \in \mathbb{R}^N$:

$$\sum_{i=1}^{N} J_i((\delta^i)^c, (\delta^{-i})^c; q_i) \geq \sum_{i=1}^{N} J_i(\delta^i, \delta^{-i}; q_i), \qquad \forall \vec{\delta} \in \mathcal{S} \tag{40}$$

For given joint quoting strategies $\vec{\delta} \in \mathcal{S}$ we denote by $\mathcal{J}$ the sum of objective functions of N market makers.

$$\mathcal{J}(\vec{\delta}; \boldsymbol{q}) = \sum_{i=1}^{N} J_i(\delta^i, \delta^{-i}; q_i) \tag{41}$$

The quantity

$$W(\boldsymbol{q}) = \mathcal{J}(\vec{\delta}^c; \boldsymbol{q}) = \sup_{\vec{\delta} \in \mathcal{S}} \mathcal{J}(\vec{\delta}; \boldsymbol{q}) \tag{42}$$

represents the *cartel* value function. As we see from Definition 4.1, computing the cartel's strategy and the corresponding value functions amounts to solving a stochastic optimal control problem for a central agent with objective function 41 in policy space $\mathcal{S}$.

Motivated by Cont et al. (2021), we show that collusion corresponds to a Pareto optimum:

Definition 4.2 (Pareto optimum). $\vec{\delta}^p = ((\delta^1)^p, \dots, (\delta^N)^p) \in \mathcal{S}$ is a Pareto-optimal policy if and only if there does not exist $\vec{\delta} \in \mathcal{S}$, such that for all $\boldsymbol{q} \in \prod_{j=1}^{N} \mathcal{Q}_j$,

$$\begin{aligned} &\forall i \in \{1, \dots, N\}, J_i(\delta^i, \delta^{-i}; q_i) \geq J_i((\delta^i)^p, (\delta^{-i})^p; q_i) \\ &\exists j \in \{1, \dots, N\}, J_j(\delta^j, \delta^{-j}; q_j) > J_j((\delta^j)^p, (\delta^{-j})^p; q_i) \end{aligned} \tag{43}$$

The following result links the concept of collusion with Pareto optima of the N market-maker system:

Proposition 4.3. *Any collusion strategy $\vec{\delta}^c = ((\delta^1)^c, \dots, (\delta^N)^c)$ in the sense of Definition 4.1 is a Pareto optimum as defined in Definition 4.2.*

Proof. By Definition 4.1, for any joint quoting strategy $\vec{\delta} \in \mathcal{S}$,

$$W(\boldsymbol{q}) \geq \mathcal{J}(\vec{\delta}; \boldsymbol{q}) = \sum_{i=1}^{N} J_i(\delta^i, \delta^{-i}; q_i) \tag{44}$$

If there exists a quoting strategy $\vec{\delta}'$ and $k \in \{1, \dots, N\}$ such that

$$J_k((\delta^i)', (\delta^{-i})'; q_i) > J_k((\delta^i)^c, (\delta^{-i})^c; q_i) \tag{45}$$

meaning that for market maker k there is a strictly better joint quoting strategy, we need to prove that $\vec{\delta}'$ cannot be an 'overall better' joint strategy for all market makers. Since

$$\sum_{j=1}^{N} J_j((\delta^i)', (\delta^{-i})'; q_i) \le W(\boldsymbol{q}) = \sum_{j=1}^{N} J_j((\delta^i)^c, (\delta^{-i})^c; q_i) \tag{46}$$

There must exists another market maker $j \neq k$ such that her objective

$$J_j((\delta^i)', (\delta^{-i})'; q_i) < J_j((\delta^i)^c, (\delta^{-i})^c; q_i)$$

which can be verified by contradiction argument with (45-46). Therefore $\vec{\delta}'$ is not a "globally better" strategy than $\vec{\delta}^c$ for all market makers. by Definition 4.2, the explicit collusion strategy $\vec{\delta}^c$ is a Pareto-optimal policy. □

Compared to (28) the dimension of joint quoting strategy is changed since under explicit collusion each market makers' ask and bid quotes are functions of $\boldsymbol{q}$, instead of their own inventory level. We assume that under explicit collusion, the execution probability for market maker i, denoted by $\tilde{f}^i_a, \tilde{f}^i_b$, are functions defined on $\mathbb{R} \times \prod_{j\neq i} \mathbb{R}^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)}$. For $\delta \in \mathbb{R}, \delta^{-i} \in \prod_{j\neq i} \mathbb{R}^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)}$,

$$\tilde{f}^i_a(\delta, \vec{\delta}^{-i}) : \mathbb{R} \times \prod_{k\neq i} \mathbb{R}^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)} \to \mathbb{R}^+, \tilde{f}^i_b(\delta, \vec{\delta}^{-i}) : \mathbb{R} \times \prod_{k\neq i} \mathbb{R}^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)} \to \mathbb{R}^+ \tag{47}$$

$\tilde{f}^i_a, \tilde{f}^i_b$ are interconnected through f^i_a, f^i_b in the following way. We take the example of ask quotes. For given joint inventory $\boldsymbol{q}$ and joint ask quoting strategies $(\delta^{a,i}_{\boldsymbol{q}}, ((\vec{\delta}^{a,j}_{\tilde{\boldsymbol{q}}})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i})$,

$$\tilde{f}^i_a(\delta^{a,i}_{\boldsymbol{q}}, ((\vec{\delta}^{a,j}_{\tilde{\boldsymbol{q}}})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i}) = f^i_a(\delta^{a,i}_{\boldsymbol{q}}, (\bar{\delta}^{a,j})_{j\neq i}) \tag{48}$$

where $\bar{\delta}^{a,j} \in \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ is a $(2\frac{H_j}{\Delta}+1)$ dimensional vector indexed by $\mathcal{Q}_j$. $\bar{\delta}^{a,j}$ is defined by taking the average of market maker j's ask quotes at joint inventories where her inventory is q_j: $\forall q_j \in \mathcal{Q}_j$, $\bar{\delta}^{a,j}_{q_j} = \frac{1}{\prod_{k=1,k\neq j}^{N}(2\frac{H_k}{\Delta}+1)} \sum_{\tilde{\boldsymbol{q}}\in\mathcal{Q}, \tilde{\boldsymbol{q}}_j = q_j} \delta^{a,j}_{\tilde{\boldsymbol{q}}}$, where $\delta^{a,j}_{\tilde{\boldsymbol{q}}}$ is the coordinate of vector $(\vec{\delta}^{a,j}_{\tilde{\boldsymbol{q}}})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}}$ at joint inventory index $\tilde{\boldsymbol{q}}$.

From dynamic programming principle, value function $W(\boldsymbol{q})$ satisfies HJB equation: $\forall \boldsymbol{q} \in \prod_{j=1}^{N} \mathcal{Q}_j$,

$$rW(\boldsymbol{q}) + \sum_{i=1}^{N} \psi_i(q_i) - \lambda^a \Delta \sup_{\vec{\delta}\in S} \sum_{i=1}^{N} \mathbb{I}(q_i > -H_i) \left[\tilde{f}^i_a\left(\delta^{a,i}_{\boldsymbol{q}}, ((\vec{\delta}^{a,j}_{\tilde{\boldsymbol{q}}})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i}\right) \left(\delta^{a,i}_{\boldsymbol{q}} - \frac{W(\boldsymbol{q}) - W(\boldsymbol{q} - \Delta \boldsymbol{e}_i)}{\Delta} \right) \right]$$

$$- \lambda^b \Delta \sup_{\vec{\delta}\in S} \sum_{i=1}^{N} \mathbb{I}(q_i < H_i) \left[\tilde{f}^i_b\left(\delta^{b,i}_{\boldsymbol{q}}, ((\vec{\delta}^{b,j}_{\tilde{\boldsymbol{q}}})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i}\right) \left(\delta^{b,i}_{\boldsymbol{q}} - \frac{W(\boldsymbol{q}) - W(\boldsymbol{q} + \Delta \boldsymbol{e}_i)}{\Delta} \right) \right] = 0 \tag{49}$$

We further justify our choice of $\tilde{f}_a^i, \tilde{f}_b^i$ from (49). If if we still formulate the central planner's problem with partially observed information, that is, applying quoting strategies as functions of single q_i and keeping f_a^i, f_b^i in (49), (49) would be otherwise an over-determined system which derives, respectively, on ask and bid sides $\prod_{j=1}^{N}(2\frac{H_i}{\Delta}+1)$ single variate optimization problem, but only $\sum_{j=1}^{N}(2\frac{H_i}{\Delta}+1)$ variables on each side ($\delta_{q_j}^{a,i}$ and $\delta_{q_j}^{b,i}$). Hence to formulate explicit collusion, we define execution probability function $\tilde{f}^i$ that is close to $f_m^i, m \in \{a,b\}$, but compatible with expanded dimension of quoting strategies.

The existence and uniqueness of solutions to (49) can be established with methods of stochastic control on graphs in Guéant and Manziuk (2020). However, solving equations (49) numerically involves optimization in high dimension space, which is fairly time consuming due to curse of dimensionality, making the problem unrealistic to solve directly. Instead we approximate the collusive spreads with a linear function of inventory levels and solve the optimization on parameters of a linear function. For simplicity we only consider homogeneous market makers, but the approach can be naturally generalized to market makers with different parametrization. The only difference is that with homogeneous market makers, we can apply shared approximation functions for their quoting strategies.

Namely for $\boldsymbol{q} \in \mathcal{Q}$, the approximated ask and bid quotes for homogeneous market makers are

$$\delta_{\boldsymbol{q}}^{a,i} \approx \xi_0^a + \sum_{k=1}^{N} \xi_k^a \cdot q_k \qquad \delta_{\boldsymbol{q}}^{b,i} \approx \xi_0^b + \sum_{k=1}^{N} \xi_k^b \cdot q_k \tag{50}$$

where $\{\xi_k^a, \xi_k^b | k \in \{0,1,\dots,N\}\}$ are the parameters to optimize.

For given joint quoting strategies $\vec{\boldsymbol{\delta}}$, denote $W^{\delta}(\boldsymbol{q})$ the value function associated to strategies $\vec{\boldsymbol{\delta}}$, then $W^{\delta}(\boldsymbol{q})$ satisfies system of linear equations.

$$\begin{aligned} rW^{\delta}(\boldsymbol{q}) + \sum_{i=1}^{N} \psi_i(q_i) - \lambda^a \Delta \sum_{i=1}^{N} \mathbb{I}(q_i > -H_i) \left[\tilde{f}_a^i \left(\delta_{\boldsymbol{q}}^{a,i}, ((\vec{\delta}_{\bar{\boldsymbol{q}}}^{a,j})_{\bar{\boldsymbol{q}} \in \mathcal{Q}})_{j \neq i} \right) \left(\delta_{\boldsymbol{q}}^{a,i} - \frac{W^{\delta}(\boldsymbol{q}) - W^{\delta}(\boldsymbol{q} - \Delta \boldsymbol{e}_i)}{\Delta} \right) \right] \\ - \lambda^b \Delta \sum_{i=1}^{N} \mathbb{I}(q_i < H_i) \left[\tilde{f}_b^i \left(\delta_{\boldsymbol{q}}^{b,i}, ((\vec{\delta}_{\bar{\boldsymbol{q}}}^{a,j})_{\bar{\boldsymbol{q}} \in \mathcal{Q}})_{j \neq i} \right) \left(\delta_{\boldsymbol{q}}^{b,i} - \frac{W^{\delta}(\boldsymbol{q}) - W^{\delta}(\boldsymbol{q} + \Delta \boldsymbol{e}_i)}{\Delta} \right) \right] = 0 \end{aligned} \tag{51}$$

We solve (51) for $(W^{\delta}(\boldsymbol{q}))_{q \in \mathcal{Q}}$ given $\vec{\boldsymbol{\delta}} \in \mathcal{S}$. For each $\boldsymbol{q}$ we seek the parameters $\xi^{a,*} = (\xi_k^{a,*})_{k \in \{1,\dots,N\}}, \xi^{b,*} = (\xi_k^{b,*})_{k \in \{1,\dots,N\}}$ that maximizes the value $W^{\delta}(\boldsymbol{q})$. The collusive spread is then approximated by

$$(\delta_{\boldsymbol{q}}^{a,i})^* \approx \xi_0^{a,*} + \sum_{k=1}^{N} \xi_k^{a,*} \cdot q_k, (\delta_{\boldsymbol{q}}^{b,i})^* \approx \xi_0^{b,*} + \sum_{k=1}^{N} \xi_k^{b,*} \cdot q_k$$

More specifically we apply a policy iteration scheme for numerically approximating the Pareto optimum. The algorithm consists of iteratively executing policy evaluation and policy improvement, where the linear approximation (50) is applied in policy improvement step. Within each iteration, policy evaluation solves a linear system of equation for a given joint quoting strategies

Algorithm 0 Policy iteration of Pareto optimum for N market makers

Input: M =number of iterations, N =number of market makers, $\tilde{f}_a^i, \tilde{f}_b^i, \lambda^a, \lambda^b$: intensity of ask and bid order flow, Δ: unit order size, ψ_i: running cost for holding inventory to market maker i

Output: Approximated Pareto optimum quoting strategy

1: Initialize quoting strategies of N market makers, denoted by $\vec{\boldsymbol{\delta}}^{(0)} = \{(\vec{\delta}^{a,i,(0)}, \vec{\delta}^{b,i,(0)})\}$.

2: **for** $m \leftarrow 0$ to M-1 **do**

3: **Policy evaluation**: Compute values $\{W^{\delta,(m+1)}(\boldsymbol{q}), \boldsymbol{q} \in \prod_{j=1}^{N} \mathcal{Q}_j\}$ by solving linear system (4.13) using joint quoting strategies $\vec{\boldsymbol{\delta}}^{(m)}$.

4: **Policy improvement**:

5: Let $(\xi_k^{a,(m+1)}), (\xi_k^{b,(m+1)})$ be unknown coefficients, assign the functional relationship:

$$\delta_{\boldsymbol{q}}^{a,i,(m+1)} \leftarrow \xi_0^{a,(m+1)} + \sum_{k=1}^{N} \xi_k^{a,i,(m+1)} \cdot q_k$$

$$\delta_{\boldsymbol{q}}^{b,i,(m+1)} \leftarrow \xi_0^{b,(m+1)} + \sum_{k=1}^{N} \xi_k^{b,i,(m+1)} \cdot q_k$$

Solve following optimization problems

$$(\xi_k^{a,(m+1)})_{k\in\{1,\dots,N\}} = \arg\max_{\xi} \sum_{\boldsymbol{q}} \left\{ \sum_{i=1}^{N} \mathbb{1}_{(q_i > -H_i)} \left[f_a^i \left(\delta_{\boldsymbol{q}}^{a,i,(m+1)}, (\delta_{\tilde{\boldsymbol{q}}}^{a,j,(m+1)})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i} \right) \left(\delta_{\boldsymbol{q}}^{a,i,(m+1)} - \frac{W^{\delta,(m+1)}(\boldsymbol{q}) - W^{\delta,(m+1)}(\boldsymbol{q} - \Delta \boldsymbol{e}_i)}{\Delta} \right) \right] \right\}$$

$$(\xi_k^{b,(m+1)})_{k\in\{1,\dots,N\}} = \arg\max_{\xi} \sum_{\boldsymbol{q}} \left\{ \sum_{i=1}^{N} \mathbb{1}_{(q_i < H_i)} \left[f_b^i \left(\delta_{\boldsymbol{q}}^{b,i,(m+1)}, (\delta_{\tilde{\boldsymbol{q}}}^{b,j,(m+1)})_{\tilde{\boldsymbol{q}}\in\mathcal{Q}})_{j\neq i} \right) \left(\delta_{\boldsymbol{q}}^{b,i,(m+1)} - \frac{W^{\delta,(m+1)}(\boldsymbol{q}) - W^{\delta,(m+1)}(\boldsymbol{q} + \Delta \boldsymbol{e}_i)}{\Delta} \right) \right] \right\}$$

6: Update $\delta_{\boldsymbol{q}}^{a,i,(m+1)}, \delta_{\boldsymbol{q}}^{b,i,(m+1)}$ using calculated $(\xi_k^{a,(m+1)})_{k\in\{1,\dots,N\}}, (\xi_k^{b,(m+1)})_{k\in\{1,\dots,N\}}$.

7: **end for**

$\vec{\boldsymbol{\delta}} \in \mathcal{S}$ based on linear Bellman equations (51).

$$M^{\delta} \cdot W^{\delta} = A^{\delta} \tag{52}$$

where M^{δ} is $\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1) \times \prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)$ matrix, $W^{\delta}, A^{\delta} \in \mathbb{R}^{\prod_{j=1}^{N}(2\frac{H_j}{\Delta}+1)}$. The data M^{δ}, A^{δ} are derived from (51). Subsequently policy improvement solves optimization problems from Bellman equations (49) given values W^{δ} calculated from policy evaluation, with unknowns being $\{\xi_k^a, k \in \{1, \dots, N\}\}$ and $\{\xi_k^a, k \in \{1, \dots, N\}\}$ from (50). The details of policy evaluation method are described in Algorithm 0.

As shown in Xiong and Cont (2021), when market makers adjust their strategies by learning from market transactions, spreads may converge to levels consistent with collusion, even in absence of any coordination across market makers. We refer to this situation as *tacit collusion* Tirole (1988). As in Calvano et al. (2020), our definition of tacit collusion is thus an operational one, referring only to price outcomes.

5 | COMPUTATION OF NASH EQUILIBRIUM VIA FICTITIOUS PLAY

Fictitious play is an iterative algorithm for computing Nash equilibrium. It is originally proposed in Brown (1949) and Brown (1951) to compute the value of two-player zero-sum finite game, where two players each play iteratively the pure best response against the empirical distribution of their opponent's historical actions. The convergence of fictitious play for two-player zero-sum finite game is proved by Robinson (1951). Monderer and Shapley (1996) proves convergence of fictitious

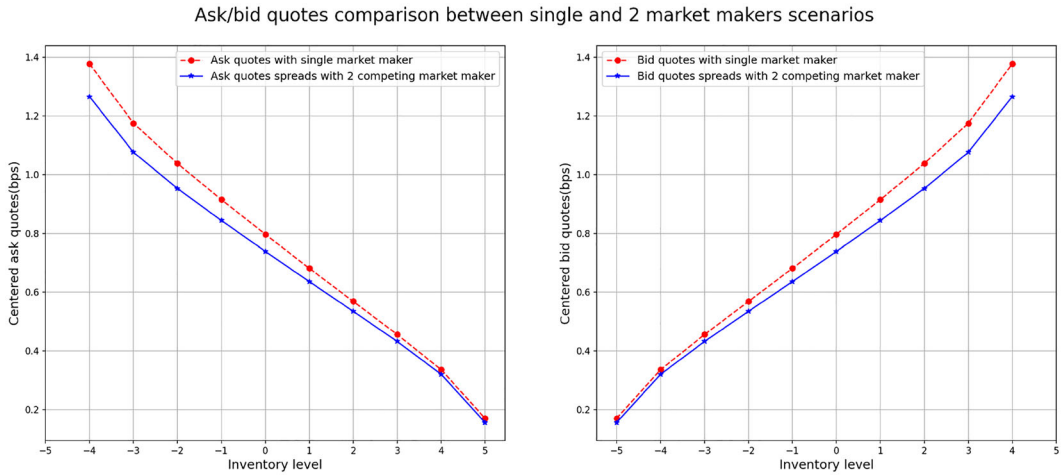


FIGURE 1 Comparison between two -market-maker equilibrium ask and bid quotes and single-market-maker's quotes. The ask and bid quotes are centered at mid price. [Color figure can be viewed at wileyonlinelibrary.com]

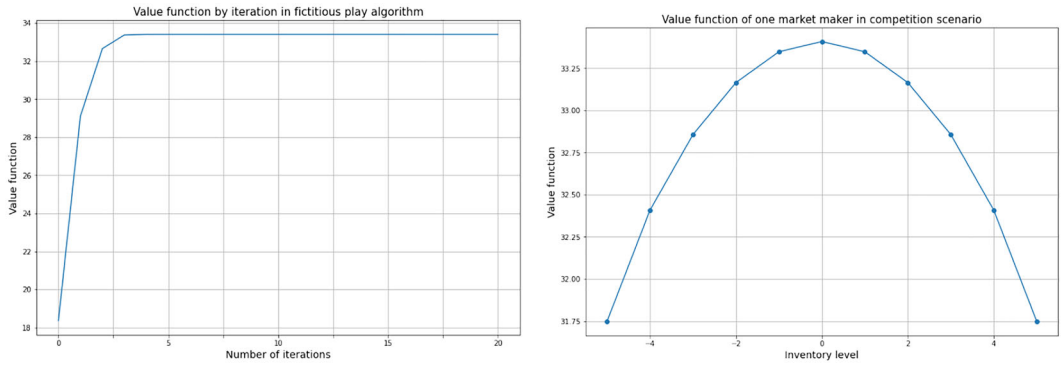


FIGURE 2 **Left**: evolution of value functions during fictitious play iterations; **Right**: equilibrium value function obtained from fictitious play algorithm. [Color figure can be viewed at wileyonlinelibrary.com]

play for potential games. However, there are no theoretical guarantee of convergence for general non-zero sum games, as is shown by a counterexample from Shapley (1962). Nevertheless, the idea of fictitious play has been a standard tool of game theory for computing practically Nash equilibrium, and has been extended to broader applications such as mean field game learning problems (e.g., Cardaliaguet & Hadikhanloo, 2017; Perrin et al., 2020). In this section, we will be focusing on practical aspects, by applying fictitious play to compute numerically Nash equilibrium for multi-agent market making problem. Our numerical results in Figures 1–3 suggest fictitious play leads to convergent quotes. We shall leave theoretical convergence analysis of fictitious play for our specific problem for future research.

Recently fictitious play has been combined with deep learning methods to find (Markovian) Nash equilibria in dynamic games (Han & Hu, 2020), with convergence analysis studied in Han et al. (2022). Hu (2021) has analyzed the convergence of deep fictitious play algorithm for finding open-loop Nash equilibria. Specifically deep fictitious play algorithm decouples N-player game

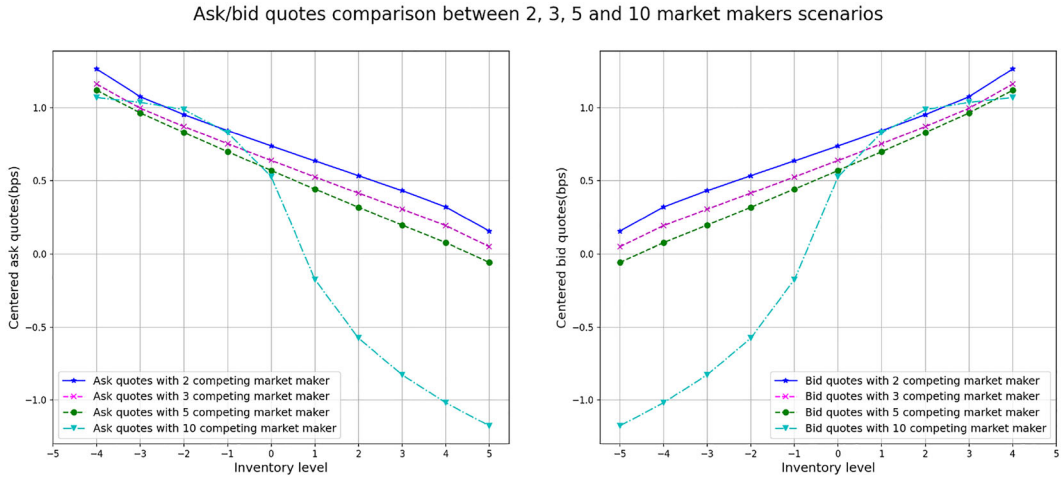


FIGURE 3 Comparison of equilibrium quotes with different number of market makers. The 0 value in vertical axis represents mid price. [Color figure can be viewed at wileyonlinelibrary.com]

into N decision problems solved iteratively, in which each player solves the optimal strategy using deep neural networks, given that competitors' strategies remain fixed as in previous iteration.

Inspired by Han and Hu (2020), the fictitious play algorithm we propose for N competing market makers problem is based on systems of equations (27) and (28). For each equation in (28), the optimization problem at a given inventory level q is a single-variate optimization problem. Hence at each iteration of fictitious play one player needs to solve single-variate optimization problem (57) for each of her inventory level, while the competitors' strategies in intensity functions are fixed as in previous iteration. This single-variate optimization allows to avoid curse of dimensionality due to number of players and inventory levels of each player, because otherwise a policy evaluation method needs to optimize simultaneously the quotes of all market makers at all inventory levels at every iteration, leading to high-dimensional optimization problems. Note that the fictitious play algorithm is designed to solve numerically Nash equilibrium, instead of simulating real competition among market makers. In practice market makers do not have information on their competitors' strategies, while in fictitious play each player optimizes for next step based on competitors' current strategies. We shall tackle the simulation through Deep Reinforcement Learning in next section.

Note that in usual fictitious play algorithm, agents play best response against the mixed strategy derived from their opponents' historical actions (Brown, 1949; Brown, 1951). In our definition best response is played solely based on competitors' last stage actions. The algorithm based on last stage information is sometimes referred to as "Iterated Best Response (IBR)" (Lanctot et al., 2017) or "Best Response Dynamics" (Roughgarden, 2016), in which an arbitrary agent is picked to update its best response to opponents' last stage actions at each iteration. We shall adhere to our setting with the name "fictitious play". This will be explained in more details in Remark 5.1.

Our fictitious play algorithm consists of iteratively executing two steps: **best response calculation** and **policy evaluation**. Policy evaluation computes the values $V_i^{\delta}(q_i)$ for joint quoting strategies based on the linear optimality equations (27). Given joint quoting strategies $(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i})$, we reformulate (27) as a system of linear equations with values $V_i^{\delta}(q_i)$ being unknown variables.

To simplify notations, we denote

$$f_a^i(\delta_{q_i}^{a,i}) = f_a^i(\delta_{q_i}^{a,i}, (\vec{\delta}^{a,j})_{j\neq i}), f_b^i(\delta_{q_i}^{b,i}) = f_b^i(\delta_{q_i}^{b,i}, (\vec{\delta}^{b,j})_{j\neq i})$$

The system of linear equations obtained is

$$\begin{cases} \left(r + \lambda^b f_b^i(\delta_{-H_i}^{b,i})\right) V_i^\delta(-H_i) - \lambda^b f_b^i(\delta_{-H_i}^{b,i}) V_i^\delta(-H_i + \Delta) = \lambda^b \Delta \delta_{-H_i}^{b,i} f_b^i(\delta_{-H_i}^{b,i}) - \psi_i(-H_i) \\ -\lambda^a f_a^i(\delta_{q_i}^{a,i}) V_i^\delta(q_i - \Delta) + \left(r + \lambda^b f_b^i(\delta_{-H_i}^{b,i}) + \lambda^a f_a^i(\delta_{H_i}^{a,i})\right) V_i^\delta(q_i) - \lambda^b f_b^i(\delta_{q_i}^{b,i}) V_i^\delta(q_i + \Delta) \\ \qquad = \lambda^a \Delta \delta_{q_i}^{a,i} f_a^i(\delta_{q_i}^{b,i}) + \lambda^b \Delta \delta_{q_i}^{b,i} f_b^i(\delta_{q_i}^{b,i}) - \psi_i(q_i), \quad \text{if } q_i \in \mathcal{Q}_i \setminus \{-H_i, H_i\} \\ -\lambda^a f_a^i(\delta_{H_i}^{a,i}) V_i^\delta(H_i - \Delta) + \left(r + \lambda^a f_a^i(\delta_{H_i}^{a,i})\right) V_i^\delta(H_i) = \lambda^a \Delta \delta_{H_i}^{a,i} f_a^i(\delta_{-H_i}^{b,i}) - \psi_i(H_i) \end{cases} \tag{53}$$

Let $\vec{V}_i^\delta := (V_i^\delta(-H_i), V_i^\delta(-H_i + \Delta), \ldots, V_i^\delta(H_i - \Delta), V_i^\delta(H_i))$, then (53) can be formulated into following matrix representation.

$$M_i \cdot \vec{V}_i = A_i \tag{54}$$

where $M_i \in \mathbb{R}^{(2\frac{H_i}{\Delta}+1)\times(2\frac{H_i}{\Delta}+1)}$, $A_i \in \mathbb{R}^{2\frac{H_i}{\Delta}+1}$.

$$M_i = \begin{bmatrix} r + \lambda^b f_b^i(\delta_{-H_i}^{b,i}) & -\lambda^b f_b^i(\delta_{-H_i}^{b,i}) & 0 & \ldots & 0 & 0 & 0 \\ -\lambda^a f_a^i(\delta_{-H_i+\Delta}^{a,i}) & r + \lambda^b f_b^i(\delta_{-H_i+\Delta}^{b,i}) + \lambda^a f_a^i(\delta_{-H_i+\Delta}^{a,i}) & -\lambda^b f_b^i(\delta_{-H_i+\Delta}^{b,i}) & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & -\lambda^a f_a^i(\delta_{H_i-\Delta}^{a,i}) & r + \lambda^b f_b^i(\delta_{H_i-\Delta}^{b,i}) + \lambda^a f_a^i(\delta_{H_i-\Delta}^{a,i}) & -\lambda^b f_b^i(\delta_{H_i-\Delta}^{b,i}) \\ 0 & 0 & 0 & \ldots & 0 & -\lambda^a f_a^i(\delta_{H_i}^{a,i}) & r + \lambda^a f_a^i(\delta_{H_i}^{a,i}) \end{bmatrix} \tag{55}$$

$$A_i = \begin{bmatrix} \lambda^b \Delta \delta_{-H_i}^{b,i} f_b^i(\delta_{-H_i}^{b,i}) - \psi_i(-H_i) \\ \lambda^a \Delta \delta_{-H_i+\Delta}^{a,i} f_a^i(\delta_{-H_i+\Delta}^{b,i}) + \lambda^b \Delta \delta_{-H_i+\Delta}^{b,i} f_b^i(\delta_{-H_i+\Delta}^{b,i}) - \psi_i(-H_i + \Delta) \\ \ldots \\ \lambda^a \Delta \delta_{H_i-\Delta}^{a,i} f_a^i(\delta_{H_i-\Delta}^{b,i}) + \lambda^b \Delta \delta_{H_i-\Delta}^{b,i} f_b^i(\delta_{H_i-\Delta}^{b,i}) - \psi_i(H_i - \Delta) \\ \lambda^a \Delta \delta_{H_i}^{a,i} f_a^i(\delta_{-H_i}^{b,i}) - \psi_i(H_i) \end{bmatrix} \tag{56}$$

Clearly the matrix M_i is diagonally dominant matrix, hence is invertible. For given strategies, we can solve for value functions

$$V_i^\delta = M_i^{-1} \cdot A_i$$

Subsequently in best response calculation each agent i solves her best response to $\boldsymbol{\delta}^{-i}$ using (57). Han and Hu (2020) applies Deep BSDE method to solve this optimization problem. We directly apply standard numerical optimization scheme since the objectives in (57) are single-variate hence not complicated with Assumption 2.3.

The details of fictitious play is presented in Algorithm 1.

Algorithm 1 Fictitious play for computation of Nash equilibrium for N market makers

Input: M =number of iterations, N =number of market makers, $f_a^i, f_b^i, \lambda^a, \lambda^b$: intensity of ask and bid order flow, Δ: unit order size, ψ_i: running cost for holding inventory to market maker i

Output: Approximated Nash equilibrium by fictitious play.

Initialize quoting strategies of N market makers, denoted by $\{(\vec{\delta}^{a,i,(0)}, \vec{\delta}^{b,i,(0)})\}$.

Compute initial values $\{V_i^{(0)}(q_i), q_i \in \mathcal{Q}_i, i \in \{1, \dots, N\}\}$ by solving linear system (5.1)

for $m \leftarrow 0$ to M-1 **do**

  Best response: Solve optimal strategy for single market maker at every inventory level:

  for $i \leftarrow 1$ to N **do**

    for $q_i \in \mathcal{Q}_i$ **do**

      Update

$$\delta_{q_i}^{a,i,(m+1)} = \arg\max_{\delta} f_a^i(\delta, (\vec{\delta}^{a,j,(m)})_{j\neq i})(\delta - \frac{V_i^{(m)}(q_i) - V_i^{(m)}(q_i - \Delta)}{\Delta})$$
$$\delta_{q_i}^{b,i,(m+1)} = \arg\max_{\delta} f_b^i(\delta, (\vec{\delta}^{b,j,(m)})_{j\neq i})(\delta - \frac{V_i^{(m)}(q_i) - V_i^{(m)}(q_i + \Delta)}{\Delta}) \tag{5.5}$$

    end for

  end for

  Policy evaluation: Compute values $\{V_i^{(m+1)}(q_i), q_i \in \mathcal{Q}_i, i \in \{1, \dots, N\}\}$ by solving linear system (5.1) using updated strategies $\{(\vec{\delta}^{a,i,(m+1)}, \vec{\delta}^{b,i,(m+1)})\}$.

end for

Remark 5.1. We are following the definition of fictitious play from Han and Hu (2020), Hu (2021) and Han et al. (2022), which is different from that originated from Brown (1949) and Brown (1951), in that the N agents update simultaneously their best responses against their competitors' pure strategies from last stage. One can alternatively choose computing best response against average of opponents' past strategies in the spirit of classical fictitious play Brown (1951). But as Hu (2021) points out, convergence with last stage information generally implies convergence with average of past strategies, and switching to the latter tends to better convergence rate for certain circumstances, but with extra computational cost (Han & Hu, 2020). For our setting since problem we study already exhibits convergence with last stage information within reasonable number of iterations, we adhere to this setting of best response using last stage information.

By Lemma A.1 there exists unique solutions $\delta_{q_i}^{a,i,(m+1)}, \delta_{q_i}^{b,i,(m+1)}$ to optimization problems in (57). Let $K_i = \sum_{j\neq i, j\in\{1,\dots,N\}} (2\frac{H_i}{\Delta} + 1)$. K_i is the total number of possible inventory levels of market maker i's competitors. We consider f_a^i and f_b^i of the form:

$$f_a^i(\delta, (\vec{\delta}^{a,j})_{j\neq i}) = \frac{1}{1+\exp(\delta)} \frac{\exp(\frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\delta_{q_j}^{a,j})}{1+\exp(\delta + \frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\delta_{q_j}^{a,j})}$$
$$f_b^i(\delta, (\vec{\delta}^{b,j})_{j\neq i}) = \frac{1}{1+\exp(\delta)} \frac{\exp(\frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\delta_{q_j}^{b,j})}{1+\exp(\delta + \frac{1}{K_i}\sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\delta_{q_j}^{b,j})} \tag{58}$$

We can also verify that the execution probabilities (58) satisfy Assumptions 2.1 and 5 when there are two market makers. We also assume each of two market makers has the same running cost function $\psi_1(q) = \psi_2(q) = \frac{1}{2} \times 0.01q^2$. The market makers have the same inventory risk limit $Q_1 = Q_2 = 5$. Order size $\Delta = 1$. Interest rate $r = 0.01$ and order flow arrival intensities $\lambda^a = \lambda^b = 2$. For simplicity we assume the units of all parameters are already scaled so that the unit of ask and bid quotes is basis point (bps). We apply fictitious play to the game with two market makers, and compare the ask/bid quotes with a benchmark model where there is only one monopolistic market maker with the ask and bid intensity $\Lambda(\delta) = \frac{1}{(1+e^{\delta})^2}$ where δ is the ask or bid quote of this monopolistic market maker. Figure 1 compares the equilibrium quotes obtained from fictitious play algorithm when there are two market makers, with the optimal centered quotes when there is one monopolistic market maker. The results show lower Nash equilibrium quotes compared to monopolistic market maker's quotes, suggesting competition introduced through the execution probabilities (58) results in more competitive quotes than monopolistic market maker case. Figure 1 also suggests skewing behavior by market makers caused by exposure to market risk of their non-zero inventories.

We also plot in Figure 2 the evolution of value function at inventory level 0 with 2 market makers to illustrate how fictitious play algorithm leads to a stable state in our experiment setting. The equilibrium value function is also shown in Figure 2. The value function achieves its maximum at 0 inventory level, this is in line with the market makers expecting higher gain by keeping inventory near 0 since they pay less running cost for holding inventory and are exposed to less market risk on their inventories.

To study the effects by number of market makers in competition, Figure 3 compares the equilibrium quotes in scenarios of two and five market makers. Figure 3 suggests that more market makers tend to have more competitive quotes than the case of two market makers. The equilibrium quotes are lower at same inventory level when number of market makers increases to 5. However, the value of quotes do not show an explicit decreasing trend on half of inventory levels when there are 10 market makers. We think that this could be explained by market makers' compensating effect from excessively crossing the spreads at the other half of inventory levels when when there are many market makers. For instance the 10 market makers tend to keep their ask quotes high at negative inventories because they quote excessively negative values at positive inventories. It is worth noting that with 5 and 10 competing market makers when inventory level is at boundaries Q and $-Q$, market makers even cross the spreads with negative ask and bid quotes. This is due to the excessive running cost for holding larger inventory so that market makers would rather change their inventory state at the cost of losing profit from making the spread. The phenomenon of crossing the spreads is even more remarkable when number of market makers reaches 10, implying that when there are many market makers they are more risk averse for holding non-zero inventory.

6 | DECENTRALIZED LEARNING AND THE EMERGENCE OF TACIT COLLUSION

The notions of Nash equilibrium and Pareto optimum described above refer to outcomes but do not attempt to describe the process through which agents arrive at their quoting strategies. In practice, market makers rely on algorithms which update their quotes based on observed market

data, and an interesting question is to understand whether the agents' strategies converges to any stationary configuration under this learning process and how such a limit may be characterized.

Market makers update their quotes dynamically and receive feedback in the form of profit from market transactions, based on which they adjust their quote. The automation of these market making algorithms naturally points to the use of Reinforcement Learning (RL) algorithms to model the resulting dynamics.

Also, in dealer markets the market makers are not allowed to communicate and make decisions based on partial information related to their own inventory, quotes and profits. These features require RL algorithm to adapt to continuous state or action spaces. Hence we focus on *decentralized multi-agent policy gradient algorithms*, which account for a type of deep reinforcement learning algorithm that allow for continuous state and action spaces.

For each market maker we introduce two types of neural networks: the critic and actor networks. Critic network is used to approximate the state-action value function $V_i^{\delta}(q_i)$ while actor network approximates optimal quoting strategies. The algorithm we apply is called decentralized Multi-Agent Deep Deterministic Policy Gradient (Decentralized MADDPG), inspired by Lowe et al. (2017) and Foerster et al. (2018). Lowe et al. (2017) and Foerster et al. (2018) proposed multi-agent actor-critic algorithm with decentralized actors and centralized critics, in which the actors are functions of each agent's local observation, and critics are functions of all agents' joint states and actions. In their algorithms critics are trained in centralized way, which implies certain communication during training steps needs to be allowed. We adapt the MADDPG algorithm to our market making model by constraining the critics to be decentralized as well.

An important feature of our Decentralized MADDPG algorithm is pre-training of critic and actor networks for initialization. Pre-training is important for the convergence of quoting strategies and has been considered by Guéant and Manziuk (2019). Recall that in Section 2 the upper bound $\Lambda(\delta)$ can be considered as the execution probability of a monopolistic market maker. We pre-train the critic networks for each market maker to the value function of a single monopolistic market maker with intensity $\Lambda(\delta)$, and pre-train the actor networks to the quoting strategies of the same monopolistic market maker, which is shown by the red curve in Figure 1. This pre-training step can be implemented by supervised learning on neural networks, since the value function and quoting strategy in monopolistic case can be explicitly calculated. We think the motif for pre-training is consistent with practical scenarios in that even though market makers do not have information on the mechanics that influence their market shares, each market maker is supposed to have a prior estimate on the general form of the execution probability when there is no other competitor. In the meantime, the actor network is initialized by pre-training so that the ask and bid quotes are within a reasonable value range. For simplicity we assume market makers know Λ.

6.1 | Reformulation to discrete-time problem

Equations (27) and (28) are local versions of the Dynamic Programming Principle. To adapt to RL simulation, we first need to formulate the multi-agent market making problem into discrete-time Bellman equations. In the following derivations, we always assume the joint quoting strategies are given and fixed $\vec{\delta} \in \prod_{j=1}^{N} (I_\delta)^{2\frac{H_j}{\Delta}+1}$. The corresponding state-action value functions are denoted by

$V_i^\delta(q_i)$ where i refers to the index of market maker, and q_i is the inventory level of market maker i at time 0. Let $\mathbb{E}_i[\cdot]$ denote the conditional expectation $\mathbb{E}[\cdot|q_0^i = q_i]$. From (9) we have

$$V_i^\delta(q_i) = \mathbb{E}_i\left[\int_0^\infty e^{-rt}\left(\delta_{q_t^i}^{a,i}N^{a,i}(dt) + \delta_{q_t^i}^{b,i}N^{b,i}(dt)\right) - \int_0^\infty e^{-rt}\psi_i(q_t^i)dt\right] \tag{59}$$

where $N^{a,i}(dt)$ and $N^{b,i}(dt)$ are order flow to market makers i whose intensities are defined in (2). Define two stopping times τ_a and τ_b denoting the arrival time of first ask and bid RFQ after 0, that is

$$\begin{aligned}\tau_a &:= \inf\{t > 0, \int_0^{\tau_a} N^a(dt) > 0\}\\ \tau_b &:= \inf\{t > 0, \int_0^{\tau_b} N^b(dt) > 0\}\end{aligned} \tag{60}$$

Let $\tau := \tau_a \wedge \tau_b$ denote the first arrival time of an RFQ received by all market makers simultaneously. From assumption on independence of ask and bid RFQs, τ_a and τ_b are independent random variables. We state a lemma on the probability distribution of τ_a and τ_b.

Lemma 6.1. *τ_a, τ_b are independent variables of exponential distribution with parameters λ_a and λ_b, respectively. Moreover,*

$$\mathbb{E}\left[\int_0^{\tau_a\wedge\tau_b} e^{-rt}dt\right] = \frac{1}{r+\lambda_a+\lambda_b}, \quad \mathbb{E}[e^{-r\tau_a}|\tau_a < \tau_b] = \mathbb{E}[e^{-r\tau_b}|\tau_b < \tau_a] = \frac{\lambda_a+\lambda_b}{r+\lambda_a+\lambda_b} \tag{61}$$

The proof of Lemma 6.1 follows easily from fundamental calculation with exponentially distributed random variables.

By Dynamic Programming Principle, we obtain

$$V_i^\delta(q_i) = \mathbb{E}_i\left[-\int_0^\tau e^{-rt}\psi_i(q_i)dt + \int_0^\tau e^{-rt}\left(\delta_{q_t^i}^{a,i}N^{a,i}(dt) + \delta_{q_t^i}^{b,i}N^{b,i}(dt)\right) + e^{-r\tau}V_i^\delta(q_\tau^i)\right] \tag{62}$$

We hereby introduce random events R_a^i, R_b^i indicating whether market maker i wins the ask/bid RFQ.

$$\begin{aligned}R_a^i &:= \{\text{Market maker } i \text{ wins the ask RFQ}\}\\ R_b^i &:= \{\text{Market maker } i \text{ wins the bid RFQ}\}\end{aligned} \tag{63}$$

Upon arrival of an RFQ, market maker i only profits if she wins the trade, that is when R_a^i or R_b^i take place. Hence we discuss different possible values for q_τ^i, with inventory risk limit taken into consideration.

If $-H_i < q_i < H_i$,

$$q_\tau^i(\omega) = \begin{cases} q_i - \Delta, & \text{if } \omega \in R_a^i \\ q_i + \Delta, & \text{if } \omega \in R_b^i \\ q_i, & \text{if } \omega \in (R_a^i)^c \cap (R_b^i)^c \end{cases} \tag{64}$$

If $q_i = -H_i$,

$$q_\tau^i(\omega) = \begin{cases} -H_i + \Delta, & \text{if } \omega \in R_b^i \\ -H_i, & \text{if } \omega \in (R_b^i)^c \end{cases} \tag{65}$$

If $q_i = H_i$,

$$q_\tau^i(\omega) = \begin{cases} H_i - \Delta, & \text{if } \omega \in R_a^i \\ H_i, & \text{if } \omega \in (R_a^i)^c \end{cases} \tag{66}$$

We have an important proposition relating $V_i^\delta(q_i)$ in terms of $\tau_a, \tau_b, R_a^i, R_b^i$. The proof of Proposition 6.2 is stated in Appendix B.

Proposition 6.2. *The state-action value function $V_i^\delta(q_i)$ satisfies*

$$\begin{aligned} V_i^\delta(q_i) = & -\frac{\psi_i(q_i)}{r + \lambda_a + \lambda_b} \\ & + \mathbb{P}(\tau_a < \tau_b)\mathbb{E}_i\left[\mathbb{I}(R_a^i)\left(e^{-r\tau_a}\delta_{q_i}^{a,i}\Delta + e^{-r\tau_a}V_i^\delta(q_i - \Delta)\right)\mathbb{I}(-H_i < q_i \le H_i) + \mathbb{I}((R_a^i)^c)e^{-r\tau_a}V_i^\delta(q_i)\middle|\tau_a < \tau_b\right] \\ & + \mathbb{P}(\tau_b < \tau_a)\mathbb{E}_i\left[\mathbb{I}(R_b^i)\left(e^{-r\tau_b}\delta_{q_i}^{a,i}\Delta + e^{-r\tau_b}V_i^\delta(q_i + \Delta)\right)\mathbb{I}(-H_i \le q_i < H_i) + \mathbb{I}((R_b^i)^c)e^{-r\tau_b}V_i^\delta(q_i)\middle|\tau_b < \tau_a\right] \end{aligned} \tag{67}$$

The Bellman equation (67) is expressed in terms of conditional probabilities. The arrival of RFQ is incorporated into the simulation of market environment, hence it is exogenous to the market makers. The market makers modeled as learning agents receive the RFQ and improve their strategies through interaction with simulated market environment. Assuming that ask and bid RFQs are independent, we have $\mathbb{P}(\tau_a < \tau_b) = \frac{\lambda_a}{\lambda_a+\lambda_b}$ and $\mathbb{P}(\tau_b < \tau_a) = \frac{\lambda_b}{\lambda_a+\lambda_b}$. We consider every iteration of RL learning algorithm as an arrival of an RFQ with ask and bid requests at probabilities $\frac{\lambda_a}{\lambda_a+\lambda_b}$ and $\frac{\lambda_b}{\lambda_a+\lambda_b}$. Upon each arrival of RFQ, the probability that market maker i fulfills the transaction is proportional to market share f_a^i or f_b^i depending on sides of the RFQ.

Recall that the state space of market maker i consists of all of her possible inventory values $\mathcal{Q}_i = \{-H_i, -H_i + \Delta, \ldots, H_i - \Delta, H_i\}$. We consider discretized time steps $t = 0, 1, \ldots$, as the arrival of RFQs. Given inventory state $q_t^i \in \mathcal{Q}_i$ at t, market maker i sets up her ask and bid quotes $(\delta_{q_t^i}^{a,i}, \delta_{q_t^i}^{b,i}) \in I_\delta \times I_\delta$. The quoting strategies $(\delta_{q_t^i}^{a,i}, \delta_{q_t^i}^{b,i})$ are alternatively denoted as functions $\delta_a^i : q \in \mathcal{Q}_i \to \mathbb{R}, \delta_b^i : q \in \mathcal{Q}_i \to \mathbb{R}$. This function representation transmits quoting strategies to neural network approximation, which will be introduced in Section 6.2.

The dealer market environment generates an ask or bid RFQ at each unit time step with probability $\frac{\lambda_a}{\lambda_a+\lambda_b}$ and $\frac{\lambda_b}{\lambda_a+\lambda_b}$.[5] The RFQ is sent to N market makers simultaneously. At time step t market maker i will set up centered ask quote $\delta_a^i = \pi_a^i(q_t^i|\theta_i^\pi)$ and centered bid quote $\delta_b^i = \pi_b^i(q_t^i|\theta_i^\pi)$. The market maker that executes the RFQ order is selected stochastically by market environment. Recall the random events R_a^i, R_b^i defined in (63). $\{R_a^i, i \in \{1, \dots, N\}\}$ form a collection of pairwise mutually exclusive events. This is the same case for $\{R_b^i, i \in \{1, \dots, N\}\}$. We have following conditional probability for events R_a^i, R_b^i.

$$\mathbb{P}(R_a^i|\tau_a < \tau_b) = \frac{f_a^i(\delta_a^i(q_t^i), \cdot)}{\sum_{j=1}^N f_a^j(\delta_a^j(q_t^j), \cdot)} \quad \text{(if RFQ is on ask side)} \tag{68}$$

$$\mathbb{P}(R_b^i|\tau_b < \tau_a) = \frac{f_b^i(\delta_b^i(q_t^i), \cdot)}{\sum_{j=1}^N f_b^j(\delta_b^j(q_t^j), \cdot)} \quad \text{(if RFQ is on bid side)}a \tag{69}$$

Again for simplicity in notations the dependence of f_a^i, f_b^i on competitors' quoting strategies are replaced by symbol "$\cdot$" in (68). We apply this probability distribution based on Assumption 2.1 with $\sum_{j=1}^N f_a^i \le 1$, $\sum_{j=1}^N f_b^i \le 1$. We do not consider the circumstance where no market maker wins the RFQ, which has probability $1 - \sum_{j=1}^N f_a^i$ for an ask RFQ and $1 - \sum_{j=1}^N f_b^i$ for a bid RFQ. Hence we normalize the execution probabilities in (68) so that they form probability distributions of random choice. Note that we hereby make a simplification that if market maker i wins the trade and her inventory is at the risk limit $\pm H_i$, then i will not execute the trade that would drive her inventory out of the risk limit. For instance when i has inventory $-H_i$ she will not quote for an ask RFQ. In this case the market environment will switch to next time step that generates a new RFQ and select a new market maker stochastically for execution. This simplification is also reflected from conditional probabilities in equation (67).

Based on equation (67), we now define reward functions in market environment. All market makers will have to pay the expected running cost

$$\mathbb{E}_i\left[-\int_0^{\tau_a \wedge \tau_b} e^{-rt} dt\right] \psi_i(q_i) = -\frac{\psi_i(q_i)}{r + \lambda_a + \lambda_b}$$

where q_i is inventory for market maker indexed by $i \in \{1, \dots, N\}$.

It is clear that at time t only the market maker who wins the RFQ receives the revenue from making the spread. Hence market maker i's reward function consists of inventory cost she has paid plus the time-discounted revenue from making the spread had she won the corresponding RFQ. Considering the risk limit and our simplification, we write formally our reward function $r_i(q_i, (\delta_a^i, \delta_b^i))$ for market maker i.

$$r_i(q_i, (\delta_a^i, \delta_b^i)) = -\frac{\psi_i(q_i)}{r + \lambda_a + \lambda_b} + \frac{(\lambda_a + \lambda_b)\Delta}{r + \lambda_a + \lambda_b}\left(\mathbb{I}(R_a^i)\mathbb{I}(-H_i < q_i \le H_i) \cdot \delta_a^i + \mathbb{I}(R_b^i)\mathbb{I}(-H_i \le q_i < H_i) \cdot \delta_b^i\right) \tag{70}$$

As far as time steps are concerned in RL simulation, we shall include t as subscript in state q and action δ, so that at time step $t = 0, 1, ...$, market maker i disposes inventory level q_t^i and set action δ_t^i. This notation should not cause ambiguity between subscripts used for inventory and

quotes in Section 2 to 3. Between time step t and $t+1$ only the inventory of the market maker winning the trade will possibly change depending on whether she reaches inventory risk limit. The transition of market maker i's inventory q_t^i to q_{t+1}^i can be summarized depending on side of RFQ.

$$q_{t+1}^i = \begin{cases} q_t^i - \Delta \mathbb{I}(R_a^i)\mathbb{I}(q_t^i > -H_i) & \text{for RFQ from ask side} \\ q_t^i + \Delta \mathbb{I}(R_b^i)\mathbb{I}(q_t^i < H_i) & \text{for RFQ from bid side} \end{cases} \tag{71}$$

(68)-(71) define a Partially Observed Markov Decision Process (POMDP). Define the discount factor $\gamma = \frac{\lambda_a+\lambda_b}{r+\lambda_a+\lambda_b}$, with strong Markov property we have the following proposition that the objective function of market maker i can be written as a discrete-time format.

Proposition 6.3. *The state-action value function $V_i^\delta(q_i)$ in (59) can be written in the following format:*

$$V_i^\delta(q_i) = \mathbb{E}_i\left[\sum_{t=0}^{\infty} \gamma^t r_i\left(q_t^i, (\delta_a^i(q_t^i), \delta_b^i(q_t^i))\right) \middle| q_0^i = q_i\right] \tag{72}$$

where $\gamma = \frac{\lambda_a+\lambda_b}{r+\lambda_a+\lambda_b}$ and reward functions $r_i(q_i, (\delta_a^i, \delta_b^i))$ are defined in (70).

The objective of market maker i is to find optimal quoting strategy $\delta_a^{i,*}, \delta_b^{i,*}$, which maximizes the expected future rewards discounted by a factor $\gamma = \frac{\lambda_a+\lambda_b}{r+\lambda_a+\lambda_b}$.

$$V_i(q_i) = \max_{\delta_a^i, \delta_b^i} \mathbb{E}_i\left[\sum_{t=0}^{\infty} \gamma^t r_i\left(q_t^i, (\delta_a^i(q_t^i), \delta_b^i(q_t^i))\right) \middle| q_0^i = q_i\right] \tag{73}$$

6.2 | The multi-agent DDPG algorithm for market makers

Now we elaborate on more details of our Decentralized Multi-agent DDPG algorithm. Our MADDPG algorithm is a type of actor-critic learning algorithm. The strategies of market maker i, $\delta_a^i : q \in \mathcal{Q}_i \to \mathbb{R}$ and $\delta_b^i : q \in \mathcal{Q}_i \to \mathbb{R}$, are approximated by neural networks $\pi_a^i(q|\theta_i^\pi), \pi_b^i(q|\theta_i^\pi)$. These are called actor networks. Similarly the state-action value functions $V_i^\delta(q)$ of market maker i are approximated by neural networks $Q_i(q, (\delta^a, \delta^b)|\theta_i^Q)$ where q is inventory level, (δ^a, δ^b) are the ask and bid quotes and θ_i^Q is the collection of network parameters. Neural networks $Q_i(q, (\delta^a, \delta^b)|\theta_i^Q)$ are named critic networks, which evaluate a given tuple of state-action combination $(q, (\delta^a, \delta^b))$. The network parameters θ_i^Q and θ_i^π are learned via interactions with the market environment, which sends RFQs to market makers, executes transaction according to market makers' execution probabilities and feedback rewards to each market makers. For critic networks Q_i Temporal Difference (TD) learning is applied for updating critic parameters θ_i^Q, while for actor networks π_a^i, π_b^i their parameters θ_i^π are updated by Stochastic Gradient Descent (SGD) step to minimize a given loss function. For simplicity in notations we omit parameters θ_i^Q and θ_i^π in neural networks Q_i and (π_a^i, π_b^i) when there is no ambiguity. Since quotes or actions (δ_a^i, δ_b^i) come from both ask and bid sides, to simplify notations we sometimes denote the action of market maker i by $\delta^i = (\delta_a^i, \delta_b^i)$.

In addition to primal critic networks Q_i and actor networks δ_a^i, δ_b^i, target critic and actor networks are introduced coupling the critics and actors in implementation. The target networks are denoted by $\tilde{Q}_i(q,(\delta^a,\delta^b)|\tilde{\theta}_i^Q)$ and $\tilde{\pi}_a^i(q|\tilde{\theta}_i^\pi), \tilde{\pi}_b^i(q|\tilde{\theta}_i^\pi)$. The introduction of target network is a common practice in implementation of deep reinforcement learning. It is intended for a more stable parameter update. While primal network parameters θ_i^Q and θ_i^π are updated at every training iteration, the target network parameters $\tilde{\theta}_i^Q$ and $\tilde{\theta}_i^\pi$ are updated slowly to make the training more stationary:

$$\begin{aligned}\tilde{\theta}_i^Q &\leftarrow \mu\theta_i^Q + (1-\mu)\tilde{\theta}_i^Q \\ \tilde{\theta}_i^\pi &\leftarrow \mu\theta_i^\pi + (1-\mu)\tilde{\theta}_i^\pi\end{aligned} \tag{74}$$

where μ represents the speed of updating target network parameters. A value μ close to 0 means very slow transition of parameters obtained from training steps into target network parameters.

At each time step t, denote m_t the side of RFQ, which takes values from $\{a,b\}$ with a for ask RFQ and b for bid RFQ. Denote by $I_t^i \in \{0,1\}$ the indicator function whether market maker i wins the RFQ. $I_t^i = 1$ suggests that market maker i wins the RFQ at time step t. The interaction between market maker i and market environment generates a tuple of data $(q_t^i, \delta_t^i, q_{t+1}^i, r_i(q_t^i,\delta_t^i), I_t^i, m_t)$. In this data tuple q_t^i is market maker i's inventory level at t, δ_t^i refers to market maker i's centered ask and bid quotes given inventory q_t^i, q_{t+1}^i is the inventory level after taking action δ_t^i, and $r_i(q_t^i,\delta_t^i)$ is the reward from environment to market maker i for state-action combination (q_t^i,δ_t^i). This data tuple is stored into an experience replay buffer of corresponding agent. Experience replay is a technique initially proposed in Mnih et al. (2013), widely practiced in reinforcement learning algorithms. The experience replay buffer is used to reduce the correlation between sample data points in online reinforcement learning which could lead to non stationary distribution of learned policy. In fact for parameter tuning, mini-batch data points are sampled from this experience replay buffer, hence unbinding sequential correlation of data samples from interactions with environment. In our experiment, we assign the replay buffer to have a fixed length with first-in-first-out queue structure.

For parameter update, each iteration consists of a Temporal Difference (TD) learning phase and a policy improvement phase. In TD learning phase parameters of critic networks θ_i^Q are updated for every $i \in \{1,\dots,N\}$, after which policy improvement updates actor network parameters θ_i^π. We first implement a mini-batched stochastic gradient descent on following loss function for parameters θ_i^Q. This loss function is based on dynamic programming equations (B.1)-(B.3) with data sampled from experience replay.

$$\begin{aligned}\mathcal{L}_i^Q(\theta_i^Q) = \mathop{\mathbb{E}}_{q_i,\delta^i,q_i',I_i}\Bigg[\Bigg(& r_i(q_i,\delta^i) + \gamma\Bigg(I_i\tilde{Q}_i(q_i',(\tilde{\pi}_a^i(q_i'),\tilde{\pi}_b^i(q_i'))\Big|\tilde{\theta}_i^Q) \\ &+(1-I_i)\tilde{Q}_i(q_i,(\tilde{\pi}_a^i(q_i),\tilde{\pi}_b^i(q_i))\Big|\tilde{\theta}_i^Q\Bigg)\Bigg) \\ &-Q_i(q_i,(\tilde{\pi}_a^i(q_i),\tilde{\pi}_b^i(q_i))|\theta_i^Q)\Bigg)^2\Bigg]\end{aligned} \tag{75}$$

where the pairs (q_i,δ^i,q_i',I_i) are sampled from the experience replay buffer that stores market maker's historical states, actions and transitions during training steps. Reward $r_i(q_i,\delta^i)$ is

calculated depending on ask or bid RFQ, based on (70). The actions $\tilde{\pi}_a^i(q_i'), \tilde{\pi}_b^i(q_i')$ are given by the target actor network of agent i. Then stochastic gradient descent can be applied on (75) to calibrate parameter θ_i^Q. Note that parameters $\tilde{\theta}_i^Q$ of target critic networks $\tilde{Q}_i$ are updated at this stage. The method applied in calibrating θ_i^Q is called Temporal Difference (TD) learning. More specifically we apply a mini-batched stochastic gradient descent. Let K be the size of mini-batch. We sample K data points from experience replay buffer, denoted by $\{(q_i^{(k)}, (\delta^i)^{(k)}, q_i'^{(k)}, I_i^{(k)}), k \in \{1, \dots, K\}\}$ a stochastic gradient descent is implemented with given learning rate α_Q.

$$\theta_i^Q \leftarrow \theta_i^Q + \alpha_Q \frac{1}{K} \sum_{k=1}^{K} \left(\hat{\beta}_k - Q_i\left(q_i^{(k)}, (\tilde{\pi}_a^i(q_i^{(k)}), \tilde{\pi}_b^i(q_i^{(k)}))\right)\right) \nabla_{\theta_i^Q} Q_i\left(q_i^{(k)}, (\tilde{\pi}_a^i(q_i^{(k)}), \tilde{\pi}_b^i(q_i^{(k)}))|\theta_i^Q\right) \tag{76}$$

where

$$\hat{\beta}_k = r_i(q_i^{(k)}, (\delta^i)^{(k)}) + \gamma\left(I_i^{(k)} \tilde{Q}_i\left(q_i'^{(k)}, \left(\tilde{\pi}_a^i(q_i'^{(k)}), \tilde{\pi}_b^i(q_i'^{(k)})\right)\middle|\tilde{\theta}_i^Q\right) + (1 - I_i^{(k)}) \tilde{Q}_i\left(q_i^{(k)}, \left(\tilde{\pi}_a^i(q_i^{(k)}), \tilde{\pi}_b^i(q_i^{(k)})\right)\middle|\tilde{\theta}_i^Q\right)\right) \tag{77}$$

When TD learning is conducted, the policy improvement step updates parameters θ_i^π to improve each market maker's quoting strategy. To calibrate the parameters of actor network θ_i^π, the loss function we use is the value of critic network.

$$\mathcal{L}_i(\theta_i^\pi) = -\mathbb{E}_{q_i}\left[Q_i\left(q_i, (\pi_a^i(q_i|\theta_i^\pi), \pi_b^i(q_i|\theta_i^\pi))\middle|\theta_i^Q\right)\right] \tag{78}$$

we apply the policy gradient for actor networks:

$$\nabla_{\theta_i^\pi} \mathcal{L}_i(\theta_i^\pi) = -\mathbb{E}_{q_i}\Big[\nabla_{\delta_a^i} Q_i\left(q_i, (\pi_a^i(q_i|\theta_i^\pi), \pi_b^i(q_i|\theta_i^\pi))\middle|\theta_i^Q\right) \nabla_{\theta_a^i} \pi_a^i(q_i|\theta_i^\pi) + \nabla_{\delta_b^i} Q_i\left(q_i, (\pi_a^i(q_i|\theta_i^\pi), \delta_b^i(q_i|\theta_i^\pi))\middle|\theta_i^Q\right) \nabla_{\theta_b^i} \pi_b^i(q_i|\theta_i^\pi)\Big] \tag{79}$$

where q_i are sampled from the same mini-batch as in TD learning phase. Stochastic gradient descent is then carried out on parameter θ_i^π with learning rate α_δ.

$$\theta_i^\pi \leftarrow \theta_i^\pi + \alpha_\delta \frac{1}{K} \sum_{k=1}^{K} \Big(\nabla_{\delta_a^i} Q_i\left(q_i^{(k)}, (\pi_a^i(q_i^{(k)}|\theta_i^\pi), \pi_b^i(q_i^{(k)}|\theta_i^\pi))\middle|\theta_i^Q\right) \nabla_{\theta_a^i} \pi_a^i(q_i^{(k)}|\theta_i^\pi) + \nabla_{\delta_b^i} Q_i\left(q_i^{(k)}, (\pi_a^i(q_i^{(k)}|\theta_i^\pi), \pi_b^i(q_i^{(k)}|\theta_i^\pi))\middle|\theta_i^Q\right) \nabla_{\theta_b^i} \pi_b^i(q_i^{(k)}|\theta_i^\pi)\Big) \tag{80}$$

When designing the learning algorithm we also take into consideration the trade-off between exploration and exploitation by introducing a exploration probability that decreases exponentially as function of iteration steps in each episode. Exploration is an essential technique for reinforcement learning to avoid being stuck on local minima. It allows RL agents to deviate from the action given by learned policy according to certain exploration rules. In our context, exploration means that the agents switch to a randomly new action contingently with an exploration probability, instead of picking the action given by actor networks. Exploration allows agents to "experience" more diversified situation to avoid being stuck on limited combinations

Algorithm 2 Decentralized Multi-agent Deep Deterministic Policy Gradient

Input: E = number of episodes, T = number of iteration steps in each episode, B = size of mini-batch. N = number of market makers, f_a^i, f_b^i: execution probabiltiies, λ^a, λ^b: intensity of ask and bid order flow, Δ: unit order size, ψ_i: running cost for holding inventory to market maker i, and hyperparameters for learning algorithm and optimization algorithm.

Output: The target actor networks $(\tilde{\pi}_a^i, \tilde{\pi}_b^i)$ of each market maker i.

Initialization of neural networks:
for i $\leftarrow$ 1 to N **do**
  Pre-train critic network Q_i and actor networks π_a^i, π_b^i to value function and quoting strategy of a single monopolistic market maker with execution probability $\Lambda(\delta)$.
  Let target networks equal to original networks. Namely $\tilde{Q}_i = Q_i$, $\pi_a^i = \tilde{\pi}_a^i$, $\pi_b^i = \tilde{\pi}_b^i$.
end for
for Episode $\leftarrow$ 1 to E **do**
  Initialize the inventory states of market makers, denoted by $(q_0^i)_{i\in\{1,\dots,N\}}$.
  for $t \leftarrow 0$ to $T-1$ **do**
    Market environment generates an ask or bid RFQ with probability $\frac{\lambda_a}{\lambda_a+\lambda_b}$ and $\frac{\lambda_b}{\lambda_a+\lambda_b}$.
    Compute action by target actor networks: $\delta_t^{a,i} = \tilde{\pi}_a^i(q_t^i), \delta_t^{b,i} = \tilde{\pi}_b^i(q_t^i)$. Exploration is considered with probability $p_0 \cdot e^{-\eta t}$.
    Market makers compete for the RFQ. I_t^i denote the indicator whether market maker i wins the RFQ.
    Next inventory level q_{t+1}^i is obtained for each market maker i.
    Data $(q_t^i, \delta_t^i, q_{t+1}^i, I_t^i)$ is stored into market maker i's replay buffer.
    if Replay buffer contains more data points than mini-batch size B **then**
      for i $\leftarrow$ 1 to N **do**
        Carry out mini-batch TD learning for critic parameters with (6.18)-(6.19).
        Carry out mini-batch Stochastic Gradient Descent for actor parameters with (6.22).
        Update target network parameters with (6.16).
      end for
    else
      Move on to next iteration $t+1$.
    end if
  end for
end for

of states and actions. More specifically we add a random Gaussian noise to actions generated by actor networks when the agents are required to explore. At the beginning of each episode, each agent has a probability of $p_0 = 5\%$ to explore on actions. This exploration probability decreases exponentially $p_t = p_0 \cdot e^{-\eta t}$ where η is a constant, t is index of iteration step. In each episode exploration is more often at the beginning with higher probability, and decreases as iteration continues.

Equations (74)–(80) define the scheme of our decentralized Multi-agent Deep Deterministic Policy Gradient (Decentralized MADDPG) algorithm. Both the critics Q_i and actors (π_a^i, π_b^i) are local functions of market maker i's own inventory q_i. This decentralized feature is an essential difference from original DDPG algorithm in Lowe et al. (2017) and Foerster et al. (2018). The interaction between market makers are realized through market shares f_a^i, f_b^i which are implicitly monitored by the market environment. As can be seen from formulation of the algorithm, training of critics and actors network only requires local information by each market maker. Hence our algorithm can provide a simulation for a scenario where all market makers apply automated learning algorithms for setting up ask and bid quotes. The simulation results will be useful for regulators to evaluate the effects of automated learning algorithms in market making. We hereby conclude the algorithm in Algorithm 2.

TABLE 1 RFQ and market makers' parameters for simulation

RFQ arrival rate $\lambda_a = \lambda_b$	Interest rate r	Order size Δ	Running cost $\psi_1(q) = \psi_2(q)$	Risk limit $Q_1 = Q_2$
2	0.01	1	$\psi_i(q) = -0.005q^2$	5

6.3 | Numerical experiments

For implementation we consider the same numerical settings as in Section 5. More specifically there are two market makers, with execution rates defined in (58). Other parameters are present in Table 1

We use fully connected neural networks for both critic Q_i and actor (δ_a^i, δ_b^i). There are three layers in each neural network with 10 hidden units in each layer. We choose Rectified Linear Unit (ReLU) function as activation function for the hidden layers. The activation for the output layer of actor network is a multiplier of tanh function. Specifically if l is the value before activation in output layer, then the final output of actor network is $5 \cdot \tanh(l)$. For critic networks linear activation is applied. We also introduce a layer normalization for each layer before passing layer outputs to activation function. Layer normalization (Ba et al., 2016) is a technique to obtain more stationary distribution when data is passed between layers. It normalizes outputs from all units of a same hidden layer to avoid vanishing gradient or gradient explosion due to extreme outputs while maintaining statistical properties of the values. We find layer normalization important to obtain stationary quotes values given by actor networks. The detailed actor-critic structure of each market maker is shown in Figure 4.

We use a standardized ADAM optimizer (Kingma & Ba, 2015) for both critic and actor networks, with learning rate 0.001, momentum decay rates $(0.9, 0.99)$ and non-zero regularization 10^{-6}. In Reinforcement Learning, an "episode" refers to a complete play of agent interacting with the environment. Within one episode the agent improves her strategy from these interactions through repeated iterations. An "episode" starts from a randomized initial state, and either terminates after certain number of iterations or when the interaction reaches certain termination condition. In our market making context, an "episode" refers to one round of repeated pricing game where market makers compete by posting centered quotes at each iteration, starting from a random initial inventory profile. Each iteration step updates the parameter of neural networks via stochastic gradient descent using data from interactions with environment. Since in dealer market the agents post centered quotes consecutively, there are no explicit termination condition to reach for each episode. Hence we set fixed number of iterations within one episode. The training step takes 500 episodes[6], where in each episode there are 500 iterations. In other words, we set $E = 500, T = 500$ for Algorithm 2. The replay buffers have fixed length of 10,000, and the mini-batch size is 32.

After the training step, we then let the market makers play a repeated pricing game using trained quoting strategies, starting from all possible combinations of initial inventory levels. The market makers quotes consecutively using the trained actor networks in the simulated market environment. This process is similar to training step however without stochastic gradient descent step. The difference between ask/bid quotes given by actor networks and equilibrium ask/bid quotes will be used as a measure for detecting collusion. Excessively higher quotes indicates higher fee charged to clients.

To summarize, we call one round of simulation which consists of following two steps:

- Training the critic and actor networks using Algorithm 2.

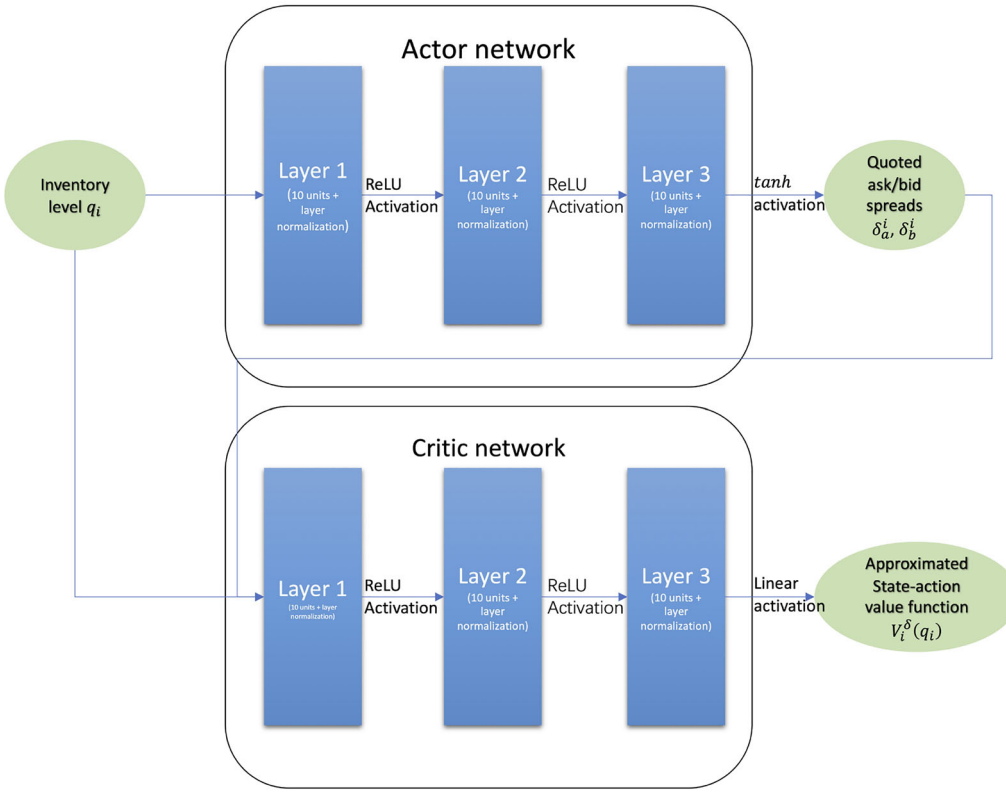


FIGURE 4 Actor-critic networks for market makers. Each hidden layer has 10 neuron units. [Color figure can be viewed at wileyonlinelibrary.com]

- Applying trained quoting strategies in automated pricing game and compare ask/bid quotes with equilibrium quotes.

We repeat 100 independent rounds of simulations to study average behavior of reinforcement learning algorithms applied in dealer market making. Figure 5 shows the average cumulative reward per episode with 95% confidence interval taken over 100 independent simulations. Cumulative reward per episode refers to the sum of rewards from 500 iterations in each episode.

We can see the average cumulative reward per episode increases during the training step, with a sharp increase during the first 10 iteration steps. This indicates that the both market makers learn to set up more profitable ask/bid quotes as training continues. There are oscillations in rewards suggesting some exploration during training. The effectiveness of Decentralized MADDPG is hence demonstrated in that the learning algorithms have indeed learn patterns advantageous to each market maker. We also find the average reward levels are similar between two competing market makers. This is explained by their homogeneity with same parameter values in Table 1, where neither market maker could gain more favorable position than her competitor.

Figure 6 shows the distribution of achieved inventory states by two market makers during the training phase. The distribution of inventory shows that different inventory values are sufficiently visited during training step. The fact that around 0 inventory is visited most frequently shows the effect from running cost function $\psi_i(q_i)$ that penalizes holding non-zero inventory. Hence Figure 6

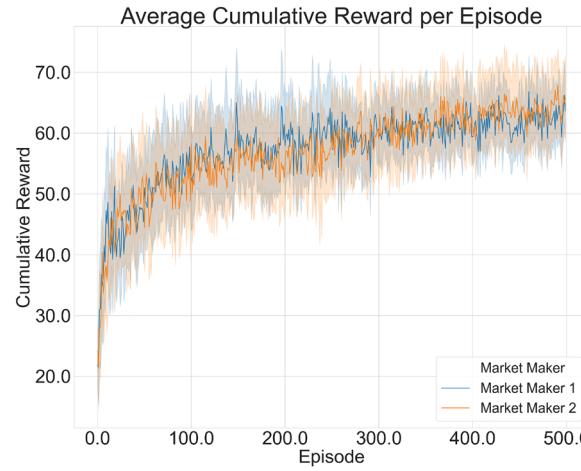


FIGURE 5 Cumulative reward per episode of market makers during training, averaged among 100 simulations with 95% confidence interval. [Color figure can be viewed at wileyonlinelibrary.com]

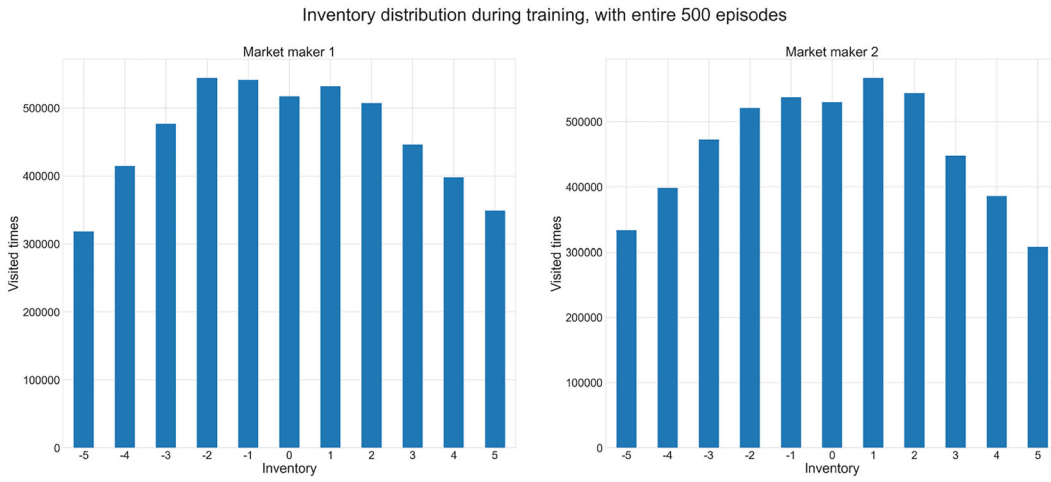


FIGURE 6 Distribution of inventory states for 2 market makers across 500 episodes. [Color figure can be viewed at wileyonlinelibrary.com]

provides evidence that exploration takes effects in our learning algorithm while the statistical property of inventory distribution is still kept.

To understand how distribution of inventory levels evolves during training we separately plot frequency of inventory aggregated from first 250 episodes and that from last 250 episodes in Figures 7 and 8. The bar plots show that as learning continues, both market makers have learnt to keep their inventories more centered around neutral inventory, demonstrated by more concentrated area around 0 inventory in last 250 episodes compared to first 250 episodes. This indicates that learning agents manage to avoid accumulating portfolios through interacting with market directly, without knowing the competition mechanism. The learned quoting strategies by decentralized MADDPG is able to take into consideration by itself the inventory constraints from running cost.

The concentration around 0 inventory presented by the learning algorithms is in line with 0 drift assumption on the asset price dynamic (1). Ganesh et al. (2019) has found that with positive (negative) drift added in price dynamic, market makers' learning algorithm learns the increasing

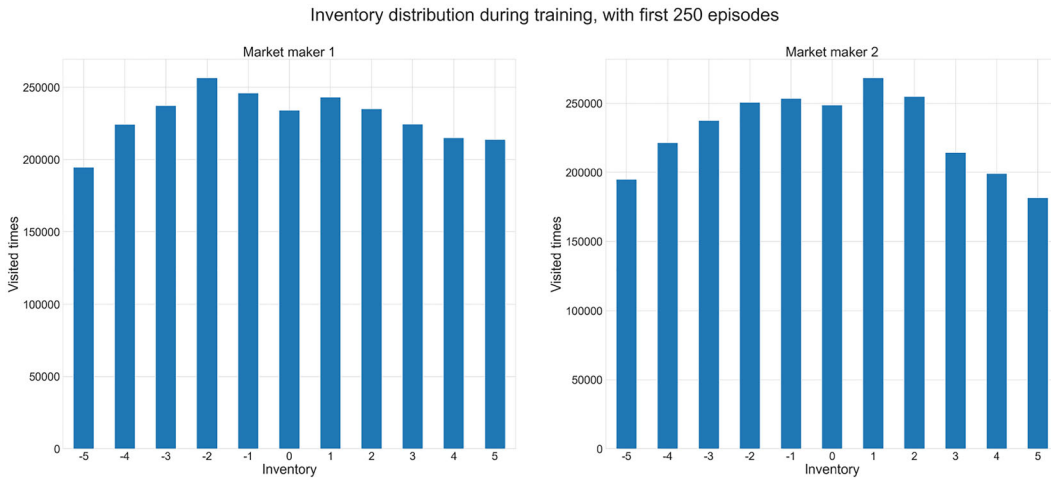


FIGURE 7 Distribution of visited inventory states of two market makers with first 250 episodes. [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 8 Distribution of visited inventory states of two market makers with last 250 episodes. [Color figure can be viewed at wileyonlinelibrary.com]

(decreasing) price trend and maintains positive (negative) inventory to gain profit. It is possible to extend our research with non-zero drift, in which we expect to observe a skewed inventory distribution from 0 given by learning algorithms, or more complicated price dynamics. We leave this topic for future research.

We now study the ask and bid quotes by actor networks after training. We obtain 100 independent quoting strategies by each market makers after 100 simulations. Figure 9 shows the average centered ask and bid quotes of the 100 quoting strategies with 95% confidence interval. The quotes are present by inventory level. The equilibrium quotes and collusive quotes are plotted as benchmark. Note that the explicit collusion ask and bid quotes are approximated by linear functions of inventory, as described in Section 4.

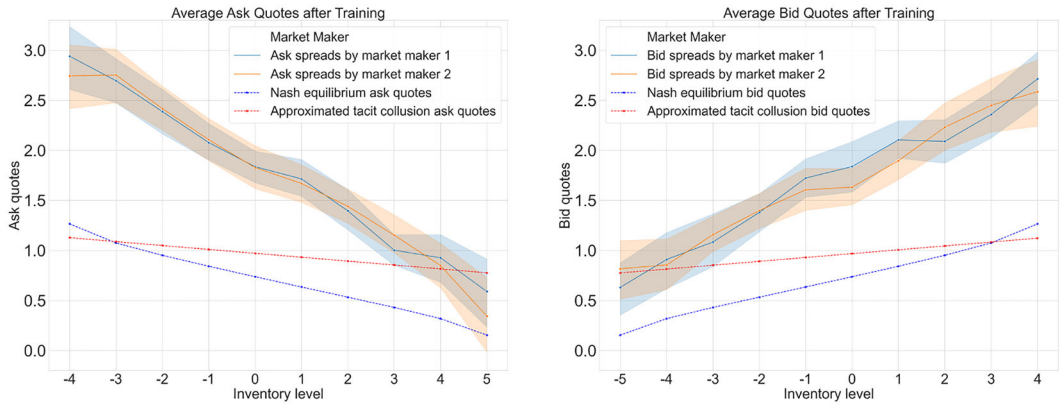


FIGURE 9 Average ask and bid quotes given by trained actor networks from 100 simulations, with 95% confidence interval. Nash equilibrium quotes are plotted in blue dashed lines, while approximated explicit collusion quotes in red dashed lines. The learned quotes are overall higher than competitive quotes in Nash equilibrium, and even higher than the collusion level across most inventory levels. [Color figure can be viewed at wileyonlinelibrary.com]

First of all it is worth noting that Decentralized MADDPG algorithm has learned spread skewing pattern in terms of inventory levels. The learned ask quote is a decreasing function of inventory while learned bid quote is increasing as inventory accumulates. Without knowing the competition mechanism the learning algorithms are successful in learning the pricing pattern close to theoretical results. From Figure 9 we see that algorithms learn to generate higher quotes than Nash equilibrium at all inventory levels. Both market makers will quote higher than equilibrium regardless of inventory based on their learned quoting strategies. We have seen that each market maker's learning algorithm only takes information specific to the market maker without access to her competitors' information at all, meaning that the algorithms are not intended to collude by design. However, both algorithms still learn to quote systematically higher than Nash equilibrium level.

For next step we simulate market making with competition using trained actor networks in the same market environment. This step is also called a "repeated game" in game theory context. In this simulation, we let two market makers start from from 0 inventory at $t = 0$, after which they quote for RFQs sent from simulated market environment, using their trained actor networks. Since we have 100 independent scenarios, we are able to analyze the average behavior in tacit collusion while avoiding bias from certain specific simulation scenario. Figure 10 shows average ask and bid quotes and cumulative profits of 100 trained actor networks at each time step during the repeated game, compared to those of Nash equilibrium and collusion strategies. We see that the quotes achieved by Decentralized MADDPG algorithms are systematically wider than the equilibrium quotes leading to higher cumulative profits earned by the learned strategies. This result is a dynamic version of Figure 9. The stationary ask and bid quotes in Figure 9 are applied in a repeated game and averaged on 100 possible inventory levels at each time step.

To further demonstrate robustness of tacit collusion in terms of initial inventories at $t = 0$, we rerun the above mentioned repeated game with all combinations of initial inventory (q_0^1, q_0^2) and calculate the average basis of 1000 time steps. The basis spreads are calculated by the difference between market makers' ask-bid spread and Nash equilibrium ask-bid spread at every time step. Note the average is imposed both on 1000 time steps and 100 independent trained actor networks. Figure 11 shows a heat map of two market makers' average basis spreads with respect to the Nash

(a) Average ask and bid quotes by trained actor networks in repeated pricing game with 1000 time steps. 0 in vertical axis refers to mid price.

(b) Average cumulative profits by trained actor networks in repeated pricing game with 1000 time steps.

FIGURE 10 Average quotes and cumulative profits of trained strategies. [Color figure can be viewed at wileyonlinelibrary.com]

equilibrium spread. In simulation the risk limit $Q_1 = Q_2 = 5$ with unit order size $\Delta = 1$, hence each market maker has 11 possible inventory levels ranging from -5 to 5. With 2 market makers there are 121 possible combinations of initial inventory levels. We see that the collusive behavior by trained actor networks is significant and robust in terms of initial states.

To examine the impact from number of competing market makers on learning results, we next conduct simulations under same settings but with different number of market makers. The form of execution probabilities depend through (6) on the number of market makers N. We run 100 independent scenarios each with 3, 5, and 10 market makers, and summarize the results in Figures 12 and 13. Figure 12 shows the average cumulative rewards during training episodes. We observe an increasing trend in scenarios with 3, 5, and 10 market makers, suggesting effective learning

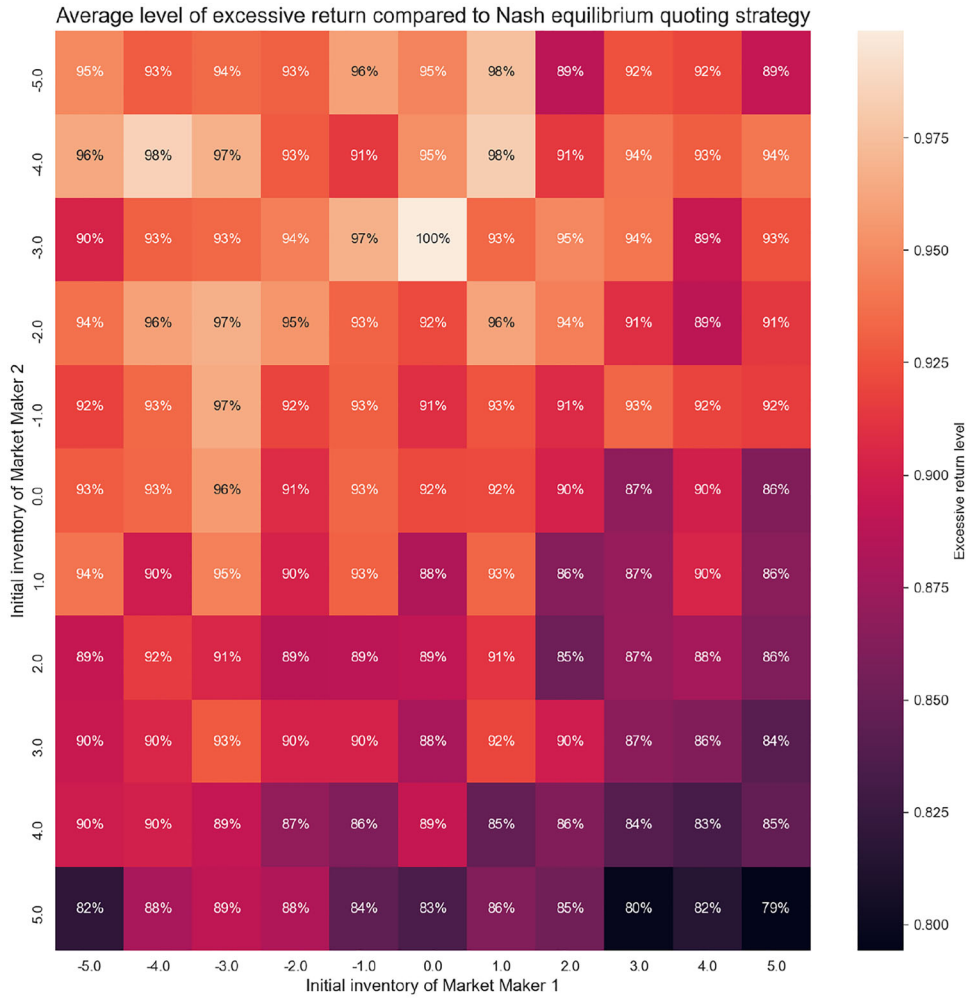
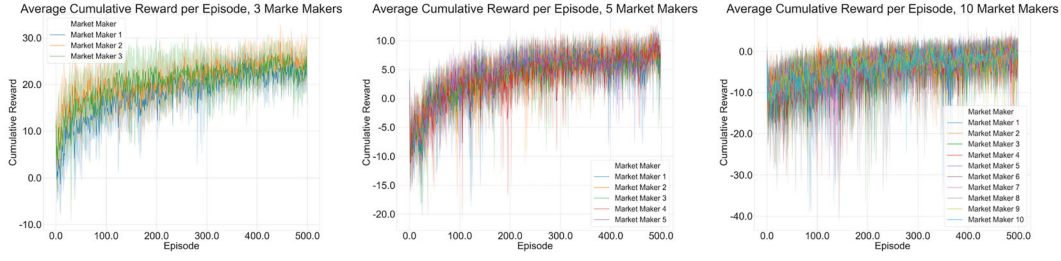


FIGURE 11 Average level of excess return in repeated pricing game with 1000 time steps, across all combinations of initial inventories. [Color figure can be viewed at wileyonlinelibrary.com]



(a) 3 Market Makers (b) 5 Market Makers (c) 10 Market Makers

FIGURE 12 Average cumulative reward per episode during training. [Color figure can be viewed at wileyonlinelibrary.com]

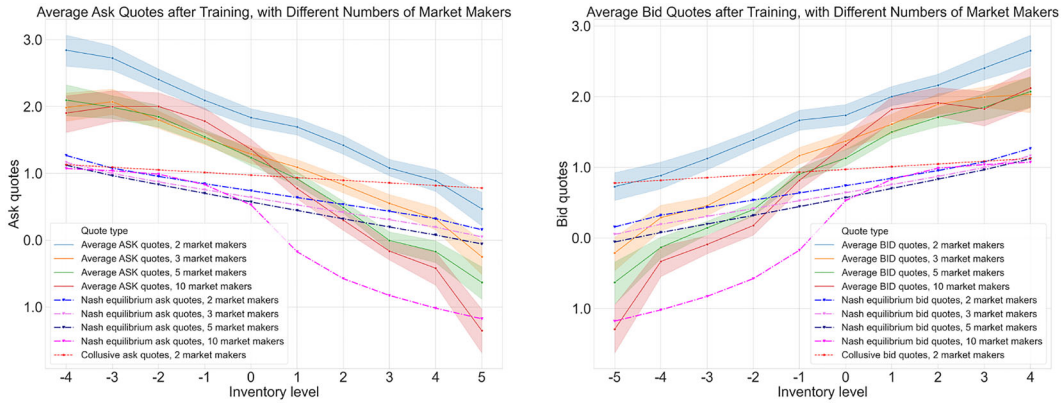


FIGURE 13 Influence of number of market makers on average ask and bid quotes. [Color figure can be viewed at wileyonlinelibrary.com]

under all three scenarios. However with more market makers the increasing trend in cumulative reward slows down, and reward per market maker is significantly lower with more market makers. With 10 market makers, the cumulative rewards are negative in majority of time. This result corresponds to intuition since more market makers induces more intense competition, hence the time for market makers to learn profitable strategies is longer.

Figure 13 presents the learned average ask and bid quotes of 2, 3, 5, and 10 competing market markers compared to corresponding Nash equilibrium quotes. We see that compared to learned ask and bid quotes with two competing market makers, the learned ask and bid quoted prices with 3, 5, and 10 market makers are globally lower at all inventory levels. This suggests that seemingly collusive phenomenon with two market makers are mitigated to some extent with more market makers. This mitigating trend is not apparent when number of market makers increases above 3. We estimate that it is due to insufficient training when there are 5 or 10 market makers, because the rewards in Figure 12 still oscillates drastically below 0 with 500 training episodes. However, the learned quotes are still above Nash equilibrium levels for most of inventory levels with 3, 5, and 10 market makers. It is worth noting that the shape of learned ask and bid quotes resembles those of Nash equilibrium quotes, especially for the case of 10 market makers. This implies that the learning algorithms have replicated the behavior of Nash equilibrium strategy but produces higher quotes leading to "tacit collusion". Another important note is that when number of market makers increases, the learned ask-bid spreads are more skewed to reduce inventory risk. With more market makers, the learning algorithm tends to be more risk averse in that the ask-bid spreads are more skewed when inventory is none zero. For example at inventory level 5, the average ask quotes of 2 market makers is 0.47. This number for 10 competing market makers is −1.35. The negative ask quotes suggests market makers are more eager to sell when their hold large long position, even at a cost of losing money. One possible explanation for this skewing behavior is that with more market makers, the probability that the learning algorithms find profitable quotes is lower hence they are more cautious to avoid holding large non-zero positions and more eager to reduce their exposure to inventory risk at such inventory levels.

To summarize this simulation using decentralized MADDPG algorithm has seen behavior similar to collusion among two competing market makers. The learning algorithm is effective in producing stylized features of ask and bid quotes, but with overall higher level than equilibrium

quotes. This tacit collusion is robust with different initial inventory levels. With more market makers, the learned algorithm tend to produce lower quotes than two market makers, and become more risk averse in that the ask-bid spreads are more skewed when market makers' inventory deviate away from 0.

ACKNOWLEDGMENTS

This research has been supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1). The authors would like to thank Xin Guo for stimulating discussions and valuable feedback. The authors are indebted to the anonymous referees and Associate Editor Renyuan Xu for their very insightful comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Wei Xiong https://orcid.org/0000-0002-9640-3650

ENDNOTES

[1] Note that the constant order size assumption could be relaxed with an order size distribution. In this case the system of coupled Hamilton–Jacobi equations will be incorporated with an integral term over order size space. Bergault and Guéant (2021) and Barzykin et al. (2023) study such extension and prove existence and uniqueness of either a classical and viscosity solution to HJB equation. But for our work the extension will impose challenges on numerical simulation and the design of reinforcement learning algorithm, with higher approximation error due to numerical integral and much longer time for learning algorithm to explore on order size space to converge. We shall leave this extension for future research and focus on current constant order size setting throughout the paper.

[2] Note that notation $\vec{\delta}^{-i}$ is exclusively used in intensity functions f_a^i and f_b^i. They represent the general variables related to others' quoting strategies , that is, $\vec{\delta}^{-i} \in \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$.

[3] The lower bound $-\delta_\infty$ is not only a practical concern but is also mathematically necessary for constraining the upper bound of intensity function as we will see in Assumption 2.1. This boundedness will be applied to validate Itô's formula on objective function in (13). Without the lower bound δ_∞ it would be challenging to validate the application of Itô's formula.

[4] Although notation $V_i^\delta(q_i)$ and $J_i(\boldsymbol{\delta}^i, \boldsymbol{\delta}^{-i}; q_i)$ represent the same quantity, we will more frequently use $V_i^\delta(q_i)$ to emphasize the functional dependence of objective functions on q_i, which also benefits better formatting the linear Bellman equation and algorithm description subsequently.

[5] This is a simplification on arrival of order flows for tractability in Reinforcement Learning simulation. The arrival time τ_a, τ_b of RFQs follows exponential distributions. Instead of simulating directly τ_a, τ_b we consider the RFQ arrivals in sense of probabilistic expectation, assuming there is an RFQ arrival at each time step in RL simulation, which is more tractable.

[6] We have also experimented training with 1000 episodes and find out that average cumulative reward curves exhibit relatively flat trend after first 500 episodes. The value of learned quotes after 1000 episodes of training are at the same level as those after 500 episodes presented in Figure 9, implying numerical convergence of learning algorithm. This comparison hence validates our choice of using 500 episodes.

REFERENCES

Abada, I., & Lambin, X. (2023). Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided? *Management Science*. https://doi.org/10.1287/mnsc.2022.4623

Ardon, L., Vadori, N., Spooner, T., Xu, M., Vann, J., & Ganesh, S. (2021). Towards a fully RL-based Market Simulator. Technical Report 1. https://doi.org/10.1145/3490354.3494372arXiv:2110.06829

Asker, J., Fershtman, C., & Pakes, A. (2021). Artificial intelligence and pricing: The impact of algorithm design. *National Bureau of Economic Research Working Paper Series*, *No. 28535*. http://www.nber.org/papers/w28535%0Ahttp://www.nber.org/papers/w28535.pdf

Assad, S., Calvano, E., Calzolari, G., Clark, R., Denicolò, V., Ershov, D., Johnson, J., Pastorello, S., Rhodes, A., Xu, L., & Wildenbeest, M. (2021). Autonomous algorithmic collusion: economic research and policy implications. *Oxford Review of Economic Policy, 37*(3), 459–478. https://doi.org/10.1093/oxrep/grab011

Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, *8*(3), 217–224. https://doi.org/10.1080/14697680701381228

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. https://doi.org/10.48550/ARXIV.1607.06450

Barzykin, A., Bergault, P., & Guéant, O. (2023). Algorithmic market making in dealer markets with hedging and market impact. *Mathematical Finance*, 33(1), 41–79. Portico. https://doi.org/10.1111/mafi.12367

Bergault, P., & Guéant, O. (2021). Size matters for OTC market makers: General results and dimensionality reduction techniques. *Mathematical Finance*, *31*(1), 279–322. https://doi.org/10.1111/mafi.12286 arXiv:1907.01225

Berge, C. (1963). *Topological spaces*. Oliver & Boyd. https://books.google.co.uk/books?id=0QJRAAAAMAAJ

Bertsekas, D. P., & Shreve, S. E. (1978). *Stochastic optimal control: The discrete time case*. Academic Press. http://www.gbv.de/dms/hbz/toc/ht000971801.pdf

Bremaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Advances in Physical Geochemistry. Springer. https://books.google.co.jp/books?id=Pk0_AQAAIAAJ

Brown, G. W. (1949). Some notes on computation of games solutions. *RAND Corporation*, RM–125–PR.

Brown, G. W. (1951). Iterative solution of games by fictitious play. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation*. Wiley.

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, *110*(10), 3267–97. https://doi.org/10.1257/aer.20190623

Cardaliaguet, P., & Hadikhanloo, S. (2017). Learning in mean field games: The fictitious play. *ESAIM - Control, Optimisation and Calculus of Variations*, *23*, 569–591. https://doi.org/10.1051/cocv/2016004

Cartea, Á., Jaimungal, S., & Ricci, J. (2014). Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, *5*(1), 415–444. https://doi.org/10.1137/130911196

Competition & Markets Authority (2021). Algorithms: How they can reduce competition and harm consumers. Technical report.

Cont, R., Guo, X., & Xu, R. (2021). Interbank lending with benchmark rates: Pareto optima for a class of singular control games. *Mathematical Finance*, *31*, 1–32. https://doi.org/10.1111/mafi.12325

Fazel, M., Ge, R., Kakade, S., & Mesbahi, M. (2018). Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning* (pp. 1467–1476). ICML.

Foerster, J. N., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2974–2982.

Galewski, M., & Rădulescu, M. (2018). On a global implicit function theorem for locally Lipschitz maps via non-smooth critical point theory. *Quaestiones Mathematicae*, *41*(4), 515–528. https://doi.org/10.2989/16073606.2017.1391353

Ganesh, S., Vadori, N., Xu, M., Zheng, H., Reddy, P., & Veloso, M. (2019). Reinforcement Learning for Market Making in a Multi-agent Dealer Market. (NeurIPS). http://arxiv.org/abs/1911.05892

Guéant, O. (2017). Optimal market making. *Applied Mathematical Finance*, *24*(2), 112–154. https://doi.org/10.1080/1350486X.2017.1342552

Guéant, O., Lehalle, C. A., & Fernandez-Tapia, J. (2013). Dealing with the inventory risk: A solution to the market making problem. *Mathematics and Financial Economics*, *7*(4), 477–507. https://doi.org/10.1007/s11579-012-0087-0

Guéant, O., & Manziuk, I. (2019). Deep reinforcement learning for market making in corporate bonds: Beating the curse of dimensionality. *Applied Mathematical Finance*, *26*(5), 387–452. https://doi.org/10.1080/1350486X.2020.1714455

Guéant, O., & Manziuk, I. (2020). Optimal control on graphs: Existence, uniqueness, and long-term behavior. *ESAIM – Control, Optimisation and Calculus of Variations*, *26*, 1–14. https://doi.org/10.1051/cocv/2019071 arXiv:1902.08926

Guo, X., & Xu, R. (2019). Stochastic games for fuel follower problem: N versus mean field game. *SIAM Journal on Control and Optimization*, *57*(1), 659–692. https://doi.org/10.1137/17M1159531

Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. *Mathematical Finance*, *1-67*. https://doi.org/10.1111/mafi.12382

Han, B. (2021). Understanding algorithmic collusion with experience replay. http://arxiv.org/abs/2102.09139

Han, J., & Hu, R. (2020). Deep fictitious play for finding Markovian Nash Equilibrium in multi-agent games. *Proceedings of Machine Learning Research*, *107*, 221–245.

Han, J., Hu, R., & Long, J. (2022). Convergence of deep fictitious play for stochastic differential games. *Frontiers of Mathematical Finance*, *1*, 287. https://doi.org/10.3934/fmf.2021011

Hettich, M. (2021). Algorithmic collusion: Insights from deep learning. *SSRN Electronic Journal*, 1–19. https://doi.org/10.2139/ssrn.3785966

Ho, T., & Stoll, H. R. (1980). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, *8*(1), 47–73. https://doi.org/10.1016/0304-405X(81)90020-9

Ho, T. S. Y., & Stoll, H. R. (1983). The Dynamics of Dealer Markets Under Competition. *The Journal of Finance*, *38*(4), 1053. https://doi.org/10.2307/2328011

Hu, R. (2021). Deep fictitious play for stochastic differential games. *Communications in Mathematical Sciences*, *19*, 325–353. https://doi.org/10.4310/CMS.2021.v19.n2.a2

Ivaldi, M., Jullien, B., Rey, P., Seabright, P., & Tirole, J. (2003). The economics of Tacit Collusion. *IDEI Working Papers, Institut d'Économie Industrielle (IDEI), Toulouse*, *186*(March). https://econpapers.repec.org/RePEc:ide:wpaper:581

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Pérolat, J., Silver, D., & Graepel, T. (2017). A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 2017-Decem (pp. 4191–4204).

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems*, *2017-Decem*, 6380–6391.

Luo, J., & Zheng, H. (2021). Dynamic Equilibrium of Market Making with Price Competition. *Dynamic Games and Applications*, *11*(3), 556–579. https://doi.org/10.1007/s13235-020-00373-w

Cartea, Á., Chang, P., Mroczka, M., & Oomen, R. (2022). Ai-driven liquidity provision in otc financial markets. *Quantitative Finance*, 1–34. https://doi.org/10.1080/14697688.2022.2130087

Cartea, Á., Chang, P., & Penalva, J. (2022). Algorithmic collusion in electronic markets: The impact of tick size. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4105954

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013. http://arxiv.org/abs/1312.5602

Monderer, D., & Shapley, L. S. (1996). Fictitious play property for games with identical interests. *Journal of Economic Theory*, *68*, 258–265. https://doi.org/10.1006/jeth.1996.0014

Perrin, S., Perolat, J., Laurière, M., Geist, M., Elie, R., & Pietquin, O. (2020). Fictitious play for mean field games: Continuous time analysis and applications. *Advances in Neural Information Processing Systems*, *2020-Decem*, 13199-13213.

Robinson, J. (1951). An iterative method of solving a game. *Annals of Mathematics*, *54*, 296–301.

Roughgarden, T. (2016). Best-response dynamics. In *Twenty lectures on algorithmic game theory* (pp. 216–229). Cambridge University Press. https://doi.org/10.1017/CBO9781316779309.017

Shapley, L. S. (1962). *On the Nonconvergence of Fictitious Play*. RAND Corporation. https://doi.org/10.7249/RM3026

Skrzypacz, A., & Hopenhayn, H. (2004). Tacit collusion in repeated auctions. *Journal of Economic Theory*, *114*(1), 153–169.

Tirole, J. (1988). *The theory of industrial organization*. MIT Press.

Waltman, L., & Kaymak, U. (2008). Q-learning agents in a Cournot oligopoly model. *Journal of Economic Dynamics and Control*, *32*(10), 3275–3293. https://doi.org/10.1016/j.jedc.2008.01.003

Xiong, W., & Cont, R. (2021). Interactions of Market Making algorithms : A study on perceived collusion. In *ICAIF '21: Proceedings of the Second ACM International Conference on AI in Finance* (pp. Article No.: 32, Pages 1–9). Association for Computing Machinery. https://doi.org/10.1145/3490354.3494397

How to cite this article: Cont, R., & Xiong, W. (2024). Dynamics of market making algorithms in dealer markets: Learning and tacit collusion. *Mathematical Finance*, *34,* 467–521. https://doi.org/10.1111/mafi.12401

APPENDIX A: PROOF OF THEOREM 3.6

From Proposition 3.4 and verification Theorem 3.5, to prove Theorem 3.6 it suffices to prove the system of HJB equation (28) admits a solution.

We define $H^i_{q_i}(\delta) := (\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})$, and give a lemma on the properties of maximum point of $H^i_{q_i}$.

Lemma A.1. *Suppose that Assumptions 2.1 and 2.3 are satisfied by intensity functions f^i_a, then for given $p^i_{q_i} \in \mathbb{R}$ and $(\vec{\delta}^{a,j})_{j\neq i} \in \prod_{j\neq i} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, function $H^i_{q_i}(\delta) := (\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})$ has a unique maximum point δ^* on $\mathbb{R}$. The maximum point δ^* satisfies*

$$\frac{\partial H^i_{q_i}}{\partial \delta}(\delta^*) = 0 \tag{A.1}$$

Proof. Using notations in (4), we have

$$\begin{aligned}(H^i_{q_i})'(\delta) &= (\delta - p^i_{q_i})\partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) + f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) \\ &= \partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) \left[\delta - p^i_{q_i} + \frac{f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})}{\partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})}\right]\end{aligned} \tag{A.2}$$

Define

$$h(\delta) = \delta - p^i_{q_i} + \frac{f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})}{\partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})}$$

Then

$$h'(\delta) = 2 - \frac{f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})\partial^2_{ii} f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})2}{\left(\partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})\right)^2} > 0$$

from Assumption 2.3. Hence $h(\delta)$ is an increasing function in δ. Again from Assumption 2.3, $\partial_1 f^i_a < 0$, $(H^i_{q_i})'(\delta)$ is a decreasing function in δ. When $\delta < p^i_{q_i}$ we must have $h(\delta) < 0$ hence $(H^i_{q_i})'(\delta) > 0$. We then prove there exists $\delta^* > p^i_{q_i}$ such that $(H^i_{q_i})'(\delta^*) = 0$ by contradiction.

First there exists $\tilde{\delta} > p^i_{q_i}$ such that $H^i_{q_i}(\tilde{\delta}) > 0$. Assume that for any δ, $(H^i_{q_i})'(\delta) > 0$. Then $H^i_{q_i}(\delta)$ is a strictly increasing function on $\mathbb{R}$. $\forall \delta \geq \tilde{\delta}$, $H^i_{q_i}(\delta) > H^i_{q_i}(\tilde{\delta}) > 0$. This contradicts with the fact that

$$\lim_{\delta\to\infty} H^i_{q_i}(\delta) = 0 \tag{A.3}$$

(A.3) is obtained from Assumption 2.1 that $0 < f_a^i(\delta, (\vec{\delta}^{a,j})) < \Lambda(\delta)$ and that $\lim_{\delta\to\infty} \delta\Lambda(\delta) = \lim_{\delta\to\infty} \Lambda(\delta) = 0$.

Therefore, there exists δ^* such that $(H_{q_i}^i)'(\delta^*) = 0$. When $\delta < \delta^*$, $(H_{q_i}^i)'(\delta) > 0$. When $\delta > \delta^*$, $(H_{q_i}^i)'(\delta) < 0$. Hence δ^* is the unique maximum point of $H_{q_i}^i$, and $(H_{q_i}^i)'(\delta) = 0$. □

We consider a family of mappings $\mathcal{T}_{\boldsymbol{p}}^a : \prod_{j=1}^N (I_\delta)^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^N (I_\delta)^{2\frac{H_j}{\Delta}+1}$, indexed by a vector $\boldsymbol{p} = (\vec{p}^1, \dots, \vec{p}^N) \in \prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ where each $\vec{p}^i = (p_q^i)_{q\in\mathcal{Q}_i} \in \mathbb{R}^{2\frac{H_i}{\Delta}+1}$ is indexed by $q \in \mathcal{Q}_i = \{-H_i, -H_i + \Delta, \dots, H_i - \Delta, H_i\}$. $\mathcal{T}_{\boldsymbol{p}}^a$ is such that $\forall \boldsymbol{\delta} = (\vec{\delta}^{a,1}, \dots, \vec{\delta}^{a,N}) \in \prod_{j=1}^N (I_\delta)^{2\frac{H_j}{\Delta}+1}$ where $\vec{\delta}^{a,i} = (\delta_q^{a,i})_{q\in\mathcal{Q}_i} \in (I_\delta)^{2\frac{H_i}{\Delta}+1}$ is a vector in $\mathbb{R}^{2\frac{H_i}{\Delta}+1}$ indexed by $\mathcal{Q}_i$, we have

$$\mathcal{T}_{\boldsymbol{p}}^a(\boldsymbol{\delta}) = \left(\left(\underset{\delta_{q_i}^{a,i} \geq -\delta_\infty}{\arg\max} \left[(\delta_{q_i}^{a,i} - p_{q_i}^i) f_a^i(\delta_{q_i}^{a,i}, (\vec{\delta}^{a,j})_{j\neq i}) \right] \right)_{q_i\in\mathcal{Q}_i} \right)_{i\in\{1,\dots,N\}} \tag{A.4}$$

From Lemma A.1, given $\boldsymbol{p}$, $\underset{\delta\in\mathbb{R}}{\arg\max}\, H_{q_i}^i(\delta)$ is unique. Hence the mapping $\mathcal{T}_{\boldsymbol{p}}^a$ is well-defined.

Remark A.2. Note that in (A.4) arg max is taken on interval $[-\delta_\infty, \infty)$, but $\mathcal{T}_{\boldsymbol{p}}^a$ is still well-defined. Since if the maximum point δ^* of function $H_{q_i}^i(\delta)$ satisfies $\delta^* \leq -\delta_\infty$, then

$$\underset{\delta\geq -\delta_\infty}{\arg\max} (\delta - p_{q_i}^i) f_a^i(\delta, (\vec{\delta}^{a,j})_{j\neq i}) = -\delta_\infty$$

We also define symmetrically the mappings $\mathcal{T}_{\boldsymbol{p}}^b$ for the bid quoting strategy side. From now on we shall focus on $\mathcal{T}_{\boldsymbol{p}}^a$, and the results follows immediately for $\mathcal{T}_{\boldsymbol{p}}^b$.

For any given $\boldsymbol{p} \in \prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, we study the existence of fixed point for $\mathcal{T}_{\boldsymbol{p}}^a$. If $\mathcal{T}_{\boldsymbol{p}}^a$ has a fixed point, we can then take $\boldsymbol{p} = ((\frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta})_{q_i\in\mathcal{Q}_i})_{i\in\{1,\dots,N\}}$ and transfer system of HJB equations (28) into a system of nonlinear equations where the unknown variables are $\{V_i(q_i), q_i \in \mathcal{Q}_i, i \in \{1, \dots, N\}\}$.

Proposition A.3. *For any given* $\boldsymbol{p} \in \prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, *there exists a nonempty compact set* $K_{\boldsymbol{p}} \subseteq \prod_{j=1}^N (I_\delta)^{2\frac{H_j}{\Delta}+1}$ *such that* $\mathcal{T}_{\boldsymbol{p}}^a(K_{\boldsymbol{p}}) \subseteq K_{\boldsymbol{p}}$.

Proof. We prove that for any $\boldsymbol{\delta}$, $\mathcal{T}_{\boldsymbol{p}}^a(\boldsymbol{\delta})$ is uniformly bounded.

Define

$$p_m = \min_{i\in\{1,\dots,N\}, q_i\in\mathcal{Q}_i} p_{q_i}^i, p_M = \max_{i\in\{1,\dots,N\}, q_i\in\mathcal{Q}_i} p_{q_i}^i$$

For given i and q_i denote the coordinate $q_i \in \mathcal{Q}_i$ of the i^{th} vector in $\mathcal{T}^a_{\boldsymbol{p}}(\boldsymbol{\delta})$ by $g^i_{q_i}$, that is,

$$g^i_{q_i} := \mathcal{T}^{a,i,q_i}_{\boldsymbol{p}}(\boldsymbol{\delta}) := \arg\max_{\delta\in\mathbb{R}}(\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) \tag{A.5}$$

From Assumption 2.1, $f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) > 0$, hence when $\delta > p^i_{q_i}$ we have $(\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) > 0$. Therefore, the maximum in (A.5) must be attained on the interval $(p^i_{q_i}, \infty)$. We obtain the lower bound p_m

$$g^i_{q_i} > p^i_{q_i} \geq p_m, \forall i \in \{1, \dots, N\}, q_i \in \mathcal{Q}_i \tag{A.6}$$

Define $\delta_m = \max(p_m, -\delta_\infty)$, then we have

$$\arg\max_{\delta\geq -\delta_\infty}(\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) \geq \delta_m \tag{A.7}$$

On the other hand, for any $\boldsymbol{\delta}$ such that $\delta^{a,j}_{q_j} \geq \delta_m$ where $q_j \in \mathcal{Q}_j, j \in \{1, \dots, N\}$, we seek an upper bound for $g^i_{q_i}$ where i and q_i are arbitrary. Replace the coordinate $\delta^{a,i}_{q_i}$ inside $\boldsymbol{\delta}$ by $\hat{\delta} \equiv p_M + 1$ and form a new vector $\hat{\boldsymbol{\delta}} \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$. We then consider the value $G^i_{q_i}$ defined by

$$G^i_{q_i} = (\hat{\delta} - p^i_{q_i}) f^i_a(\hat{\delta}, (\vec{\delta}^{a,j})_{j\neq i}) \tag{A.8}$$

From Assumption 2.3, f^i_a is increasing function in $\delta^j_{q_j}, \forall q_j \in \mathcal{Q}_j, \forall j \neq i$, and $\delta^j_{q_j} \geq p_m, \forall q_j \in \mathcal{Q}_j, \forall j \neq i$, we have

$$G^i_{q_i} = (\hat{\delta} - p^i_{q_i}) f^i_a(\hat{\delta}, (\vec{\delta}^{a,j})_{j\neq i}) \geq (\hat{\delta} - p^i_{q_i}) f^i_a(\hat{\delta}, (p_m)_{q_j\in\mathcal{Q}_j, j\neq i}) \geq f^i_a(\hat{\delta}, (p_m)_{q_j\in\mathcal{Q}_j, j\neq i}) \tag{A.9}$$

From Assumption 2.1, the upper bound function Λ of f^i_a satisfies $\lim_{\delta\to\infty}\Lambda(\delta)\delta = 0$. We can also derive $\lim_{\delta\to\infty}\Lambda(\delta) = 0$. There exists $\delta_M > \max(p_M + 1, -\delta_\infty)$ such that

$$f^i_a(p_M + 1, (p_m)_{q_j\in\mathcal{Q}_j, j\neq i}) > \max_{q_i\in\mathcal{Q}_i}\Lambda(\delta_M)(\delta_M - p^i_{q_i}) > \max_{q_i\in\mathcal{Q}_i} f^i_a(\delta_M, (\vec{\delta}^{a,j})_{j\neq i})(\delta_M - p^i_{q_i}) \tag{A.10}$$

Combining (A.9) and (A.10) we obtain

$$(\hat{\delta} - p^i_{q_i}) f^i_a(\hat{\delta}, (\vec{\delta}^{a,j})_{j\neq i}) > \max_{q_i\in\mathcal{Q}_i} f^i_a(\delta_M, (\vec{\delta}^{a,j})_{j\neq i})(\delta_M - p^i_{q_i}) \tag{A.11}$$

Hence the maximum point δ^* of function $H^i_{q_i}(\delta) := (\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})$ can only be achieved on interval $[p_m, \delta_M]$, since from Lemma A.1 $H^i_{q_i}$ is increasing function on $[p_m, \delta^*]$ and decreasing function on $[\delta^*, \infty)$. Therefore we have

$$\delta^* \leq \delta_M$$

Since i and q_i are arbitrary we finally obtain

$$g^i_{q_i} \leq \delta_M, \forall q_i \in \mathcal{Q}_i, i \in \{1, \dots, N\} \tag{A.12}$$

Therefore, the compact set $K_{\boldsymbol{p}} = \prod_{j=1}^{N} [\delta_m, \delta_M]^{(2\frac{H_j}{\Delta}+1)} \subseteq \prod_{j=1}^{N} (I_\delta)^{2\frac{H_j}{\Delta}+1}$ satisfies

$$\mathcal{T}_{\boldsymbol{p}}^{a}(K_{\boldsymbol{p}}) \subseteq K_{\boldsymbol{p}}$$

□

We next prove that $\mathcal{T}_{\boldsymbol{p}}^{a}$ is continuous on $K_{\boldsymbol{p}}$. To proceed we first introduce the notions of upper and lower hemicontinuity for set-valued functions (or in other name, correspondence). We denote a correspondence that maps from A to subsets of B by $\Gamma : A \Rightarrow B$ such that $\forall x \in A, \Gamma(x) \subseteq B$.

Definition A.4. (Upper hemicontinuity) A correspondence $\Gamma : A \Rightarrow B$ is upper hemicontinuous at $a \in A$, if for any open neighborhood V of $\Gamma(a)$ (i.e. $\Gamma(a) \subseteq V$, there exists a neighborhood U of a, such that for any $x \in U, \Gamma(x) \subseteq V$).

Definition A.5. (Lower hemicontinuity) A correspondence $\Gamma : A \Rightarrow B$ is lower hemicontinuous at $a \in A$, if for any open set V such that $V \cap \Gamma(a) \neq \emptyset$, there exists a neighborhood U of a, such that for any $x \in U, \Gamma(x) \cap V \neq \emptyset$.

We will need below Berge's Maximum Theorem (Berge, 1963) for the continuity of arg max function.

Lemma A.6. *(Berge's Maximum Theorem) A function $f : X \times \Theta \to \mathbb{R}$ is continuous on $X \times \Theta$. A correspondence $D : \Theta \Rightarrow X$ is compact-valued, i.e. $\forall \theta \in \Theta, D(\theta)$ is a compact subset of X. Define the maximum function $f^*(\theta) = \sup\{f(x,\theta), x \in D(\theta)\}$, and $D^* : \Theta \Rightarrow X$ by $D^*(\theta) = \arg\sup\{f(x,\theta), x \in D(\theta)\} = \{x \in D(\theta) : f(x,\theta) = f^*(\theta)\}$. If D is both upper and lower hemicontinuous at θ, then $f^*(\theta)$ is continuous, and $D^*(\theta)$ is upper hemicontinuous with nonempty and compact values.*

For a single-valued mapping, we have following lemma connecting upper hemicontinuity and the continuity of function.

Lemma A.7. *Let X, Y be 2 topological spaces, and $\Gamma : X \Rightarrow Y$ be single-valued mapping. If Γ is upper hemicontinuous, then Γ is also a continuous as a function $\Gamma : X \to Y$.*

Proof. The proof is straightforward. For $x \in X$, let $V \subseteq Y$ be a open set containing $f(x)$, i.e. $\{f(x)\} \subseteq V$. Since Γ is upper hemicontinuous and V is a neighborhood with $\{f(x)\} \subseteq V$, from Definition A.4 there exists a neighborhood U of x, such that for any $u \in U, \{\Gamma(u)\} \subseteq V$. Since U is a neighborhood of x, there exists an open set O satisfying $x \in O, O \subseteq U$. Moreover, $\forall u \in O, \Gamma(u) \in V$. Hence Γ is continuous in $x \in X, \forall x \in X$. Therefore, $\Gamma : X \to Y$ is a continuous function. □

Proposition A.8. *For any given $\boldsymbol{p} \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, Let $K_{\boldsymbol{p}} \subseteq \prod_{j=1}^{N} (I_\delta)^{2\frac{H_j}{\Delta}+1}$ be the compact set defined in Proposition A.3. Then $\mathcal{T}_{\boldsymbol{p}}^{a} : K_{\boldsymbol{p}} \to K_{\boldsymbol{p}}$ is continuous.*

Proof. It suffices to verify the continuity of $\mathcal{T}_{\boldsymbol{p}}^{a,i,q_i}$ for given index i, q_i, where

$$\mathcal{T}_{\boldsymbol{p}}^{a,i,q_i}(\boldsymbol{\delta}) := \underset{\delta \geq -\delta_\infty}{\arg\max}(\delta - p_{q_i}^i) f_a^i(\delta, (\vec{\delta}^{a,j})_{j\neq i})$$

in other words to prove $\mathcal{T}_{\boldsymbol{p}}^{a,i,q_i}$ is continuous in terms of $(\vec{\delta}^{a,j})_{j\neq i}$. Write function $H_{q_i}^i(\delta) := (\delta - p_{q_i}^i) f_a^i(\delta, (\vec{\delta}^{a,j})_{j\neq i})$. From Lemma A.1, $\arg\max$ of $H_{q_i}^i(\delta)$ exists and is unique for any $i \in \{1, \dots, N\}, q_i \in \mathcal{Q}_i$. Hence $\mathcal{T}_{\boldsymbol{p}}^a$ is well-defined as a single-valued mapping on $K_{\boldsymbol{p}}$.

$H_{q_i}^i$ is continuous in terms of $(\vec{\delta}^{a,j})_{j\neq i}$, and $\arg\max_{\delta} H_{q_i}^i(\delta)$ is taken on a compact set, denoted by $K_{\boldsymbol{p}}^{i,q_i} \subseteq \mathbb{R}$. Hence the conditions in Lemma A.6 are satisfied. In fact, we take $f = H_{q_i}^i$, $X = \mathbb{R}$, $\Theta = \prod_{j\neq i} \mathbb{R}^{2\frac{H_i}{\Delta}+1}$, $x = \delta \in X$, $\theta = (\vec{\delta}^{a,j})_{j\neq i} \in \Theta$. And $D(\theta) \equiv K'$ where K' is the projection of $K_{\boldsymbol{p}}$ on subspace $\prod_{j\neq i} \mathbb{R}^{2\frac{H_i}{\Delta}+1}$. Then $D(\theta) = K'$ is also a compact set. D as a constant mapping is both upper and lower hemicontinuous. Therefore, from Lemma A.6 $\mathcal{T}_{\boldsymbol{p}}^{a,i,q_i}$ is continuous as function of $(\vec{\delta}^{a,j})_{j\neq i}$. Combining all coordinates $q_i \in \mathcal{Q}_i$ and $i \in \{1, \dots, N\}$, we obtain $\mathcal{T}_{\boldsymbol{p}}^{a,i,q_i}$ is a continuous mapping on $K_{\boldsymbol{p}}$. □

Proposition A.9. *For any given* $\boldsymbol{p} \in \prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, *the mapping* $\mathcal{T}_{\boldsymbol{p}}^a : K_{\boldsymbol{p}} \to K_{\boldsymbol{p}}$ *has a fixed point. That is, there exists* $\boldsymbol{\delta}_{\boldsymbol{p}} \in \prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ *such that* $\mathcal{T}_{\boldsymbol{p}}^a(\boldsymbol{\delta}_{\boldsymbol{p}}) = \boldsymbol{\delta}_{\boldsymbol{p}}$.

Proof. From Proposition A.3, $K_{\boldsymbol{p}}$ is a compact set in $\prod_{j=1}^N \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, hence is closed. By construction in proof of Proposition A.3, $K_{\boldsymbol{p}}$ is also a convex set. From Proposition A.8, $\mathcal{T}_{\boldsymbol{p}}^a$ is continuous on $K_{\boldsymbol{p}}$. Then by Schauder's fixed-point theorem, $\mathcal{T}_{\boldsymbol{p}}^a$ has a fixed point in $K_{\boldsymbol{p}}$. □

To proceed we also need the uniqueness of the fixed point $\boldsymbol{\delta}_{\boldsymbol{p}}$ and the continuity of the map $\boldsymbol{p} \mapsto \boldsymbol{\delta}_{\boldsymbol{p}}$. To derive these results we use the following global implicit function theorem (Galewski & Rădulescu, 2018):

Lemma A.10. *Let* $F \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$ *be a* C^1 *mapping which satisfies*

- $\forall y \in \mathbb{R}^m$ *the function* $\phi_y(x)$ *defined by* $\phi_y(x) = \frac{1}{2}||F(x,y)||^2$ *is coercive, i.e.*

$$\lim_{||x||\to\infty} \phi_y(x) = +\infty$$

- *The Jacobian matrix* $\partial_x F(x,y)$ *is non-singular for any* $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$.

Then there exists a unique function $f : \mathbb{R}^m \to \mathbb{R}^n$ *such that* $f \in C^1(\mathbb{R}^m, \mathbb{R}^n)$ *and*

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m, F(x,y) = 0\} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m, x = f(y)\}.$$

Proposition A.11. *For any* $\boldsymbol{p} \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, *the fixed point* $\boldsymbol{\delta}_{\boldsymbol{p}}$ *from Proposition A.9 is unique and the mapping* $\boldsymbol{\delta}_{\boldsymbol{p}} = \boldsymbol{\delta}(\boldsymbol{p}) : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^{N} (I_\delta)^{2\frac{H_j}{\Delta}+1}$ *is continuous in* $\boldsymbol{p}$.

Proof. Given i and $q_i \in \mathcal{Q}_i$, from Lemma A.1, the maximal point of $H^i_{q_i}(\delta) = (\delta - p^i_{q_i}) f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i})$ satisfies first order condition:

$$(\delta - p^i_{q_i})\partial_1 f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) + f^i_a(\delta, (\vec{\delta}^{a,j})_{j\neq i}) = 0 \tag{A.13}$$

Define a mapping $\mathcal{L}^i_{q_i}(\boldsymbol{\delta}, \boldsymbol{p}) : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \times \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \mathbb{R}$.

$$\mathcal{L}^i_{q_i}(\boldsymbol{\delta}, \boldsymbol{p}) = -\frac{f^i_a(\delta^{a,i}_{q_i}, (\vec{\delta}^{a,j})_{j\neq i})}{\partial_1 f^i_a(\delta^{a,i}_{q_i}, (\vec{\delta}^{a,j})_{j\neq i})} - \delta^{a,i}_{q_i} + p^i_{q_i} \tag{A.14}$$

Then define mapping $\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{p}) = ((\mathcal{L}^i_{q_i})_{q_i\in\mathcal{Q}_i})_{i\in\{1,\dots,N\}} : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \times \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$. We then compute the gradient of $\mathcal{L}$.

$$\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_i}} = -2 + \frac{\partial^2_{ii} f^i_a f^i_a}{(\partial_1 f^i_a)^2} \tag{A.15}$$

$$\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_j}} = 0, \forall q_j \neq q_i \tag{A.16}$$

$$\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,j}_{q_j}} = -\frac{(\partial_1 f^i_a)(\partial^j_{q_j} f^i_a) - f^i_a(\partial^j_{q_j}\partial_1 f^i_a)}{(\partial_1 f^i_a)^2}, \forall j \neq i \tag{A.17}$$

Then we have

$$\left|\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_i}}\right| - \sum_{q_j\neq q_i}\left|\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_j}}\right| - \sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\left|\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,j}_{q_j}}\right| = \frac{2(\partial_1 f^i_a)^2 - \partial^2_{ii} f^i_a f^i_a - \sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j}\left|(\partial_1 f^i_a)(\partial^j_{q_j} f^i_a) - f^i_a(\partial^j_{q_j}\partial_1 f^i_a)\right|}{(\partial_1 f^i_a)^2} \tag{A.18}$$

From Assumption 2.3, we have $|\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_i}}| - \sum_{q_j\neq q_i} |\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,i}_{q_j}}| - \sum_{j\neq i}\sum_{q_j\in\mathcal{Q}_j} |\frac{\partial \mathcal{L}^i_{q_i}}{\partial \delta^{a,j}_{q_j}}| > 0$ Hence the Jacobian matrix $\nabla_{\boldsymbol{\delta}}\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{p})$ is diagonally dominant, hence is non-singular. Therefore, $\nabla_{\boldsymbol{\delta}}\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{p})$ is bijective for any $(\boldsymbol{\delta}, \boldsymbol{p}) \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \times \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$.

Now it remains to prove $\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{p})$ is coercive. Given a sequence $\{\boldsymbol{\delta}^{(n)}\} \subseteq \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ such that $||\boldsymbol{\delta}^{(n)}|| \to \infty$, there must exists a subsequence $\{(\delta^{a,i_n}_{q_{i_n}})^{(k_n)}\}$ such that $\lim_{n\to\infty} |(\delta^{a,i_n}_{q_{i_n}})^{(k_n)}| = \infty$. Other-

wise there exists a constant $M > 0$, such that for any $K > 0$ there exists $k > K$ and $|(\delta^i_{q_i})^{(k)}| < M$ holds for any i, q_i. Hence $||\boldsymbol{\delta}^{(k)}|| < M\sqrt{\prod_{i=1}^{N}(2\frac{H_i}{\Delta}+1)}$. This contradicts with $||\boldsymbol{\delta}^{(n)}|| \to \infty$.

When $(\delta^{a,i_n}_{q_{i_n}})^{(k_n)} \to -\infty$, since $-\frac{f^i_a(\delta^{a,i}_{q_i},(\vec{\delta}^{a,j})_{j\neq i})}{\partial_1 f^i_a(\delta^{a,i}_{q_i},(\vec{\delta}^{a,j})_{j\neq i})} > 0$ we have

$$\mathcal{L}^{i_n}_{q_{i_n}}(\boldsymbol{\delta}^{(k_n)},\boldsymbol{p}) = -\frac{f^i_a((\delta^{a,i_n}_{q_{i_n}})^{(k_n)},((\vec{\delta}^{a,j})^{(k_n)})_{j\neq i})}{\partial_1 f^i_a((\delta^{a,i_n}_{q_{i_n}})^{(k_n)},((\vec{\delta}a,j)^{(k_n)})_{j\neq i})} - (\delta^{a,i_n}_{q_{i_n}})^{(k_n)} + p^{i_n}_{q_{i_n}} > -(\delta^{a,i_n}_{q_{i_n}})^{(k_n)} + p^{i_n}_{q_{i_n}} \to \infty \tag{A.19}$$

When $(\delta^{a,i_n}_{q_{i_n}})^{(k_n)} \to +\infty$, from Assumption 2.3 let $Q = \lim_{\delta\to+\infty}\frac{f^i_a(\delta,\cdot)}{\partial_1 f^i_a(\delta,\cdot)} < \infty$ there exists $R > 0$ such that for any $n > R$ we have

$$\begin{aligned}\mathcal{L}^{i_n}_{q_{i_n}}(\boldsymbol{\delta}^{(k_n)},\boldsymbol{p}) &= -\frac{f^i_a((\delta^{a,i_n}_{q_{i_n}})^{(k_n)},((\vec{\delta}^{a,j})^{(k_n)})_{j\neq i})}{\partial_1 f^i_a((\delta^{a,i_n}_{q_{i_n}})^{(k_n)},((\vec{\delta}^{a,j})^{(k_n)})_{j\neq id})} - (\delta^{a,i_n}_{q_{i_n}})^{(k_n)} + p^{i_n}_{q_{i_n}} \\ &\leq -Q + 1 - (\delta^{a,i_n}_{q_{i_n}})^{(k_n)} + p^{i_n}_{q_{i_n}} \to -\infty\end{aligned} \tag{A.20}$$

When there are two subsequences of $(\delta^{a,i_n}_{q_{i_n}})^{(k_n)} \to +\infty$ converging, respectively, to $+\infty$ and $-\infty$ from (A.19) and (A.20) we still have

$$\lim_{||\boldsymbol{\delta}^{k_n}||\to\infty} ||\mathcal{L}(\boldsymbol{\delta}^{k_n},\boldsymbol{p})|| = +\infty$$

Therefore, $\mathcal{L}(\boldsymbol{\delta},\boldsymbol{p})$ satisfies the conditions in Lemma A.10, hence there exists unique mapping $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{p})$ which is C^1 in $\boldsymbol{p}$. Define $\boldsymbol{\delta}_{\boldsymbol{p}} = \max(\boldsymbol{\delta}(\boldsymbol{p}), -\delta_\infty)$ then $\boldsymbol{\delta}_{\boldsymbol{p}}$ is continuous in $\boldsymbol{p}$. □

We can now prove Theorem 3.6.

Proof. (**Theorem** 3.6) Denote $\vec{V} = ((V_i(q_i))_{q_i\in\mathcal{Q}_i})_{i\in\{1,\dots,N\}}$ is an unknown vector that we want to solve in space $\prod_{j=1}^{N}\mathbb{R}^{2\frac{H_j}{\Delta}+1}$. Then we take a specific $\hat{\boldsymbol{p}} \in \prod_{j=1}^{N}\mathbb{R}^{2\frac{H_j}{\Delta}+1}$ such that $\hat{p}^i_{q_i} = \frac{V_i(q_i)-V_i(q_i-\Delta)}{\Delta}$. From Proposition A.9 we can express the fixed point of $\mathcal{T}^a_{\hat{\boldsymbol{p}}}$ as a function of $\hat{\boldsymbol{p}}$, that is, a function of $\vec{V}$. We denote this fixed point by $\boldsymbol{\delta}(\vec{V}) = ((\delta^{a,i}_{q_i}(\vec{V}))_{q_i\in\mathcal{Q}_i})_{i\in\{1,\dots,N\}}$. Note that by Proposition A.11 $\boldsymbol{\delta}(\vec{V})$ is unique given $\vec{V}$, and is continuous in $\vec{V}$.

Equation (28) can be written as

$$\begin{aligned}&rV_i(q_i) + \psi_i(q_i) - \mathbb{I}(q_i > -H_i)\lambda^a\Delta\left[f^i_a\left(\delta^{a,i}_{q_i}(\vec{V})),(\vec{\delta}^{a,j}(\vec{V}))\right)_{j\neq i}\right)\left(\delta^{a,i}_{q_i}(\vec{V})) - \frac{V_i(q_i)-V_i(q_i-\Delta)}{\Delta}\right)\right] \\ &\quad - \mathbb{I}(q_i < H_i)\lambda^b\Delta\left[f^i_b\left(\delta^{b,i}_{q_i}(\vec{V})),(\vec{\delta}^{b,j}(\vec{V}))\right)_{j\neq i}\right)\left(\delta^{b,i}_{q_i}(\vec{V})) - \frac{V_i(q_i)-V_i(q_i+\Delta)}{\Delta}\right)\right] = 0\end{aligned} \tag{A.21}$$

To prove there exists a solution to equation (28), it suffices to prove there exists a vector $\vec{V} = ((V_i(q_i))_{q_i\in\mathcal{Q}_i})_{i\in\{1,\dots,N\}} \in \prod_{j=1}^{N}\mathbb{R}^{2\frac{H_j}{\Delta}+1}$ satisfying the system of nonlinear equations (A.21).

We define $\mathbb{R}$-valued mappings $\mathcal{H}^{a,i}_{q_i}$ and $\mathcal{H}^{b,i}_{q_i}$ defined on $\prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, such that for $\vec{V} \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$,

$$\mathcal{H}^{a,i}_{q_i}(\vec{V}) = \mathbb{I}(q_i > -H_i) f^i_a\left(\delta^{a,i}_{q_i}(\vec{V}), (\vec{\delta}^{a,j}(\vec{V}))j \neq i\right)\left(\delta^{a,i}_{q_i}(\vec{V}) - \frac{V_i(q_i) - V_i(q_i - \Delta)}{\Delta}\right) \quad \text{(A.22)}$$

$$\mathcal{H}^{b,i}_{q_i}(\vec{V}) = \mathbb{I}(q_i < H_i) f^i_b\left(\delta^{b,i}_{q_i}(\vec{V}), (\vec{\delta}^{b,j}(\vec{V}))j \neq i\right)\left(\delta^{b,i}_{q_i}(\vec{V}) - \frac{V_i(q_i) - V_i(q_i + \Delta)}{\Delta}\right) \quad \text{(A.22)}$$

Then $\mathcal{H}^{a,i}_{q_i}$ and $\mathcal{H}^{b,i}_{q_i}$ form mappings $\mathcal{H}^a, \mathcal{H}^b : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ when we group together $q_i \in \mathcal{Q}_i$ and $i \in \{1, \dots, N\}$, defined by

$$\mathcal{H}^a(\vec{V}) = ((\mathcal{H}^{a,i}_{q_i}(\vec{V}))_{q_i \in \mathcal{Q}_i})_{i \in \{1,\dots,N\}}, \mathcal{H}^b(\vec{V}) = ((\mathcal{H}^{b,i}_{q_i}(\vec{V}))_{q_i \in \mathcal{Q}_i})_{i \in \{1,\dots,N\}} \quad \text{(A.23)}$$

From equation (A.21) and notations in (A.23) we define mapping $\Phi : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$, where $\Phi(\vec{V}) = ((\Phi^i_{q_i}(\vec{V}))_{q_i \in \mathcal{Q}_i})_{i \in \{1,\dots,N\}}$

$$\Phi^i_{q_i}(\vec{V}) = (1 - r)V_i(q_i) - \psi_i(q_i) + \lambda^a \Delta \mathcal{H}^{a,i}_{q_i}(\vec{V}) + \lambda^b \Delta \mathcal{H}^{b,i}_{q_i}(\vec{V}) \quad \text{(A.24)}$$

From Proposition 3.2, any equilibrium value function $V_i(q_i)$ is uniformly bounded by a constant M. Hence the vector $\vec{V}$ can be restricted on a convex and compact set $K \subseteq \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$. $\psi_i(q_i)$ is also uniformly bounded $\forall q_i \in \mathcal{Q}_i, \forall i \in \{1, \dots, N\}$ by $\max_{i \in \{1,\dots,N\}} \Psi_i$ defined in the proof of Proposition 3.2. We now prove that $\mathcal{H}^{a,i}_{q_i}(\vec{V})$ and $\mathcal{H}^{b,i}_{q_i}(\vec{V})$ are also uniformly bounded. From (A.23), we obtain

$$|\mathcal{H}^{a,i}_{q_i}(\vec{V})| \leq \left|\delta^{a,i}_{q_i}(\vec{V}))\Lambda(\delta^{a,i}_{q_i}(\vec{V}))\right| + \left|\Lambda(\delta^{a,i}_{q_i}(\vec{V})))\right|\frac{2L}{\Delta} \quad \text{(A.25)}$$

where L is the uniform bound of $V_i(q_i)$ defined in Proposition 3.2. From Proposition A.3 $\delta^{a,i}_{q_i}(\vec{V}))$ takes value from a compact set $K^i_{q_i}$, the functions $\Lambda(\delta)$ and $\delta\Lambda(\delta)$ are continuous in δ, hence they are bounded on the compact set $K^i_{q_i}$. Therefore, there exists a constant $C^i_{q_i}$ such that

$$|\mathcal{H}^{a,i}_{q_i}(\vec{V})| \leq C^i_{q_i}$$

We then take $C := \max_{i,q_i} C^i_{q_i}$, then C is the uniform upper bound for $|\mathcal{H}^{a,i}_{q_i}(\vec{V})|$ regardless of q_i and i. We can similarly prove $|\mathcal{H}^{b,i}_{q_i}(\vec{V})|$ is also uniformly bounded by a positive constant.

Therefore $\forall q_i \in \mathcal{Q}_i, \forall i \in \{1, \dots, N\}$, the mapping $\Phi^i_{q_i}(\vec{V})$ is uniformly bounded by an closed interval $I^i_{q_i}$ regardless of $\vec{V}$. Define set $A = \prod_{i=1}^{N}(\prod_{q_i \in \mathcal{Q}_i} I^i_{q_i})$ then $A \in \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ and A is a convex and compact set. Meanwhile $\Phi(A) \subseteq (A)$.

Finally by Proposition A.9 $\delta_{q_i}^{a,i}(\vec{V})$ and $\delta_{q_i}^{b,i}(\vec{V})$ are continuous functions of $\vec{V}$, then $\mathcal{H}_{q_i}^{a,i}(\vec{V})$ and $\mathcal{H}_{q_i}^{b,i}(\vec{V})$ are also continuous functions of $\vec{V}$. Therefore $\Phi(\vec{V}) : \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1} \to \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$ is also continuous mapping in terms of $\vec{V}$. Applying Brouwer's fixed point theorem, Φ has a fixed point $\vec{V}^* \in A \subseteq \prod_{j=1}^{N} \mathbb{R}^{2\frac{H_j}{\Delta}+1}$.

$$\Phi(\vec{V}^*) = \vec{V}^* \tag{A.26}$$

$\vec{V}^*$ satisfies the system of linear equations (A.21). Hence $\vec{V}^*$ satisfies system of HJB equations (28). By verification theorem Proposition 3.5, $\vec{V}^*$ is the equilibrium value functions of N market makers, whereas $\boldsymbol{\delta}(\vec{V}^*)$ is the joint quoting strategy under Nash equilibrium. □

APPENDIX B: PROOF OF PROPOSITION 6.2

From Lemma 6.1, the running cost of market maker is

$$\mathbb{E}_i\left[-\int_0^{\tau_a\wedge\tau_b} e^{-rt}\psi_i(q_i)dt\right] = -\frac{\psi_i(q_i)}{r+\lambda_a+\lambda_b}$$

Hence from (62) we obtain for $-H_i < q_i < H_i$:

$$\begin{aligned}
V_i^\delta(q_i) &= -\frac{\psi_i(q_i)}{r+\lambda_a+\lambda_b} + \mathbb{E}_i\Big[\mathbb{I}(R_a^i)\mathbb{I}(\tau_a<\tau_b)\big(e^{-r\tau_a}\delta_{q_i}^{a,i}\Delta + e^{-r\tau_a}V_i^\delta(q_i-\Delta)\big)\\
&\quad + \mathbb{I}(R_b^i)\mathbb{I}(\tau_b<\tau_a)\Big(e^{-r\tau_b}\delta_{q_i}^{b,i}\Delta + e^{-r\tau_b}V_i^\delta(q_i+\Delta)\Big) + \mathbb{I}((R_a^i)^c\cap(R_b^i)^c)e^{-r\tau}V_i^\delta(q_i)\Big]\\
&= -\frac{\psi_i(q_i)}{r+\lambda_a+\lambda_b} + \mathbb{P}(\tau_a<\tau_b)\mathbb{E}_i\left[\mathbb{I}(R_a^i)\big(e^{-r\tau_a}\delta_{q_i}^{a,i}\Delta + e^{-r\tau_a}V_i^\delta(q_i-\Delta)\big) + \mathbb{I}((R_a^i)^c)e^{-r\tau_a}V_i^\delta(q_i)\middle|\tau_a<\tau_b\right]\\
&\quad + \mathbb{P}(\tau_b<\tau_a)\mathbb{E}_i\left[\mathbb{I}(R_b^i)\big(e^{-r\tau_b}\delta_{q_i}^{a,i}\Delta + e^{-r\tau_b}V_i^\delta(q_i+\Delta)\big) + \mathbb{I}((R_b^i)^c)e^{-r\tau_b}V_i^\delta(q_i)\middle|\tau_b<\tau_a\right]
\end{aligned} \tag{B.1}$$

For $q_i = -H_i$

$$\begin{aligned}
V_i^\delta(-H_i) &= -\frac{\psi_i(-H_i)}{r+\lambda_a+\lambda_b} + \mathbb{E}_i\Big[\mathbb{I}(\tau_a<\tau_b)\big(e^{-r\tau_a}V_i^\delta(-H_i)\big)\\
&\quad + \mathbb{I}(\tau_b<\tau_a)\Big(\mathbb{I}(R_b^i)(e^{-r\tau_b}\delta_{-H_i}^{b,i}\Delta + e^{-r\tau_b}V_i^\delta(-H_i+\Delta)) + \mathbb{I}((R_b^i)^c)e^{-r\tau_b}V_i^\delta(-H_i)\Big)\Big]\\
&= -\frac{\psi_i(-H_i)}{r+\lambda_a+\lambda_b} + \mathbb{P}(\tau_a<\tau_b)\mathbb{E}_i\left[e^{-r\tau_a}V_i^\delta(-H_i)\middle|\tau_a<\tau_b\right]\\
&\quad + \mathbb{P}(\tau_b<\tau_a)\mathbb{E}_i\left[\mathbb{I}(R_b^i)\Big(e^{-r\tau_b}\delta_{-H_i}^{b,i}\Delta + e^{-r\tau_b}V_i^\delta(-H_i+\Delta)\Big) + \mathbb{I}((R_b^i)^c)e^{-r\tau_b}V_i^\delta(-H_i)\middle|\tau_b<\tau_a\right]
\end{aligned} \tag{B.2}$$

For $q_i = H_i$

$$V_i^\delta(H_i) = -\frac{\psi_i(H_i)}{r+\lambda_a+\lambda_b} + \mathbb{E}_i\big[\mathbb{I}(\tau_b<\tau_a)\big(e^{-r\tau_b}V_i^\delta(H_i)\big)$$

$$+\mathbb{I}(\tau_a < \tau_b)\Big(\mathbb{I}(R_a^i)(e^{-r\tau_a}\delta_{H_i}^{a,i}\Delta + e^{-r\tau_a}V_i^\delta(H_i-\Delta)) + \mathbb{I}((R_a^i)^c)e^{-r\tau_a}V_i^\delta(H_i)\Big)\Big]$$

$$= -\frac{\psi_i(H_i)}{r+\lambda_a+\lambda_b} + \mathbb{P}(\tau_b < \tau_a)\mathbb{E}_i\left[e^{-r\tau_b}V_i^\delta(H_i)\middle|\tau_b < \tau_a\right]$$

$$+\mathbb{P}(\tau_a < \tau_b)\mathbb{E}_i\left[\mathbb{I}(R_a^i)\Big(e^{-r\tau_a}\delta_{H_i}^{a,i}\Delta + e^{-r\tau_a}V_i^\delta(H_i-\Delta)\Big) + \mathbb{I}((R_a^i)^c)e^{-r\tau_a}V_i^\delta(H_i)\middle|\tau_a < \tau_b\right] \tag{B.3}$$

By combining equations (B.1)–(B.3) with indicator functions $\mathbb{I}(-H_i < q_i \le H_i)$ and $\mathbb{I}(-H_i \le q_i < H_i)$ we obtain (67).