Spoofing and Manipulating Order Books with Learning Algorithms

Álvaro Cartea[a,b], Patrick Chang[a,b], Gabriel García-Arenas[a,b]

[a] *Oxford-Man Institute of Quantitative Finance*
[b] *Mathematical Institute, University of Oxford*

Abstract

We propose a dynamic model of the limit order book to derive conditions to test if a trading algorithm will learn to manipulate the order book. Our results show that as a market maker becomes more tolerant to bearing inventory risk, the learning algorithm will find optimal strategies that manipulate the book more frequently. Manipulation occurs to induce mean reversion in inventory to an optimal level and to execute round-trip trades with limit orders at a higher probability than was otherwise likely to occur; spoofing is a special case when the market maker prefers that manipulative limit orders are not filled. The conditions are tested with order book data from Nasdaq and we show that market conditions are conducive for an algorithm to learn to manipulate the order book. Finally, when two market makers use learning algorithms to trade, their algorithms can learn to coordinate their manipulation.

Keywords: Market Microstructure, Market Making, Market Manipulation, Spoofing, Learning Algorithms, Inventory Model

1. Introduction

There is growing concern that unintended behavior may arise when decision making is delegated to artificial intelligence algorithms. Recently, the OECD and the Dutch Authority of Financial Markets (AFM) expressed concerns about algorithms learning to manipulate financial markets

We are grateful to Álvaro Arroyo, Laura Ballota, Bruno Biais, Ales Cerny, Joanne Chen, Ryan Donnelly, Jerome Detemple, Bruno Dupire, Fayçal Drissi, Rob Graumans, Leyla Han, Sebastian Jaimungal, Dror Kenett, Steven Kou, Joon-Suk Lee, Charles-Albert Lehalle, Albert Menkveld, Per Mykland, Marcel Nutz, Leandro Sánchez-Betancourt, Maxime Sauzet, Jonathan Sokobin, Lucy White, Fernando Zapatero, and Xunyu Zhou for helpful comments. We also thank seminar participants at Vrije Universiteit Amsterdam, Questrom School of Business, University of Chicago, FINRA, Columbia University, Princeton University, Fields Institute, University of Warwick, ICAIF 2023, ADRiO-EWGCFM 2023, Bayes Business School, and the Oxford Victoria Seminar. Patrick Chang and Gabriel García-Arenas acknowledge financial support from the Oxford-Man Institute.

(see OECD, 2021; AFM, 2023).[1] In this paper, we derive conditions to test if an algorithm will learn to manipulate the market through manipulative quote-based strategies such as spoofing.

A manipulative quote-based strategy consists of submitting limit orders to both sides of the order book when the objective is either to buy or to sell an asset. If the objective is to buy an asset, the strategy submits a large sell limit order that will be cancelled, and posts a limit order on the bid which is the one intended to result in a transaction. The large ask order is a manipulative order that tilts the order book and creates misleading information about the sell pressure of the asset. Market participants interpret the increase in sell pressure as an expected drop in the price of an asset, so a sell-heavy tilt in the book is followed by an increase in the arrival rate of sell orders that cross the spread in anticipation of a price drop. With the increase in the number of liquidity taking orders, the probability of buying the asset with a limit order is higher than was otherwise likely to occur because market participants will trade on the misleading signal. Similarly, if the objective is to sell an asset, then a manipulative order on the bid creates buy pressure that market participants interpret as an expected increase in the price of an asset, which allows one to sell an asset with a limit order at a higher probability than was otherwise likely to occur.

Quote-based manipulation relies on the change in behavior elicited by a manipulative order, and this change in behavior can be explained with the asymmetric information model of Glosten (1994) and a non-zero tick size in the order book. The step-function theory of Fox et al. (2021) explains that asymmetry in the volumes posted on the best bid and the best ask is interpreted by market participants as good or bad news about the asset. Specifically, when a sell limit order for a large number of shares arrives at a price equal to the existing best offer and there is no increase in the bids at the best bid price, market participants tend to react as if bad news arrived about the asset. Similarly, upon the arrival of a bid for a large number of shares at a price equal to the best bid and there is no increase in the orders at the best offer price, market participants react as if good news arrived about the asset. Therefore, when there is an imbalance between the liquidity posted at the best bid and the best ask quotes, market participants tend to interpret this as a signal to trade in a particular direction, buy or sell, in anticipation of a change in the price of the asset.

We summarize the volume imbalance between limit orders resting on the bid and on the ask sides of the book as buy-heavy, sell-heavy, and neutral. The rates with which market orders, limit orders, and cancellations arrive at the market depend on the tilt of the book; thus, the probability

[1] More generally, regulatory bodies around the world are concerned about market manipulation with trading algorithms, and they introduced legislation to address this concern. In the EU, RTS 6 and 7 require firms to test their trading algorithms so they do not behave in an unintended manner or contribute to disorderly trading conditions. In the US, the SEC approved FINRA's rule that requires algorithmic trading developers to register as securities traders, and are therefore subject to the SEC and FINRA rules that govern their trading activities.

with which limit orders are executed depends on the volume imbalance of the book. Specifically, our empirical results with data from Nasdaq show that the fill probability of a sell limit order is highest (lowest) when the book is buy-heavy (sell-heavy), and the fill probability of a buy limit is highest (lowest) when the book is sell-heavy (buy-heavy). Quote-based manipulation is profitable because traders can manipulate the tilt of the book to buy or to sell an asset with a limit order at a higher probability than was otherwise likely to occur.

To analyze if algorithms can learn to manipulate the book, we develop a dynamic model where the market maker interacts with the limit order book at discrete time intervals for an infinite trading horizon.[2] The market maker is non-myopic and is averse to holding high levels of inventory. Specifically, her objective (i.e., optimality criterion) is to maximize the present value of her expected wealth, while penalizing exposure to inventory risk. The market maker provides liquidity at the best bid and the best ask prices, and she delegates decision making to a learning algorithm to find an optimal trading strategy.[3] As with most learning algorithms, the marker maker's algorithm learns a stationary Markov strategy.[4] Here, the Markov strategy depends on her level of inventory and the state of the limit order book, which is given by its volume imbalance (i.e., tilt of the book). To understand unintended behavior that may emerge, we do not focus on the behavior of a particular learning algorithm. Instead, we analyze the decision framework of learning algorithms, so our results and testable conditions apply to any learning algorithm that finds an optimal stationary Markov strategy.

In our analysis, the market maker does not endow the algorithm with an action that manipulates the order book. Instead, we focus on how an innocuous set of actions leads to manipulation when individual actions are sequenced in a particular order. Unintentional manipulation emerges because the learning algorithm dynamically maximizes the market maker's optimality criterion. Indeed, manipulation in our setting is unintentional, but it is the best course of action when the algorithm learns the optimal strategy. This is different from unintended behavior that arises when an algorithm fails to optimize the optimality criterion. In such cases, the unintended behavior differs on a case-by-case basis and depends on the idiosyncratic assumptions of the learning algorithm.

In our model, manipulation occurs when a large limit order is placed at time t on the side of the book that counters one's objective to buy or sell an asset, and the following action at time

[2]We focus on an infinite trading horizon because most learning algorithms are designed for this setting.

[3]There are several reasons why a market maker would delegate decision making to an algorithm. For example, the rise of high-frequency trading means that delegating decision making to an algorithm is necessary for a market maker to remain competitive.

[4]See Puterman (1994), Szepesvári (2010), and Sutton and Barto (2018) for examples of generic learning algorithms. See also Calvano et al. (2020, 2021), and Abada and Lambin (2023) for examples of learning algorithms that have been studied in the context of algorithmic collusion.

$t+1$ is to place a limit order on the side of the book that aligns with one's objective to buy or sell an asset. To derive conditions to test if an algorithm will learn to manipulate the order book, we characterize the optimal stationary Markov strategy as a function of the value of the market maker's inventory aversion parameter for each state of the Markov strategy, i.e., for each pair of inventory level and volume imbalance regime.[5] The optimal strategy manipulates the book when the optimal action in the current state (i.e., inventory and volume imbalance pair) is a manipulative order (i.e., a large limit order in the "wrong" direction that will be cancelled), and the subsequent state prescribes an optimal action of placing a limit order on the side of the book that is intended to result in a transaction to complete the manipulative sequence.

Our main result provides sufficient conditions on the limit order book to test if an algorithm can learn to manipulate the order book. If certain conditions hold and a trader can tilt the book with manipulative orders, then there is a range of values of the inventory aversion parameter where the algorithm will learn to manipulate the book. In particular, we show that as the market maker becomes more tolerant to bearing inventory risk, the learning algorithm is more likely to learn manipulative strategies. The conditions depend only on the parameters of the model, and are applicable to any limit order book, e.g., Euronext, LSE, Nasdaq, NYSE. Our results show that market conditions in Nasdaq are conducive for algorithms to learn optimal strategies that manipulate the order book. In all the stocks we consider, we find that an algorithm will always learn to manipulate the order book for a range of values of the inventory aversion parameter.

One of the consequences associated with quote-based manipulation is that the manipulative order can get "caught out", i.e., the manipulative order inadvertently leads to a transaction. Our model and the learning algorithms account for this possibility. The market maker's decision to manipulate the order book balances the tradeoff between the probability of a fill of the manipulative order, the increase in inventory risk, and profits from round-trip trades due to the manipulative order. Often, when making markets, inventory levels deviate from the preferred inventory position.[6] In our model, the longer and further one deviates away from the preferred inventory position, the more severe is the penalty arising from inventory risk aversion, so the optimal strategy increases the pressure to ensure mean reversion to the preferred level of inventory. Thus, the market maker's strategy balances the trade-off between (i) buying or selling an asset to revert to the preferred inventory position at standard fill probabilities (i.e., without manipulating the book), or (ii) posting a manipulative order to manipulate the fill probabilities through the tilt of the book, which exposes the strategy to deviate further from the preferred inventory position (at least temporarily), but it

[5] Our characterization follows a similar spirit to the characterization of the optimal order choice in Parlour (1998).
[6] In our model, the preferred inventory position is zero when the fundamental value of the asset is a martingale.

also exposes the strategy to a round-trip trade. With this trade-off in mind, it is clear that a manipulative strategy becomes dynamically optimal and a more frequent optimal strategy when the market maker is less averse to holding high levels of inventory.

Counter-intuitive to the goal of quote-based manipulation, it is not always bad for the market maker to receive a fill on her manipulative order. Indeed, there are situations in which the preference is for the manipulative order to be filled because the expectation is to unwind the acquired position very quickly. In particular, we analyze if the market maker prefers that her manipulative order is filled and we find two driving forces behind the manipulation. One, the manipulation is optimal because it can lead to a manipulative round-trip trade that, in expectation, will be completed faster than otherwise. In these cases, the manipulative order is submitted with the preference for it to be filled and to unwind it immediately with an increased probability due to the tilt in the book caused by the manipulative order. Two, the manipulation is optimal because it increases the chances that the market maker's inventory will revert to a preferred position. In these cases, the preference is that the manipulative order is not inadvertently filled, which is more commonly understood as spoofing. That is, the quote-based manipulation we study includes (i) spoofing as a refinement when the preference is that the manipulative order is not inadvertently filled, and (ii) manipulation for a round-trip trade as a refinement when the preference is that the manipulative order is filled.

Additionally, our model shows that as the quoted spread narrows, learning algorithms will be less likely to manipulate the order book. Specifically, as the quoted spread decreases, the range of values of the inventory aversion parameter where manipulation is optimal decreases. In the limit, when the quoted spread is zero, an algorithm will not learn to manipulate the book because manipulation is suboptimal. Of course, theory shows that the quoted spread is positive even if the tick size is zero.[7] Nonetheless, the insight is that (i) if the profits from using limit orders are negligible and the costs from using market orders are negligible, then it is more efficient to use market orders to revert to the preferred inventory position, and (ii) if the expected profit from the opportunistic round-trip trade, where one leg is a manipulative order, does not outweigh the penalty imposed to manage the inventory risk, then quote-based manipulation is not optimal.

We extend our results in three directions. One, derive testable conditions to determine if manipulative strategies are learned when a manipulative order does not always succeed in manipulating the book. Our results show that if a manipulative order meaningfully affects the probability of tilting the book, then our testable conditions continue to hold. Two, we use backward induction

[7] See for example Stoll (1978), Ho and Stoll (1981), Copeland and Galai (1983), Glosten and Milgrom (1985), Amihud and Mendelson (1986), and Glosten (1994).

to solve numerically for an optimal strategy and find that algorithms can also learn to manipulate the order book when the trading horizon is finite. Finally, we study the effect of introducing a competing market maker. We find that if both market makers train their algorithms offline, then their algorithms either coordinate or mis-coordinate depending on their initial inventory level. On the other hand, if both market makers train their algorithms online, then their algorithms learn to coordinate by either riding the manipulative sequences of each other or by allowing one market maker to ride the other market maker's manipulative sequences to avoid mis-coordination.

In the literature, traders can attempt to manipulate the market in several ways. Studies focus on information-based manipulation and trade-based manipulation. Information-based manipulation occurs when the manipulator releases misleading information (see for example Bagnoli and Lipman, 1996; Van Bommel, 2003; Vila, 1989), whereas trade-based manipulation occurs when the manipulator buys or sells an asset to effect changes in the price (see for example Allen and Gale, 1992; Allen and Gorton, 1992; Chakraborty and Yılmaz, 2004a,b).

On the other hand, spoofing is a particular case of quote-based manipulation that has received little analysis (see Fox et al., 2021). One exception is Williams and Skrzypacz (2021) who extend the setup of Glosten and Milgrom (1985) to show that spoofing can occur in equilibrium, and to study the equilibrium consequences. In our paper, instead of using an information-based model, we use an inventory model to analyze if algorithms can learn to manipulate the order book. Similar to the approach in Ho and Stoll (1981), we propose a model of the market dynamics that is consistent with empirical stylized facts, where the market dynamics do not necessarily derive from principles of individual economic behavior. The purpose of our paper is not to explain the underlying economic reasoning of the market dynamics, but to use the market dynamics to derive conditions to test if an algorithm will learn to to manipulate the market.

We focus on the broader quote-based manipulation because existing legislation does not explicitly outlaw spoofing.[8] Spoofing is illegal because the manipulative order (also known as a spoof order) creates misleading information to buy or to sell an asset with a higher probability than was otherwise likely to occur. The manipulative order is what is considered illegal under existing legislation, and the reason why spoofing is illegal. Thus, our results allow us to understand and test for quote-based manipulation, but also include spoofing as a particular case. Moreover, our model allows us to make the key distinction between spoofing and other forms of quote-based manipulation based on one's preference of a fill of a manipulative order.

Our results have implications for how a rational market maker should behave within the market dynamics of our model. Indeed, our work is a discrete-time analogue to papers that use stochastic

[8]The exception is the Dodd-Frank (2010) Act, but the Act only applies to the US commodities market.

optimal control and continuous-time models to derive algorithmic trading strategies. For example, Cartea et al. (2020) derive a manipulative quote-based strategy to acquire or liquidate a large position. A key difference is that they explicitly encode a manipulative action, whereas manipulation is optimal but unintentional in our model because the market maker does not endow the algorithm with an action that manipulates the book. Also, their paper is not about learning algorithms.

On the empirical side, Lee et al. (2013) use a proprietary dataset with trader identification from the Korea Exchange to show that spoofing achieves substantial extra profits and spoofing tends to target stocks with higher return volatility, lower market capitalization, lower price level, and lower managerial transparency. Wang (2019) uses data from the Taiwan Futures Exchange to show that market participants spoof the order book in stocks that exhibit high volumes of trading, high volatility, and high prices. Wang also shows that spoofing increases the volume of trading, increases the volatility of prices, and increases the quoted spread. Our empirical results complement their findings because we find that market conditions from Nasdaq are such that algorithms will learn to manipulate the order book.

Finally, our work is closer to the literature that studies the unintentional effects of algorithms that learn to collude (see for example Calvano et al., 2020, 2021).[9] Our approach is similar to that in Cartea et al. (2023) who prove that algorithms can learn to collude. They analyze the equilibria that can be learned and prove convergence to collusive equilibria; whereas in this paper, we analyze the decision framework where algorithms learn optimal strategies and derive testable conditions to determine if algorithms will learn to manipulate the order book. Additionally, similar to the algorithmic collusion literature, we find that algorithms can also learn to coordinate their manipulation when they learn together.

The remainder of the paper proceeds as follows. The next section shows the relationship between volume imbalance and the behavior of market participants. Section 3 presents our model and Section 4 derives the optimal strategy and testable conditions to determine if a manipulative strategy can be learned. Section 5 analyzes the mechanics of quote-based manipulation, its relation to spoofing, and how the parameters of the model affect quote-based manipulation as the optimal strategy. Section 6 tests the conditions with order book data from Nasdaq. Section 7 relaxes model assumptions and extends our testable conditions. Finally, Section 8 studies manipulation with multiple market makers and Section 9 concludes with some regulatory implications.

[9] See also Cartea et al. (2022a,b), Colliard et al. (2022), and Dou et al. (2023) for studies on algorithmic collusion in financial markets.

2. Volume Imbalance and Order Book Activity

This section uses data from Nasdaq for April 2023 to illustrate the relationship between volume imbalance and the activity in the limit order book that is central to quote-based manipulation. For each trading day, we remove the first and last 15 minutes to exclude behavior in the limit order book during the opening and closing auctions.

Volume imbalance at time t is given by

$$\omega_t = \frac{V_t^b - V_t^a}{V_t^b + V_t^a} \in (-1, 1)\,, \tag{1}$$

where $V_t^b\,, V_t^a > 0$ are the liquidity posted at the best bid and the best ask, respectively, at time t. Volume imbalance summarizes the tilt of the limit order book, so when ω_t is close to 1 there is a strong buy pressure and when ω_t is close to -1 there is a strong sell pressure. To simplify our subsequent analysis, we discretize volume imbalance into three regimes: buy-heavy (BH) when $\omega_t \in (1/3, 1)$, neutral (N) when $\omega_t \in [-1/3, 1/3]$, and sell-heavy (SH) when $\omega_t \in (-1, -1/3)$.

Table 1: Arrival rates of market orders (MOs) for April 2023.

Ticker	Buy MO arrival rates (per second)			Sell MO arrival rates (per second)		
	SH	N	BH	SH	N	BH
AAPL	0.060	0.176	**0.525**	**0.606**	0.179	0.058
AMZN	0.067	0.168	**0.447**	**0.456**	0.167	0.065
CSCO	0.008	0.025	**0.101**	**0.104**	0.033	0.012
INTC	0.014	0.042	**0.138**	**0.139**	0.036	0.013
MSFT	0.297	0.350	**0.461**	**0.475**	0.352	0.286
TSLA	0.532	0.677	**0.773**	**0.750**	0.635	0.529

Table 1 presents the arrival rates (per second) of buy and sell market orders in each volume imbalance regime for the several assets, and Table 2 presents the average volume of the market orders that arrive. The arrival rate of buy (sell) market orders is highest when the book is buy-heavy (sell-heavy). However, the average volume of buy (sell) market orders is lowest when the book is buy-heavy (sell-heavy). Nevertheless, the net effect (i.e., arrival rate times average volume) is that there are more buy (sell) transactions when the book is buy-heavy (sell-heavy).

Similarly for limit orders, Table 3 presents the arrival rates (per second) of buy and sell limit orders in each volume imbalance regime for the assets we consider, and Table 4 presents the average volume of the limit orders that arrive. The arrival rates of buy limit orders are higher than

Table 2: Average volume of market orders (MOs) for April 2023.

Ticker	Buy MO average volume			Sell MO average volume		
	SH	N	BH	SH	N	BH
AAPL	145.58	111.16	**62.27**	**64.29**	103.11	135.29
AMZN	205.95	108.59	**61.68**	**60.92**	107.78	181.50
CSCO	149.88	182.33	**110.30**	**111.74**	161.72	131.67
INTC	212.79	256.93	**143.67**	**134.62**	266.80	227.45
MSFT	72.17	46.41	**20.66**	**21.25**	45.57	70.14
TSLA	132.27	56.33	**26.55**	**25.35**	54.34	120.54

Table 3: Arrival rates of limit orders (LOs) for April 2023.

Ticker	Buy LO arrival rates (per second)			Sell LO arrival rates (per second)		
	SH	N	BH	SH	N	BH
AAPL	4.245	7.000	4.129	4.285	6.970	4.303
AMZN	4.367	7.346	4.381	4.637	7.600	4.213
CSCO	0.567	1.345	0.951	0.882	1.416	0.641
INTC	1.090	2.386	1.829	1.686	2.330	1.106
MSFT	3.631	4.126	4.452	4.830	4.268	3.582
TSLA	3.913	4.628	3.309	3.560	4.784	4.088

Table 4: Average volume of limit orders (LOs) for April 2023.

Ticker	Buy LO average volume			Sell LO average volume		
	SH	N	BH	SH	N	BH
AAPL	98.19	109.51	**112.32**	**115.30**	109.50	97.40
AMZN	87.95	96.32	**101.95**	**101.53**	95.44	87.83
CSCO	228.77	271.07	**307.59**	**319.84**	267.68	234.34
INTC	298.83	372.36	**415.56**	**415.82**	364.61	292.29
MSFT	42.93	44.81	**45.69**	45.25	**45.73**	44.76
TSLA	49.64	55.64	**60.97**	**66.90**	61.38	57.94

those of sell limit orders when the book is buy-heavy, and the arrival rates of sell limit orders are higher than those of buy limit orders when the book is sell-heavy. On the other hand, the average volume of buy (sell) limit orders is largest when the book is buy-heavy (sell-heavy). The net effect (i.e., arrival rate times average volume) is that there are more buy (sell) limit orders than sell (buy) limit orders when the book is buy-heavy (sell-heavy).

Table 5: Arrival rates of limit order cancellations for April 2023.

Ticker	Arrival rates of limit buy cancellation (per second)			Arrival rates of limit sell cancellation (per second)		
	SH	N	BH	SH	N	BH
AAPL	3.092	6.037	3.785	3.946	6.154	3.334
AMZN	3.594	6.464	3.631	3.885	6.734	3.529
CSCO	0.398	1.021	0.701	0.682	1.054	0.452
INTC	0.818	1.812	1.308	1.193	1.755	0.801
MSFT	3.673	3.561	3.082	3.316	3.756	3.719
TSLA	2.641	3.475	2.824	3.066	3.607	2.777

Table 6: Average volume of limit order cancellations for April 2023.

Ticker	Average volume of limit buy cancellations			Average volume of limit sell cancellations		
	SH	N	BH	SH	N	BH
AAPL	92.25	109.29	**112.92**	**116.77**	111.55	91.26
AMZN	78.53	95.69	**106.34**	**103.48**	94.28	79.87
CSCO	172.01	281.82	**370.26**	**378.71**	281.96	179.55
INTC	230.90	393.75	**499.81**	**489.77**	392.68	228.14
MSFT	23.17	41.34	**63.64**	**63.67**	41.85	23.55
TSLA	29.17	50.51	**71.24**	**82.05**	54.89	35.21

Similarly, Table 5 presents the arrival rates (per second) of limit buy and limit sell cancellations in each volume imbalance regime for the three assets we consider, and Table 6 presents the average volume of limit order cancellations that arrive.

Volume imbalance clearly influences the behavior of market participants. This is consistent with the work of Harris and Panchapagesan (2005), who find that market participants condition their quotation behavior on volume imbalance. These empirical findings are further supported by a survey sent to Dutch algorithmic trading firms, where AFM found that trading algorithms use between 100 and 1000 features, and volume imbalance is one of the key features (see AFM, 2023).

Finally, Table 7 reports the probability that a limit order on the best bid or best ask is filled within the next five seconds, one second, and half of a second, respectively. These fill probabilities account for the effect of time-priority; see Section 6 for details about the estimation procedure. This link between the fill probability and volume imbalance regimes makes quote-based manipulation viable and profitable. By manipulating the volume imbalance and tilting the order book, the

Table 7: Fill probabilities.

Ticker	Side	5 seconds			1 second			0.5 seconds		
		SH	N	BH	SH	N	BH	SH	N	BH
AAPL	Ask	0.4393	0.4782	**0.5819**	0.1048	0.1286	**0.1910**	0.0449	0.0579	**0.0928**
	Bid	**0.6210**	0.5207	0.4687	**0.2196**	0.1499	0.1180	**0.1121**	0.0697	0.0518
AMZN	Ask	0.4155	0.4651	**0.5669**	0.1008	0.1232	**0.1903**	0.0451	0.0566	**0.0933**
	Bid	**0.5587**	0.4570	0.4201	**0.1767**	0.1228	0.1044	**0.0863**	0.0566	0.0479
CSCO	Ask	0.0674	0.1005	**0.1768**	0.0093	0.0151	**0.0384**	0.0040	0.0060	**0.0198**
	Bid	**0.1851**	0.1034	0.0713	**0.0400**	0.0154	0.0109	**0.0201**	0.0064	0.0048
INTC	Ask	0.0970	0.1384	**0.2353**	0.0158	0.0222	**0.0561**	0.0071	0.0095	**0.0274**
	Bid	**0.2124**	0.1314	0.1116	**0.0501**	0.0211	0.0161	**0.0251**	0.0089	0.0070
MSFT	Ask	0.5577	0.5838	**0.6286**	0.2033	0.2135	**0.2529**	0.1037	0.1095	**0.1333**
	Bid	**0.6101**	0.5739	0.5605	**0.2339**	0.2081	0.2041	**0.122**	0.1064	0.1048
TLSA	Ask	0.6051	0.6280	**0.6729**	0.2618	0.2751	**0.3222**	0.1459	0.1517	**0.1876**
	Bid	**0.668**	0.6201	0.6047	**0.3137**	0.2660	0.2661	**0.1820**	0.1482	0.1480

probability of buying or selling the asset with a limit order is higher than it was otherwise likely to occur.

3. The Model

Here, we present our dynamic model of the limit order book where the market maker provides liquidity at the best bid and best ask. The market maker is non-myopic and interacts with the limit order book at discrete times $t = 0, 1, 2, ..., +\infty$. The market maker delegates the decision making process to a learning algorithm that can be trained online or offline.

3.1. Setup

Framework. We model the decision process of the market maker as a Markov decision process $\mathcal{M} = \langle \mathcal{S}, (\mathcal{A}_{\mathbf{s}})_{\mathbf{s}\in\mathcal{S}}, p, (u_{\mathbf{s}})_{\mathbf{s}\in\mathcal{S}}, \delta \rangle$. Let $\mathbf{s} \in \mathcal{S}$ denote the state, where the set $\mathcal{S}$ is finite, and let $\mathcal{A}_{\mathbf{s}}$ denote the finite set of actions for the market maker in state $\mathbf{s}$. The state evolves according to the transition function $p : \mathcal{S} \times \mathcal{A}_{\mathbf{s}} \to \Delta(\mathcal{S})$, where $\Delta(\mathcal{S})$ is the set of probability measures on $\mathcal{S}$. We denote by $p(\mathbf{s}'|\mathbf{s}, a)$ the probability that the subsequent state is $\mathbf{s}'$ given that the current state is $\mathbf{s}$ and action a is played. At every time step t, the payoff is given by a utility function $u : \mathcal{S} \times \mathcal{A}_s \times \mathcal{S} \to \mathbb{R}$, where $|u(\mathbf{s}, a, \mathbf{s}')| < \infty$ for all $a \in \mathcal{A}_{\mathbf{s}}$ and $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. The payoff from the utility function $u(\mathbf{s}, a, \mathbf{s}')$ depends on the transition from state $\mathbf{s}$ to state $\mathbf{s}'$ under action a. Finally, $\delta \in [0, 1)$ is the parameter with which the market maker discounts the future stream of payoffs.

Strategies. The market maker uses an algorithm to learn a time-invariant strategy that depends only on the state $\mathbf{s}$, i.e., we consider stationary Markov strategies. A stationary Markov strategy describes a mapping to a set of probability measures on $\mathcal{A}_\mathbf{s}$ for each state $\mathbf{s}$, i.e., $\sigma \in \Sigma^{SM} = \prod_{\mathbf{s}\in\mathcal{S}} \Delta(\mathcal{A}_\mathbf{s})$ such that $\sigma : \mathcal{S} \to \Delta(\mathcal{A}_\mathbf{s})$. Similarly, a stationary pure Markov strategy describes a mapping to an action $\mathcal{A}_\mathbf{s}$ for each state $\mathbf{s}$, i.e., $\sigma \in \Sigma^{SPM} = \prod_{\mathbf{s}\in\mathcal{S}} \mathcal{A}_\mathbf{s}$ such that $\sigma : \mathcal{S} \to \mathcal{A}_\mathbf{s}$. More generally, a strategy is a mapping from the set of all possible histories to a set of probability measures on $\mathcal{A}_\mathbf{s}$, i.e., $\sigma \in \Sigma$ such that $\sigma : \mathcal{H} \to \Delta(\mathcal{A}_\mathbf{s})$, where $\mathcal{H} = \cup_{t=0}^{\infty} \mathcal{H}_t$ and $\mathcal{H}_t$ satisfies the recursion $\mathcal{H}_t = \mathcal{H}_{t-1} \times \mathcal{S} \times \cup_{\mathbf{s}\in\mathcal{S}} \mathcal{A}_\mathbf{s}$ with $\mathcal{H}_0 = \mathcal{S}$. In general, a strategy need not be time-invariant.

The restriction to stationary Markov strategies is essential for a learning algorithm to find an optimal strategy because it significantly reduces the space of possible strategies. Hence, an algorithm does not need to search for an optimal strategy over the space of all possible (history-dependent) contingency plans. Our focus on stationary Markov strategies is not restrictive because classical results (given below) show that there exists a stationary pure Markov strategy that achieves the same optimality criteria as an optimal strategy from Σ. Indeed, most learning algorithms search for an optimal strategy in the space of stationary Markov strategies.

Optimality Criteria. The value of a strategy $\sigma \in \Sigma^{SM}, \Sigma^{SPM}, \Sigma$ that starts in state $\mathbf{s}$ is the continuation value of the strategy from state $\mathbf{s}$, i.e., the expected discounted stream of payoffs from implementing strategy σ is given by

$$\upsilon_\mathbf{s}(\sigma) = \mathbb{E}_\sigma \left[\sum_{t=0}^{\infty} \delta^t \, u(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) \,\middle|\, \mathbf{s}_0 = \mathbf{s} \right], \tag{2}$$

where the expectation in (2) is with respect to the strategy σ. That is, actions are sampled from σ and the expectation is taken over $p(\mathbf{s}_{t+1}|\mathbf{s}_t, a_t)$.

For a fixed value of the discount parameter $\delta \in [0, 1)$, the optimal continuation value is given by $\upsilon^*_\mathbf{s} = \sup_{\sigma\in\Sigma} \upsilon_\mathbf{s}(\sigma)$. Existence and uniqueness of $\upsilon^* = (\upsilon^*_\mathbf{s})_{\mathbf{s}\in\mathcal{S}}$ is guaranteed by Theorem 6.2.5a in Puterman (1994). Therefore, an optimal strategy $\sigma^* \in \Sigma^{SM}, \Sigma^{SPM}, \Sigma$ exists if $\upsilon^*_\mathbf{s} = \upsilon_\mathbf{s}(\sigma^*) \geq \upsilon_\mathbf{s}(\sigma)$ for all $\mathbf{s} \in \mathcal{S}$ and all $\sigma \in \Sigma$. Crucially, Theorem 6.2.10 in Puterman (1994) guarantees that there exists an optimal stationary pure Markov strategy $\sigma^* \in \Sigma^{SPM}$, such that $\upsilon_\mathbf{s}(\sigma^*) \geq \upsilon_\mathbf{s}(\sigma)$ for all $\mathbf{s} \in \mathcal{S}$ and all $\sigma \in \Sigma$.

Therefore, with these theoretical guarantees, we ignore all strategies that are history-dependent contingency plans, and for the remainder of the paper, a strategy refers to a stationary Markov strategy, and a pure strategy refers to a stationary pure Markov strategy.

3.2. Trading Environment

We present our model of the limit order book. Many of our assumptions are for tractability purposes and conform with the market dynamics described in Section 2. We use the midpoint of the bid-ask spread as a proxy for the fundamental value of the asset Z. At each time point, the value of the asset either goes to $Z + \varphi$ with probability $\beta \in (0, 1)$, or goes down to $Z - \varphi$ with probability $1 - \beta$, where $\varphi > 0$ is the tick size. The fundamental value of the asset is a martingale when $\beta = 0.5$, and it drifts up or down when $\beta > 0.5$ or $\beta < 0.5$, respectively.

States. The set of states $\mathcal{S}$ is the Cartesian product of a set of environmental variables Ω and the inventory of the market maker $\mathcal{Q}$, i.e., $\mathcal{S} = \Omega \times \mathcal{Q}$. We restrict the level of inventory to the set $\mathcal{Q} = \{-\bar{q}, ..., 0, ..., \bar{q}\}$, where $\bar{q}$ is some positive integer. The set Ω contains a finite number of environmental variables which are relevant features of the order book and that affect the payoffs the market maker receives. Here, the elements of Ω are the three regimes of volume imbalance in the limit order book, i.e., $\omega \in \Omega = \{BH, N, SH\}$, because quote-based manipulation is the focus of our analysis.

As discussed in Section 2, volume imbalance affects the behavior of market participants, and therefore affects the probability with which a limit order is filled. To capture this effect, let $p^b_\omega \in (0, 1)$ and $p^a_\omega \in (0, 1)$ denote the probability that a limit buy and a limit sell order, respectively, are filled in $[t, t + 1)$ for each regime $\omega \in \Omega$.[10] These fill probabilities account for the effect of time-priority in the limit order book and need not sum to unity. Our empirical results in Section 2 (see Table 7) show that the fill probabilities of bids and offers are similar when the book is neutral (i.e., $p^b_N \approx p^a_N$), the fill probabilities of offers are higher when the book is buy-heavy (i.e., $p^b_{BH} \ll p^a_{BH}$), and the fill probabilities of bids are higher when the book is sell-heavy (i.e., $p^b_{SH} \gg p^a_{SH}$).

Actions. The market maker does not endow the algorithm with an action that manipulates the order book. Instead, we focus on how an innocuous set of actions leads to manipulation when

[10] In addition to affecting the fill probabilities, volume imbalance has substantial explanatory power in predicting future price movements (see e.g., Harris and Panchapagesan, 2005; Cao et al., 2009). Therefore, submitting a manipulative order may change the volume imbalance regime, which may affect future price movements; however, it should not affect the fundamental value of the asset because manipulative orders contain no real information about the fundamentals. In our model, we do not model how volume imbalance affects future price movements because we use the midpoint of the bid-ask spread as a proxy for the fundamental value of the asset. This prevents unnecessary complications that arise when computing the change in wealth for the payoffs received as a consequence of submitting a manipulative order. Our analysis focuses specifically on how manipulative orders affect the fill probabilities, and how the fill probabilities affect the payoffs at the next period. If manipulative orders were to affect future price movements (temporarily), then quote-based manipulation would become more prevalent in our model because a manipulative order will increase the profits of round-trip trades. In Section 5, we show that the profit from round-trip trades is a key factor to determine if quote-based manipulation is optimal.

individual actions are sequenced in a particular order. That is, quote-based manipulation occurs when the set of actions produces unintended manipulative behavior as a consequence of a learning algorithm dynamically optimizing the market maker's optimality criteria.

The set of actions at time t consists of:

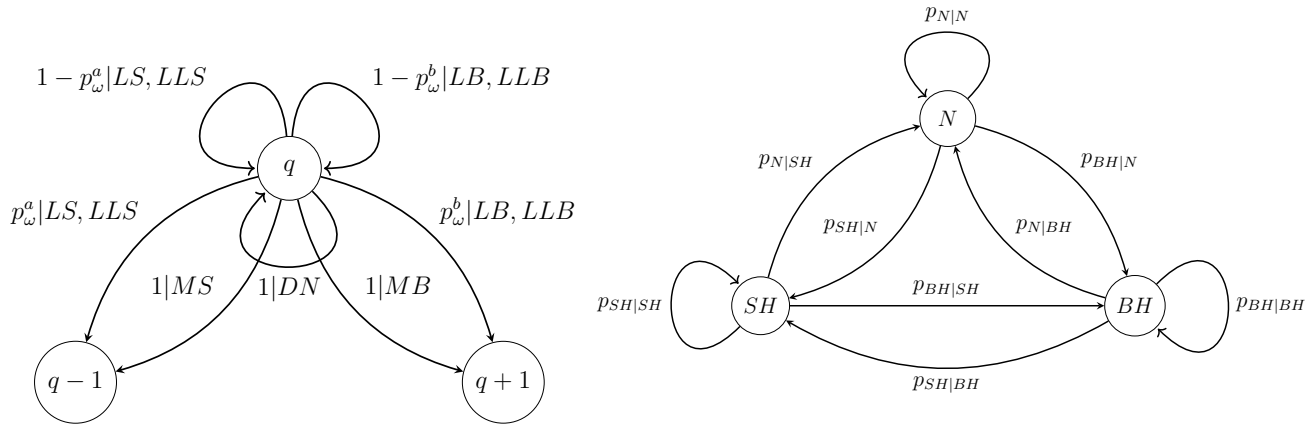
- Submit a buy limit order (LB) on the best bid or a sell limit order (LS) on the best offer for one unit of the asset. If the limit order is not executed between $[t, t+1)$, then the order is cancelled before the start of $t+1$.
- Submit a large buy limit order (LLB) on the best bid or a large sell limit order (LLS) on the best offer and cancel the order before the start of $t+1$.
- Submit a market order to buy (MB) or to sell (MS) one unit of the asset.
- Do nothing (DN).

When a large limit order is submitted, the probability that the manipulative order is filled is low. Recall that when the large limit buy (limit sell) tilts the book to buy-heavy (sell-heavy), Table 1 shows that the arrival rate of sell (buy) market orders is very low. Therefore, we assume that at most one unit of the large limit order can be executed. This assumption allows us to validate that a particular sequence of actions has the intention to manipulate; see the analysis in Section 5.

We do not model the strategic cancellation of limit orders for the sake of analytical tractability.[11] Rather, we focus on when algorithms learn to create misleading information to buy or to sell an asset with a higher probability than was otherwise likely to occur, which is the key feature that enables quote-based manipulation. Explicitly cancelling a limit order, i.e., sending an instruction to the exchange to cancel an outstanding limit order, is not necessary in Nasdaq if the limit order is submitted with a time-in-force between t and $t+1$ because the limit order expires at time $t+1$.

The set of actions available to the market maker depends on the state $\mathbf{s} \in \mathcal{S}$. The market maker has access to the full set of actions $\mathcal{A}_{\mathbf{s}} = \{LB, LS, LLB, LLS, MB, MS, DN\}$ when $\mathbf{s} = (\omega, q)$ for all $\omega \in \Omega$ and for all $q \in \mathcal{Q} \setminus \{\bar{q}, -\bar{q}\}$. At the boundaries of the inventory constraint, the market maker does not buy or does not sell any additional assets. That is, the set of actions reduces to $\mathcal{A}_{\mathbf{s}} = \{LS, LLS, MS, DN\}$ when $\mathbf{s} = (\omega, \bar{q})$ for all $\omega \in \Omega$, and to $\mathcal{A}_{\mathbf{s}} = \{LB, LLB, MB, DN\}$ when $\mathbf{s} = (\omega, -\bar{q})$ for all $\omega \in \Omega$.

[11] Including cancellation of an order as part of the action prevents us from including an additional state variable that tracks if a market maker has an outstanding order in the book. Additionally, the fill probability between time $[t, t+1)$ and $[t+1, t+2)$ for a limit order submitted at time t is not the same unless the fill probability is Markov.



(a) Transition dynamics and probabilities conditional on the action starting from state $q \in \mathcal{Q}$.

(b) Transition dynamics and probabilities conditional on small orders or no orders, i.e., $a \in \{LB, LS, MB, MS, DN\}$.

Figure 1: State transition diagram of (a) the level of inventory, and (b) the volume imbalance regime.

In this setup, individual actions are not manipulative. Although a large limit order can tilt the book, the action is not manipulative because there is no advantage to be gained from the action alone; rather, a large limit order will only be manipulative when the action following a large limit order aims to profit from the tilt created. For this reason, we focus on a manipulative sequence of actions. We formalize and refine this later in Definition 1.

Transition Dynamics. We present the transition dynamics over the set $\mathcal{Q}$ and Ω separately to simplify the presentation, and recall that the set of states is $\mathcal{S} = \Omega \times \mathcal{Q}$.

The transition dynamics of the inventory level is intuitive. If an action resulted in a buy transaction, then the level of inventory increases by one unit, i.e., $q_{t+1} = q_t + 1$; if an action resulted in a sell transaction, then the level of inventory decreases by one unit, i.e., $q_{t+1} = q_t - 1$; and if an action resulted in no transaction, then the level of inventory stays the same, i.e., $q_{t+1} = q_t$. The transition probabilities depend on the action taken and the fill probabilities (for limit order submissions). Figure 1a presents the state transition diagram when the current state is $q \in \mathcal{Q}$. The edges indicate the transition probability conditional on the action taken.

The transition probabilities of the volume imbalance regime depend on two distinct cases: when a small or no order is submitted, i.e., $a \in \{LB, LS, MB, MS, DN\}$, and when a large order is submitted, i.e., $a \in \{LLB, LLS\}$. When a small or no order is submitted, volume imbalance evolves according to the baseline dynamics in Figure 1b because unit orders and no orders have little to no impact on the liquidity at the best bid-ask quotes, and hence have little to no impact on the volume imbalance in (1).

On the other hand, when a large limit order is submitted, the volume of the buy (sell) limit order is large enough to tilt the volume imbalance regime to the buy-heavy (the sell-heavy) regime. If we ignore the change in behavior of other market participants, then the volume imbalance regime changes at time t and reverts back before time $t+1$ because of the timing of submission and cancellation of large limit orders posted by the market maker. However, in Section 2, we saw that market participants adjust their own liquidity provision. Empirically, we have the following observations: (i) the average volume of limit orders and limit order cancellations is similar in size under each volume imbalance regime, but the arrival rates of limit orders are higher than the arrival rates of limit order cancellations, and (ii) there are more buy (sell) limit orders than sell (buy) limit orders when the book is buy-heavy (sell-heavy). Therefore, the empirical results show that market participants adjust their own liquidity provision to submit more limit orders to the same side of the book as the market maker's large limit order. However, market participants can never react instantaneously and will have a delay when reacting to a large limit order from the market maker. Such delay can occur for a number of reasons, for example, other market participants do not have the monitoring capabilities (i.e., latency, see Aquilina et al., 2021), or do not have the infrastructure to react immediately.

The effect of the delay and change in liquidity provision is that the volume imbalance regime at time $t+1$ corresponds to the change in regime caused by the market maker's large limit order at time t. Therefore, when a large limit order is submitted, the volume imbalance regime moves to the buy-heavy (the sell-heavy) regime with probability one at the next time step, i.e., $p(BH \,|\, \omega, LLB) = 1$ and $p(SH \,|\, \omega, LLS) = 1$ for all $\omega \in \Omega$, and the change in fill probabilities comes into effect at time $t+1$ (as a consequence of the delay); Section 7.1 below considers large limit order that do not always succeed in tilting the book.

These stylized facts are important for unintentional quote-based manipulation to occur. They create the necessary temporal link between past actions and future actions of the market maker so that quote-based manipulation becomes dynamically optimal as a sequence of actions. Thus, quote-based manipulation emerges as an optimal sequence of actions even if the market maker does not encode quote-based manipulation as a possible action into the learning algorithm.

Utility. The market maker is averse to holding inventory and maximizes the present value of her wealth. The wealth $W = X + Z\,q$ of the market maker is the sum of her cash position X and the marked-to-market value of the inventory $Z\,q$, where Z is the fundamental value of the asset that is proxied by the midpoint of the bid-ask spread. To cast the market maker's objective into the

optimization problem of a learning algorithm, we write the one-step utility as

$$u(\mathbf{s}, a, \mathbf{s}') = Y(\mathbf{s}, a, \mathbf{s}') - \alpha\,(q')^2\,, \tag{3}$$

where $q' \in \mathcal{Q}$ is the inventory after action a and $\alpha > 0$ is the inventory aversion parameter.[12] The quadratic penalty ensures that the utility function is concave in the level of inventory. Hence, the inventory aversion parameter α affects the willingness of the market maker to take on larger levels, long or short, of inventory. For example, as the value of α increases, the market maker is more averse to inventory risk so she is less willing to increase the level of inventory, long or short. On the other hand, $Y(\mathbf{s}, a, \mathbf{s}')$ is the change in wealth as a consequence of action $a \in \mathcal{A}_\mathbf{s}$ in state $\mathbf{s} \in \mathcal{S}$. For a value of the tick size $\varphi > 0$, the expected change in wealth from state $\mathbf{s} = (\omega, q) \in \mathcal{S}$ and taking action $a \in \mathcal{A}_\mathbf{s}$ is given by

$$\mathbb{E}\left[Y(\mathbf{s}, a, \mathbf{s}')\right] = \begin{cases} p^b_\omega\,\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q + p^b_\omega\,\varphi) & \text{for } a = \{LB, LLB\}\,, \\ p^a_\omega\,\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q - p^a_\omega\,\varphi) & \text{for } a = \{LS, LLS\}\,, \\ -\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q + \varphi) & \text{for } a = MB\,, \\ -\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q - \varphi) & \text{for } a = MS\,, \\ (2\,\beta - 1)\,\varphi\,q & \text{for } a = DN\,, \end{cases}$$

where the expectation is taken with respect to the fundamental value of the asset Z and the fill probabilities p^b_ω, p^a_ω of limit orders, and $\vartheta > 0$ is the expected quoted spread.

With this one-step utility function, the optimal continuation value is given by

$$\sup_{\sigma \in \Sigma} \mathbb{E}_\sigma \left[\sum_{t=0}^{\infty} \delta^t \left(Y(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) - \alpha\, q^2_{t+1} \right) \middle| \, \mathbf{s}_0 = \mathbf{s} \right], \tag{4}$$

which corresponds to maximizing the present value of wealth subject to a running inventory penalty.[13] The optimization problem in (4) is similar to the optimization problem posed in O'Hara and Oldfield (1986) where they assume a negative exponential for the one-step utility function. Our choice of (3) leads to a clear interpretation of the continuation value, and it also simplifies our

[12] The units of α are such that (4) is in units of wealth. In our model, the value of α is allowed to range between $(0, \infty)$, but in practice, as a rule-of-thumb, the value of α must be several orders of magnitudes smaller than the expected quoted spread, i.e., $\alpha \ll \vartheta$, otherwise the market maker will not be willing to make markets.

[13] In (4), the payoffs received depends on the realization of the inventory level at the next time step; however, the payoffs are discounted from the start of the period. The timing of the payoffs is constructed to fit within the framework of learning algorithms, where the payoffs are assumed to be immediate. This construction presents no issue in (4) because of the expectation operator, and the timing is also consistent with the model of O'Hara and Oldfield (1986).

analysis in the next section.

The objective to maximize wealth subject to a running inventory penalty is closely related to robustness and ambiguity aversion from Hansen and Sargent (2007). Specifically, Cartea et al. (2017) show in a related problem that the market maker's objective of maximizing wealth subject to a running inventory penalty is equivalent to a risk-neutral to inventory market maker who is ambiguous to the drift of the fundamental value of the asset.

Finally, the optimality criterion in (3) produces behavior consistent with inventory models. The behavior of the optimal strategy depends on the level of inventory, and there is a preferred inventory position (the preferred inventory position is zero when the fundamental price is a martingale), which is consistent with the results of Amihud and Mendelson (1986).[14] The optimal strategy also prefers to sell if inventory is long and prefers to buy if inventory is short, which is consistent with the results of Stoll (1978) and Ho and Stoll (1981).

For simplicity, the remainder of the paper assumes that the fundamental value of the asset is a martingale, i.e., $\beta = 0.5$, so that the expected one-step utility from state $\mathbf{s} = (\omega, q) \in \mathcal{S}$ and taking action $a \in \mathcal{A}_{\mathbf{s}}$ is given by

$$\bar{u}(\mathbf{s}, a) = \begin{cases} p^b_\omega \, \vartheta/2 - \alpha \, p^b_\omega \, (q+1)^2 - \alpha \, (1 - p^b_\omega) \, q^2 & \text{for } a = \{LB, LLB\}\,, \\ p^a_\omega \, \vartheta/2 - \alpha \, p^a_\omega \, (q-1)^2 - \alpha \, (1 - p^a_\omega) \, q^2 & \text{for } a = \{LS, LLS\}\,, \\ -\vartheta/2 - \alpha \, (q+1)^2 & \text{for } a = MB\,, \\ -\vartheta/2 - \alpha \, (q-1)^2 & \text{for } a = MS\,, \\ -\alpha \, q^2 & \text{for } a = DN\,. \end{cases}$$

4. Theory

This section derives conditions to test if quote-based manipulation will occur when decision making is delegated to a learning algorithm. Our analysis focuses on $q \neq 0$, where the overall intention is to buy or to sell the asset so that inventory reverts to the preferred position $q = 0$. We omit the case when $q = 0$ because the overall intention to buy or to sell is unclear.

Throughout the paper, we maintain the following assumptions on the parameters of our model.

Assumption *The expected quoted spread and the inventory aversion parameter are both greater than zero (i.e., $\vartheta > 0$ and $\alpha > 0$), the fill probabilities $p^b_\omega \in (0,1)$ and $p^a_\omega \in (0,1)$ for all $\omega \in \Omega$, and the market maker is not myopic, i.e., $\delta > 0$.*

[14] See also Madhavan and Smidt (1993) and Hasbrouck and Sofianos (1993) who examine this preferred inventory position in more detail.

4.1. Optimal Strategy

The optimal strategy satisfies Bellman's optimality equations, so the optimal action in each state $\mathbf{s} = (\omega, q) \in \mathcal{S}$ is given by

$$a^* = \arg\max_{a \in \mathcal{A}_{\mathbf{s}}} \left\{ \bar{u}(\mathbf{s}, a) + \delta \sum_{\omega' \in \Omega} p(\omega'|\omega, a) \sum_{q' \in \mathcal{Q}} p(q'|q, a)\, \upsilon^*_{\omega', q'} \right\}, \tag{5}$$

where υ^* is the optimal continuation value. The optimal action is non-myopic and balances the immediate expected payoff $\bar{u}(\mathbf{s}, a)$ with the expected future stream of discounted payoffs. Hence, the optimal action takes into account how the current action will affect the transition to subsequent states $\mathbf{s}$, and therefore accounts for how the current action will affect future actions.

To gain some insight into the optimal strategy, the following lemma shows that the optimal continuation value υ^* decreases in value as the level of inventory deviates further away from zero. Therefore, the optimal continuation value υ^* achieves its maximum value at zero inventory.

Lemma 1 *For $q \geq 0$, the optimal continuation value $\upsilon^*_{\omega,q}$ is non-increasing as q increases, i.e., when $0 \leq q' \leq q$, we have $\upsilon^*_{\omega,q} \leq \upsilon^*_{\omega,q'}$ for all $\omega \in \Omega$. Similarly, for $q \leq 0$, the optimal continuation value $\upsilon^*_{\omega,q}$ is non-decreasing as q increases, i.e., when $q' \leq q \leq 0$, we have $\upsilon^*_{\omega,q'} \leq \upsilon^*_{\omega,q}$ for all $\omega \in \Omega$.*

The result is intuitive because the market maker's aversion to higher levels of inventory and because the fundamental value of the asset is a martingale. One implication is that the preferred level of inventory is zero because it leads to the highest expected stream of discounted payoffs; hence, optimal strategies will induce mean reversion to zero inventory. The concavity of the optimal continuation value $\upsilon^*_{\omega,q}$ as a function of inventory holds for each volume imbalance regime ω. The relationship between the optimal continuation value $\upsilon^*_{\omega,q}$ and the volume imbalance regimes is key for spoofing to be optimal.

The following lemma eliminates suboptimal actions, which reduces the number of actions to consider when solving (5).

Lemma 2 *For all volume imbalance regimes $\omega \in \Omega$, the following two statements hold. The actions do nothing (i.e., DN) and market buy order (i.e., MB) are suboptimal if $q > 0$. The actions do nothing (i.e., DN) and market sell order (i.e., MS) are suboptimal if $q < 0$.*

The following proposition characterizes the optimal action for each state $\mathbf{s} = (\omega, q)$ when $q \neq 0$. The optimal action is characterized in terms of the value of the inventory aversion parameter α, i.e., the willingness to hold inventory. Figure 2 shows the optimal action as a function of the

value of the inventory aversion parameter α when $q > 0$. For each pair of neighboring actions in Figure 2, there is a cutoff value of α such that the optimal strategy is indifferent between the two actions because they yield the same expected stream of discounted payoffs. Hence, for a given value of α, the optimal action is one that maximizes the expected stream of discounted payoffs.

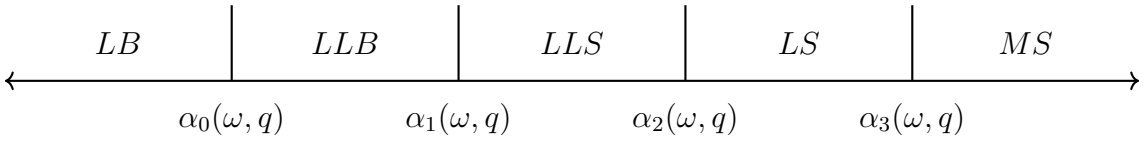


Figure 2: Optimal action choice for each state $\mathbf{s} = (\omega, q)$ for $q > 0$.

For the remainder of the paper, we denote $x \vee y = \max\{x, y\}$, and $x \wedge y = \min\{x, y\}$.

Proposition 1 *Let $q > 0$ and assume $p^a_{SH} < p^a_N < p^a_{BH}$ holds. Then, for each state $\mathbf{s} = (\omega, q)$, there exist cutoff values of the inventory aversion parameter $\alpha_0(\omega, q) < \alpha_1(\omega, q) < \alpha_2(\omega, q) < \alpha_3(\omega, q)$ such that the optimal stationary pure Markov strategy $\sigma^* \in \Sigma^{SPM}$ is given by*

$$\sigma^*(\omega, q) = \begin{cases} LB & \text{if } \alpha \in \left(0, 0 \vee \alpha_0(\omega, q)\right), \\ LLB & \text{if } \alpha \in \left(0 \vee \alpha_0(\omega, q), 0 \vee \alpha_1(\omega, q)\right), \\ LLS & \text{if } \alpha \in \left(0 \vee \alpha_1(\omega, q), 0 \vee \alpha_2(\omega, q)\right), \\ LS & \text{if } \alpha \in \left(0 \vee \alpha_2(\omega, q), 0 \vee \alpha_3(\omega, q)\right), \\ MS & \text{if } \alpha \in \left(0 \vee \alpha_3(\omega, q), +\infty\right). \end{cases}$$

Similarly, let $q < 0$ and assume $p^b_{SH} > p^b_N > p^b_{BH}$ holds. Then, for each state $\mathbf{s} = (\omega, q)$, there exist cutoff values of the inventory aversion parameter $\alpha_0(\omega, q) < \alpha_1(\omega, q) < \alpha_2(\omega, q) < \alpha_3(\omega, q)$ such that the optimal stationary pure Markov strategy $\sigma^* \in \Sigma^{SPM}$ is given by

$$\sigma^*(\omega, q) = \begin{cases} LS & \text{if } \alpha \in \left(0, 0 \vee \alpha_0(\omega, q)\right), \\ LLS & \text{if } \alpha \in \left(0 \vee \alpha_0(\omega, q), 0 \vee \alpha_1(\omega, q)\right), \\ LLB & \text{if } \alpha \in \left(0 \vee \alpha_1(\omega, q), 0 \vee \alpha_2(\omega, q)\right), \\ LB & \text{if } \alpha \in \left(0 \vee \alpha_2(\omega, q), 0 \vee \alpha_3(\omega, q)\right), \\ MB & \text{if } \alpha \in \left(0 \vee \alpha_3(\omega, q), +\infty\right). \end{cases}$$

In general, the cutoff values $\alpha_0(\omega, q)$, $\alpha_1(\omega, q)$, and $\alpha_3(\omega, q)$ are different for positive and negative inventory. However, the cutoff values are the same for positive and negative of inventory in the particular case when $p^a_{BH} = p^b_{SH}$, $p^a_{SH} = p^b_{BH}$, and $p^a_N = p^b_N$, and when the transition

probability matrix for the Markov chain given in Figure 1b is a centrosymmetric matrix (i.e., a matrix that is symmetric about its center).

For $q > 0$ and $\alpha = \alpha_0(\mathbf{s})$, the optimal strategy is indifferent between a buy limit order and a large buy limit order in state $\mathbf{s}$; similarly, for $q < 0$ and $\alpha = \alpha_0(\mathbf{s})$, the optimal strategy is indifferent between a sell limit order and a large sell limit order in state $\mathbf{s}$. A similar interpretation applies to the remaining cutoff values $\alpha_1(\mathbf{s})$, $\alpha_2(\mathbf{s})$, and $\alpha_3(\mathbf{s})$. The conditions in the proposition include the maximum operator because the inventory aversion parameter is strictly positive and there is no guarantee that the cutoff values are positive. The conditions on the fill probabilities are not restrictive because they hold for all assets and all the timescales considered (see Table 7).

The proposition is intuitive because the optimal strategy induces mean reversion to zero inventory. For example, when there is less willingness to hold larger values of inventory (i.e., for large values of α), the action preference favors actions that aim to sell the asset and reduce the level of inventory when $q > 0$, and the action preference favors actions that aim to buy the asset and increase the level of inventory when $q < 0$. Similarly, for a fixed value of α, as the level of inventory deviates away from zero, the cutoff value $\alpha_1(\omega, q)$ shifts closer to zero (the value of $\bar{\alpha}_1(\omega, q)$ in (6) decreases as the absolute value of q increases). Hence, the action preference favors actions that aim at selling the asset to reduce the level of inventory when $q > 0$, and that aim at buying the asset to increase the level of inventory when $q < 0$.

4.2. Quote-Based Manipulation

To derive the conditions to test if algorithms will learn to manipulate the order book, we first define quote-based manipulation in terms of the set of actions available. Recall that an explicit consideration of our model is that the market maker does not endow her algorithm with an individual action that manipulates the book; rather, manipulation occurs in our model when a particular combination of actions is sequenced together. The following definition includes a refinement that also accounts for the market maker's objective to buy or to sell the asset.

Definition 1 (Manipulation) *For $q > 0$, quote-based manipulation occurs if a large buy limit order is placed at time t, and it is followed at time $t + 1$ by a unit or large sell limit order. Similarly, for $q < 0$, quote-based manipulation occurs if a large sell limit order is placed at time t, and it is followed at time $t + 1$ by a unit or large buy limit order.*

The sequence of actions in Definition 1 is manipulative because the optimal strategy intends to revert to the preferred inventory position of zero, so sending a large buy limit order when the market maker is long or a large sell limit order when the market maker is short counters

this objective. Therefore, if these sequences arise in optimality, then the market maker must be profiting through this manipulative order by tilting the book in her favor for future actions.

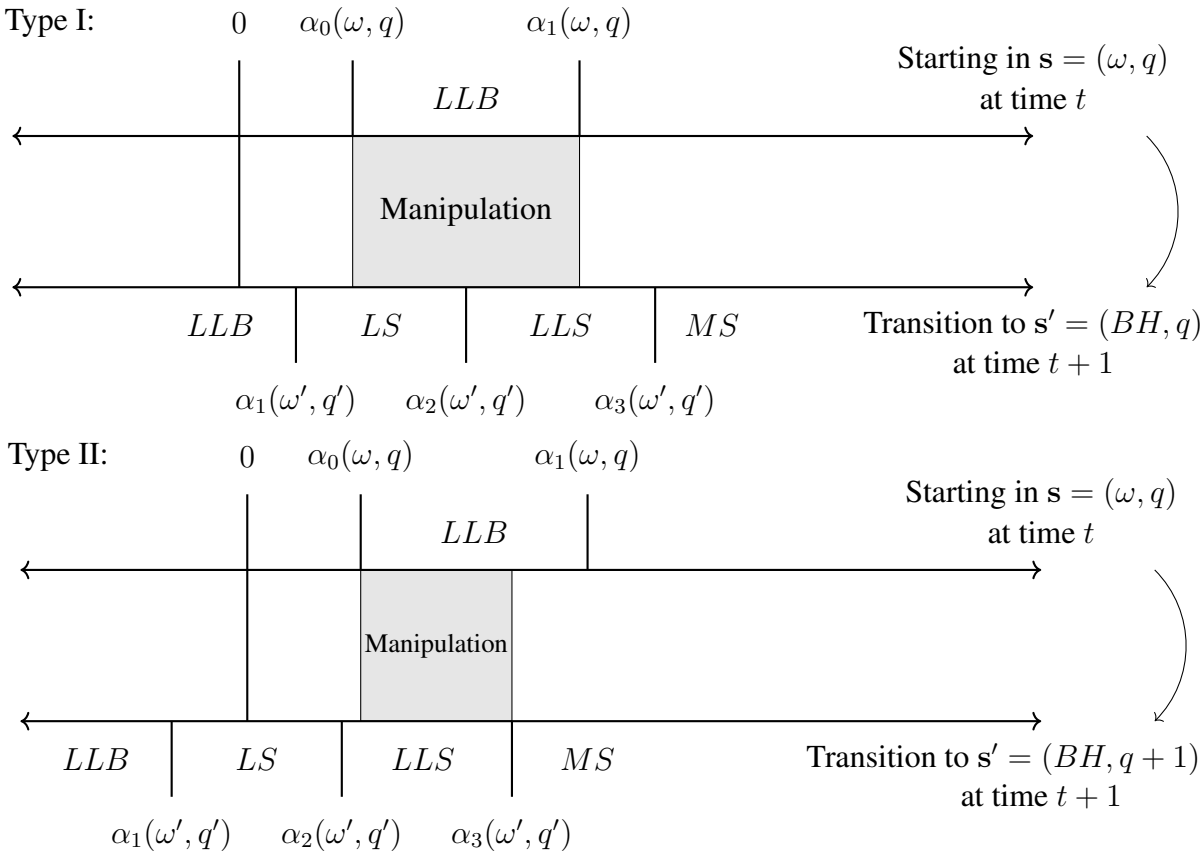


Figure 3: Quote-based manipulation is optimal when the value of the inventory aversion parameter α lies within the shaded region for $q > 0$. The top panel describes type I manipulation, and the bottom panel describes type II manipulation.

One consequence of quote-based manipulation is that the manipulative order may get caught out and lead to a transaction. We distinguish this possibility into two cases. Type I manipulation occurs when the manipulative order is not caught out, i.e., it is not filled. Specifically, type I manipulation is the sequence initiated by LLB in state $\mathbf{s} = (\omega, q)$ and followed by LS or LLS in state $\mathbf{s}' = (BH, q)$ when $q > 0$, or the sequence initiated by LLS in state $\mathbf{s} = (\omega, q)$ and followed by LB or LLB in state $\mathbf{s}' = (SH, q)$ when $q < 0$. On the other hand, type II manipulation occurs when the manipulative order is caught out, i.e., it gets filled. Specifically, type II manipulation is the sequence initiated by LLB in state $\mathbf{s} = (\omega, q)$ and followed by LS or LLS in state $\mathbf{s}' = (BH, q + 1)$ when $q > 0$, or the sequence initiated by LLS in state $\mathbf{s} = (\omega, q)$ and followed by LB or LLB in state $\mathbf{s}' = (SH, q - 1)$ when $q < 0$.

Figure 3 illustrates the two types of manipulation when $q > 0$. Here, manipulation occurs if the

value of the inventory aversion parameter α is within the shaded region. We say that manipulation occurs in state $\mathbf{s} = (\omega, q)$ if the optimal action is to submit a large buy limit order in state $\mathbf{s} = (\omega, q)$, and when the market transitions to the subsequent state $\mathbf{s}'$ the optimal strategy prescribes to submit either a sell limit order or a large sell limit order.

We formalize the definition of quote-based manipulation in the context of an optimal strategy. We adopt the standard convention that the interval $(x, y) = \emptyset$ if $x \geq y$, and recall that $x \vee y = \max\{x, y\}$, and $x \wedge y = \min\{x, y\}$.

Definition 2 (Manipulative Strategy) *If there exists a state* $\mathbf{s} = (\omega, q)$ *such that*

(i) $$\emptyset \neq I_1(\mathbf{s}) = \begin{cases} \Big(0 \vee \alpha_0(\omega, q) \vee \alpha_1(BH, q), \big(0 \vee \alpha_1(\omega, q)\big) \wedge \big(0 \vee \alpha_3(BH, q)\big)\Big) & \text{if } q > 0\,, \\ \Big(0 \vee \alpha_0(\omega, q) \vee \alpha_1(SH, q), \big(0 \vee \alpha_1(\omega, q)\big) \wedge \big(0 \vee \alpha_3(SH, q)\big)\Big) & \text{if } q < 0\,, \end{cases}$$

or

(ii) $$\emptyset \neq I_2(\mathbf{s}) = \begin{cases} \Big(0 \vee \alpha_0(\omega, q) \vee \alpha_1(BH, q+1), \big(0 \vee \alpha_1(\omega, q)\big) \wedge \big(0 \vee \alpha_3(BH, q+1)\big)\Big) & \text{if } q > 0\,, \\ \Big(0 \vee \alpha_0(\omega, q) \vee \alpha_1(SH, q-1), \big(0 \vee \alpha_1(\omega, q)\big) \wedge \big(0 \vee \alpha_3(SH, q-1)\big)\Big) & \text{if } q < 0\,. \end{cases}$$

Then, if the value of the inventory aversion parameter $\alpha \in I_1(\mathbf{s})$*, the optimal strategy is a manipulative strategy, where type I manipulation occurs in states* $\mathbf{s}$ *where condition (i) is satisfied. Similarly, if the value of the inventory aversion parameter* $\alpha \in I_2(\mathbf{s})$*, the optimal strategy is a manipulative strategy, where type II manipulation occurs in states* $\mathbf{s}$ *where condition (ii) is satisfied.*

The intervals $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$ describe conditions for quote-based manipulation to be dynamically optimal as a sequence of actions. Specifically, if $\alpha \in I_1(\mathbf{s})$ or $\alpha \in I_2(\mathbf{s})$, then manipulation occurs in state $\mathbf{s}$; if $\alpha \in I_1(\mathbf{s})$, then manipulation occurs in state $\mathbf{s}$ when the manipulative order is not filled; if $\alpha \in I_2(\mathbf{s})$, then manipulation occurs in state $\mathbf{s}$ when the manipulative order is filled; and if $\alpha \in I_1(\mathbf{s}) \cap I_2(\mathbf{s})$, then manipulation occurs in state $\mathbf{s}$ regardless of whether the manipulative order was filled or not.

4.3. Testable Conditions

To obtain testable conditions that apply to a wide variety of learning algorithms, we make the following assumption.

Assumption A *The learning algorithm used by the market maker learns an optimal stationary pure Markov strategy* $\sigma^* \in \Sigma^{SPM}$.

This assumption is not restrictive and allows us to analyze the framework where algorithms learn optimal strategies so that the testable conditions we derive apply to generic learning algo-

rithms. The premise and objective of designing a learning algorithm is to learn an optimal stationary pure Markov strategy. Indeed, the most popular offline learning algorithms (such as policy iteration and value iteration) and online learning algorithms (such as Q-learning and SARSA) satisfy this assumption (see for example Puterman, 1994; Sutton and Barto, 2018).[15]

Clearly, when Assumption A and the conditions for manipulation in Definition 2 hold, the algorithm will learn a manipulative strategy. The issue is that the intervals $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$ may be empty. Additionally, the intervals depend on the cutoff values, which in turn, depend on the parameters of the model and on the optimal continuation value v^*. The following theorem provides sufficient conditions to determine if algorithms will learn manipulative strategies. These conditions depend only on the fill probabilities of the limit orders.

Theorem 1 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. If the conditions*

$$p^b_{BH} < p^a_{BH} \tag{C1}$$

and

$$p^a_{SH} < p^b_{SH} \tag{C2}$$

hold, then $I_1(\mathbf{s}) \neq \emptyset$ and $I_2(\mathbf{s}) \neq \emptyset$ for all $\mathbf{s} \in \mathcal{S}$ such that (i) $\mathbf{s} = (SH, q > 0)$, (ii) $\mathbf{s} = (BH, q < 0)$, and (iii) $\mathbf{s} = (N, q)$ for either $q > 0$ or $q < 0$.

These testable conditions have strong implications summarized in the following corollary.

Corollary 1 *Let the following hold: (i) Assumption A, (ii) $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$, and (iii) (C1) and (C2). Then, for any state $\mathbf{s} = (\omega, q)$ outlined in Theorem 1, there exist values of the inventory aversion parameter α for which the algorithm will learn a type I manipulation strategy in state $\mathbf{s}$; similarly, there exist values of the inventory aversion parameter α for which the algorithm will learn a type II manipulation strategy in state $\mathbf{s}$.*

In short, if the fill probabilities satisfy certain conditions, then there are values of the inventory aversion parameter α where an algorithm will learn a manipulative strategy in which type I and/or type II manipulation occurs in state $\mathbf{s}$. Whether both types or only one type is learned depends on the value of the inventory aversion parameter α and if the intervals $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$ overlap.

The role of the conditions in Theorem 1 is intuitive. The conditions for the fill probabilities of the ask and of the bid are so that the ordering of action preferences in Proposition 1 hold. For $q > 0$,

[15] See Watkins and Dayan (1992) for a convergence proof of Q-learning and Singh et al. (2000) for a convergence proof of SARSA. See also Hambly et al. (2023), Ning et al. (2021), and Spooner et al. (2018) for applications of learning algorithms in financial markets that satisfy this assumption.

condition (C1) ensures that buy limit orders are not optimal when the book is buy-heavy, while condition (C2) ensures that there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence with a large buy limit order in sell-heavy. Similarly, for $q < 0$, condition (C2) ensures that sell limit orders are not optimal when the book is sell-heavy, while condition (C1) ensures that there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence with a large sell limit order in buy-heavy.

For the neutral regime, conditions (C1) and (C2) ensure that the manipulative sequence will be completed for positive inventory and negative inventory, respectively. Although we show that there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence, we do not know if the manipulative sequence is initiated with a large buy limit order for $q > 0$, or if the manipulative sequence is initiated with a large sell limit order for $q < 0$. The following theorem imposes stronger conditions to resolve this indeterminacy.

Theorem 2 *Let* $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$, (C1) *and* (C2) *hold, and let*

$$\begin{aligned} p^a_{BH} - p^b_{SH} &< \min\left\{(p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{p_{N|BH}}{p_{BH|BH}}\,,\ (p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{p_{N|N}}{p_{BH|N}}\right\} \\ p^b_{SH} - p^a_{BH} &< \min\left\{(p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{p_{N|SH}}{p_{SH|SH}}\,,\ (p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{p_{N|N}}{p_{SH|N}}\right\} \end{aligned} \tag{C3}$$

hold.

2.1 If $(p^b_N - p^a_N) > \dfrac{\delta}{1+\delta}(p^b_{SH} - p^a_{BH})$ *holds, then* $I_1(\mathbf{s}) \neq \emptyset$ *and* $I_2(\mathbf{s}) \neq \emptyset$ *for all states* $\mathbf{s} = (N, q > 0)$.

2.2 If $(p^a_N - p^b_N) > \dfrac{\delta}{1+\delta}(p^a_{BH} - p^b_{SH})$ *holds, then* $I_1(\mathbf{s}) \neq \emptyset$ *and* $I_2(\mathbf{s}) \neq \emptyset$ *for all states* $\mathbf{s} = (N, q < 0)$.

Condition (C3) is a formal condition to describe that the values of the fill probabilities p^a_{BH} and p^b_{SH} are similar, i.e., $p^a_{BH} \approx p^b_{SH}$. This condition simplifies the analysis to determine if the manipulative sequence is initiated in the neutral regime with a large buy limit order for $q > 0$, or if the manipulative sequence is initiated with a large sell limit order for $q < 0$. On the other hand, the conditions $(p^b_N - p^a_N) > \frac{\delta}{1+\delta}(p^b_{SH} - p^a_{BH})$ and $(p^a_N - p^b_N) > \frac{\delta}{1+\delta}(p^a_{BH} - p^b_{SH})$ describe the condition to determine if manipulation occurs when inventory is long or short in the neutral regime. These conditions are such that one inequality will always hold, but never both or neither.

One can think of the fill probabilities in the neutral regime as the short-term incentives associated with the immediate payoffs, and the fill probabilities in the heavy regimes as the long-term

incentives associated with discounted future payoffs. The intuition behind these conditions is clear. If the signs of $p_N^b - p_N^a$ and $p_{SH}^b - p_{BH}^a$ are different, then the short-term and long-term incentives align. On the other hand, if the signs of $p_N^b - p_N^a$ and $p_{SH}^b - p_{BH}^a$ are the same, then the short-term and long-term incentives are not aligned, so the determining factor is the tradeoff between short-term and long-term incentives.

For example, for manipulation to occur in the neutral regime for $q > 0$, it must be that there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence with a large buy limit order. If $p_N^b > p_N^a$ and $p_{SH}^b < p_{BH}^a$, then the signs of $p_N^b - p_N^a$ and $p_{SH}^b - p_{BH}^a$ are different, so the short-term and long-term incentives align because the immediate payoff from a buy limit order is greater than that of a sell limit order, and the future payoffs from being in buy-heavy (proxied with p_{BH}^a) are better than the future payoffs from being in sell-heavy (proxied with p_{SH}^b). Thus, it is clear that there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence with a large buy limit order because the incentives align. On the other hand, if $p_N^b > p_N^a$ and $p_{SH}^b > p_{BH}^a$, then the signs of $p_N^b - p_N^a$ and $p_{SH}^b - p_{BH}^a$ are the same, so the short-term and long-term incentives are not aligned. In this case, if the immediate payoff outweighs the discounted future payoffs, then there are values of the inventory aversion parameter α for which it is optimal to initiate the manipulative sequence with a large buy limit order.

Both Theorems 1 and 2 are sufficient (but not necessary) conditions for the intervals $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$ to exist. Hence, the theorems provide conditions to test when a manipulative strategy could be optimal; however, failure to satisfy the conditions does not mean that a manipulative strategy cannot be optimal. Moreover, the theorems do not specify the values of the inventory aversion parameter α for which a manipulative strategy is optimal. To narrow the search for values of the inventory aversion parameter α where an algorithm will learn to manipulate the book, we derive an interval $I'(\mathbf{s})$ that contains both $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$, which uses the following upper bound.

Lemma 3 *Let $m \in (0,1)$ be the minimum element of the transition probability matrix for the Markov chain given in Figure 1b. Then for all $\omega, \omega' \in \Omega$ and $q \in \mathcal{Q}$, we have $\upsilon^*_{\omega,q} - \upsilon^*_{\omega',q} \leq \vartheta/m$.*

With this lemma, the following proposition characterizes the interval $I'(\mathbf{s})$ that contains both $I_1(\mathbf{s})$ and $I_2(\mathbf{s})$, and does not depend on the optimal continuation value υ^*.

Proposition 2 *Let $p_{SH}^a < p_N^a < p_{BH}^a$ and $p_{SH}^b > p_N^b > p_{BH}^b$ hold. For all $\mathbf{s} = (\omega, q)$ when $q \neq 0$, the interval $I'(\mathbf{s}) = \big(0, \bar{\alpha}_1(\omega, q)\big)$ is such that $I_1(\mathbf{s}) \subset I'(\mathbf{s})$ and $I_2(\mathbf{s}) \subset I'(\mathbf{s})$.*

The interval $I'(\mathbf{s})$ is characterized in terms of the upper bound of the cutoff value $\alpha_1(\omega, q)$

given by

$$\alpha_1(\omega, q) \leq \bar{\alpha}_1(\omega, q) = \begin{cases} \Big[\left(p^b_\omega - p^a_\omega\right) \vartheta/2 + \vartheta/m \Big] \Big[p^a_\omega \left(2\,q - 1\right) + p^b_\omega \left(2\,q + 1\right) \Big]^{-1} & \text{if } q > 0\,, \\ \Big[\left(p^a_\omega - p^b_\omega\right) \vartheta/2 + \vartheta/m \Big] \Big[- p^a_\omega \left(2\,q - 1\right) - p^b_\omega \left(2\,q + 1\right) \Big]^{-1} & \text{if } q < 0\,, \end{cases} \tag{6}$$

for all $\omega \in \Omega$. The value of $\bar{\alpha}_1(\omega, q)$ is strictly positive, depends only on parameters of the model, and is easy to compute. The interval $I'(\mathbf{s})$ narrows the search for values of the inventory aversion parameter α for which an algorithm will learn a manipulative strategy.

Indeed, $\alpha \in I'(\mathbf{s})$ is a necessary but not a sufficient condition for an algorithm to learn a manipulative strategy in state $\mathbf{s}$. Specifically, not all values of $\alpha \in I'(\mathbf{s})$ will lead to manipulation in state $\mathbf{s}$ because not all values of $\alpha \in I'(\mathbf{s})$ lie within $I_1(\mathbf{s})$ or $I_2(\mathbf{s})$. Nonetheless, if $\alpha \notin I'(\mathbf{s})$, then an algorithm cannot learn a strategy that manipulates the order book in state $\mathbf{s}$. In the next section, we use this condition to analyze how parameters of the model affect an algorithm's ability to learn to manipulate the order book.

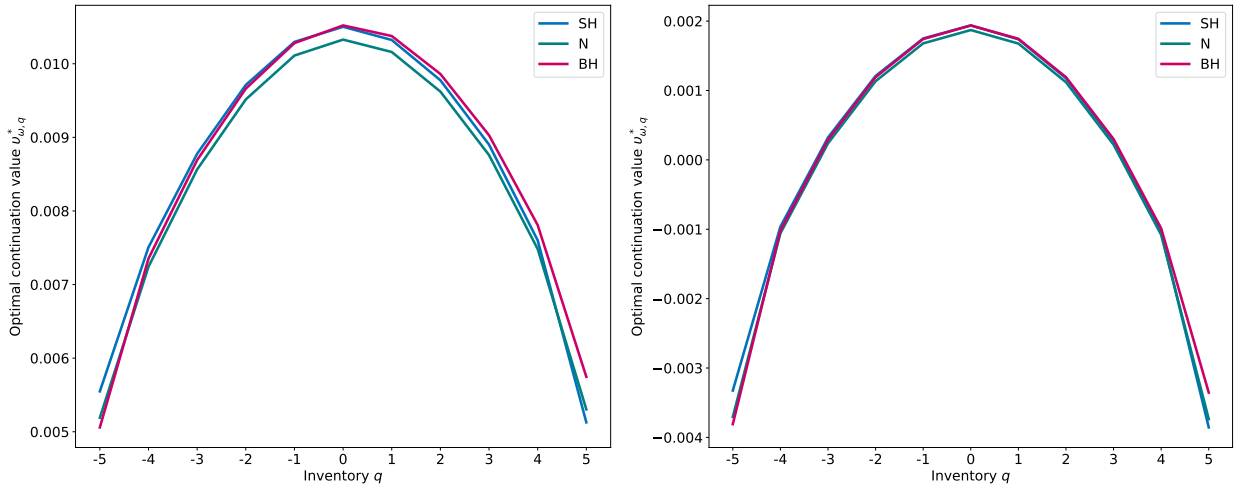
5. Understanding Manipulation and Spoofing

This section analyzes the mechanics of what makes quote-based manipulation dynamically optimal and how parameters of the model affect the optimality of manipulative strategies.

5.1. Workings of Manipulation

Mechanics of Manipulation. Figure 4 illustrates the mechanics behind manipulation in optimality with data from AMZN and CSCO at 0.5 second decision intervals. The model parameters are estimated with the dataset from Section 2; the estimation procedure is discussed below in Section 6. With a discount factor of $\delta = 0.95$, we solve for the optimal strategy with the policy iteration algorithm and solve for the optimal continuation values $v^*_{\omega,q}$ for each volume imbalance regime as a function of the inventory level. The inventory aversion parameter is $\alpha = 10^{-5}$.

For each volume imbalance regime, there is a gravitational pull towards zero inventory because $v^*_{\omega,q}$ achieves its maximum at $q = 0$. However, the optimal continuation values $v^*_{\omega,q}$ as a function of inventory q differs based on the volume imbalance regimes, i.e., there is asymmetry in the volume imbalance regimes. This asymmetry is key for quote-based manipulation to arise. Specifically, consider AMZN at $\omega = SH$ with $q = 1$, and focus only on the expected discounted future payoffs. If the market maker tries to revert to zero inventory with a sell limit order, then the expected discounted future payoffs is an expectation across the optimal continuation values $v^*_{\omega,q}$ for all regimes and for inventory levels $q = 0$ and $q = 1$. On the other hand, if the market maker uses a manipulative order (a large buy limit order), then the expected discounted future payoffs is an



(a) AMZN (b) CSCO

Figure 4: Optimal continuation values $v^*_{\omega,q}$.

expectation across the optimal continuation values $v^*_{\omega,q}$ for buy-heavy and for inventory levels $q = 1$ and $q = 2$. This difference in the expected discounted future payoffs induced by the asymmetry in the volume imbalance regimes is a key factor that makes a manipulative strategy dynamically optimal. Specifically, if this difference outweighs the gravitational pull to zero inventory and the immediate payoffs, then a manipulative strategy becomes dynamically optimal.

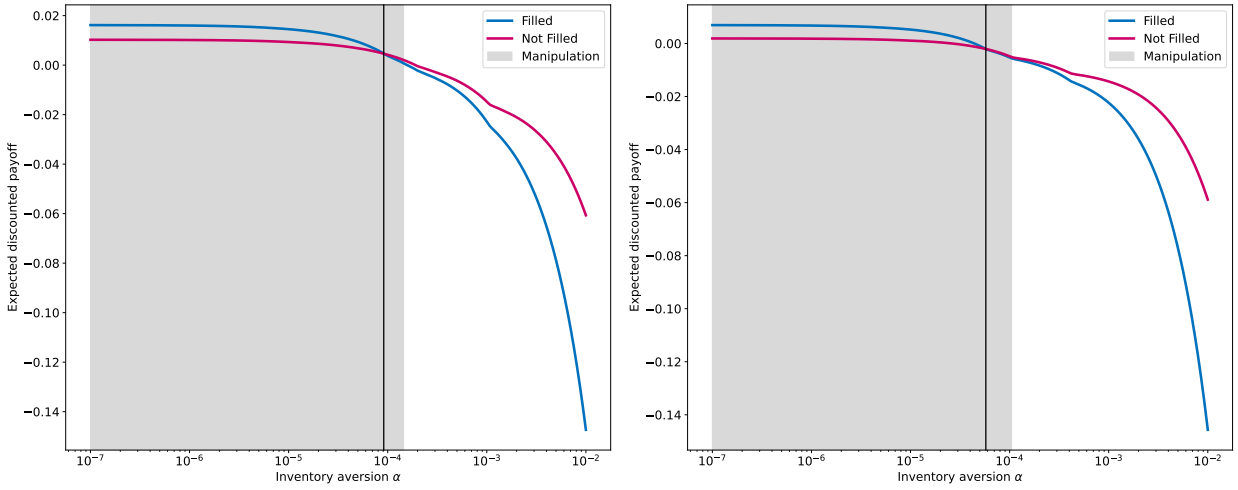
The concavity of these curves decreases as the value of the inventory aversion parameter α decreases. A decrease in the concavity increases the difference in the expected discounted future payoffs induced by the asymmetry; hence, quote-based manipulation is more likely to become dynamically optimal as the market maker becomes more tolerant to bearing inventory risk.

Spoofing and Fill Preferences. Counter-intuitive to the motivation to manipulate the book, we find that getting caught out with a manipulative order is not always suboptimal. Indeed, there are situations in which manipulation occurs with the preference for the manipulative order to get filled. To show this, we analyze if the market maker prefers that her manipulative order is caught out, or if she prefers that her manipulative order does not get caught out. We compute

$$v_{\mathbf{s}}(LLB, \text{filled}) = \mathbb{E}_{\sigma^*}\left[\sum_{t=0}^{\infty} \delta^t \, u(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) \,\middle|\, \mathbf{s}_0 = \mathbf{s}, a_0 = LLB, q_1 = q_0 + 1\right],$$

$$v_{\mathbf{s}}(LLB, \text{not filled}) = \mathbb{E}_{\sigma^*}\left[\sum_{t=0}^{\infty} \delta^t \, u(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) \,\middle|\, \mathbf{s}_0 = \mathbf{s}, a_0 = LLB, q_1 = q_0\right],$$

which corresponds to the expected stream of discounted payoffs conditional on a manipulative order getting filled or not for $q > 0$. If $\upsilon_{\mathbf{s}}(LLB, \text{filled}) > \upsilon_{\mathbf{s}}(LLB, \text{not filled})$, then the market maker prefers that her manipulative order is caught out. Conversely, if $\upsilon_{\mathbf{s}}(LLB, \text{filled}) < \upsilon_{\mathbf{s}}(LLB, \text{not filled})$, then the market prefers that her manipulative order does not get caught out.



(a) AMZN (b) CSCO

Figure 5: Expected stream of discounted payoffs when the manipulative order is filled or not for $\mathbf{s} = (SH, q = 2)$.

Figure 5 illustrates this preference. The shaded region denotes the values of the inventory aversion parameter where manipulation is optimal, i.e., $I_1(\mathbf{s}) \cup I_2(\mathbf{s})$. The vertical black line denotes the cutoff value of α for which the preference switches. Within the shaded area, when α is to the left of the vertical line, the market maker prefers that her manipulative order is caught out; and to the right of the vertical line, the market prefers that her manipulative order does not get caught out.

A market maker who prefers a fill of her manipulative order may seem counter-intuitive when the motivation to manipulate the book is to manage inventory risk and revert to the preferred inventory position. However, managing inventory risk is only one part of the optimization problem to determine if manipulation is optimal. The market maker may prefer that her manipulative order is filled because the manipulative order increases the probability to complete a round-trip trade; i.e., the additional profit from the round-trip trade outweighs the costs to manage inventory risk.

Based on the fill preferences, we further refine quote-based manipulation into two forms: manipulation for a round-trip trade where manipulation occurs with the preference that the manipulative order is filled, and spoofing where manipulation occurs with the preference that the manipulative order is not filled. This refinement is different from type I and type II manipulation, which

determines the optimality of the manipulation sequence. The refinement of fill preferences allow us to determine the intention behind a market makers manipulative order, which allows us to make the distinction between spoofing and manipulation for a round-trip trade.

One might argue that manipulation for a round-trip trade is not market manipulation but rather a byproduct of making markets. In our model, we assume (based on empirical results) that at most one unit of the large limit order can be executed. We make this assumption so that the market maker has the same expected one-step utility if she sends a limit order for one unit. Therefore, if she were to make markets without manipulating the book, then she would not use a large limit order. Given that a limit order for one unit (in the same direction as the manipulative order) is never suboptimal (see Figure 6 in Appendix B), it means that the market maker only sends a large limit order to tilt the book to complete a round-trip trade that, in expectation, will be completed faster than otherwise.

5.2. Model Parameters and Manipulation

Building on the insights behind the driving forces of quote-based manipulation, we formalize how parameters of the model affect an algorithm's ability to learn to manipulate the order book. The results rely on Proposition 2 so that if $\alpha \notin I'(\mathbf{s})$, then an algorithm cannot learn a strategy that manipulates the order book in state $\mathbf{s}$.

Inventory Aversion. The following proposition shows that if the market maker is sufficiently averse to holding larger levels of inventory, then their algorithm will not learn to manipulate the order book.

Proposition 3 *Let Assumption A hold and let* $p^a_{SH} < p^a_N < p^a_{BH}$ *and* $p^b_{SH} > p^b_N > p^b_{BH}$ *hold. If the market maker's inventory aversion parameter is such that*

$$\alpha > \max_{\omega \in \Omega,\, q \in \{-1,1\}} \{\bar{\alpha}_1(\omega, q)\}\,, \tag{7}$$

then the algorithm will not learn to manipulate the order book for any state $\mathbf{s} = (\omega, q \neq 0)$.

For a fixed volume imbalance regime ω, the upper bound of the cutoff $\bar{\alpha}_1(\omega, q)$ decreases monotonically as the absolute value of q increases. When the value of the inventory aversion parameter α satisfies (7), then $\alpha \notin I'(\mathbf{s})$ for all states $\mathbf{s} = (\omega, q)$ where $q \neq 0$, and the algorithm will not learn a manipulative strategy for all $q \neq 0$.

Conversely, if the value of the inventory aversion parameter does not satisfy the inequality in (7), then $\alpha \in I'(\mathbf{s})$ for some states $\mathbf{s} = (\omega, q)$ where $q \neq 0$. Here, there is a possibility that an algorithm will learn a manipulative strategy, but it is not guaranteed.

The result is intuitive if we consider the factors that make a manipulative strategy dynamically optimal. When initiating the manipulative sequence with a large limit order, there is a possibility that the large limit order is filled. Therefore, if the market maker is sufficiently averse to holding larger levels of inventory, then the cost associated with a manipulative order getting filled is too high for manipulation to be optimal; thus, the algorithm will not learn to manipulate the order book.

Quoted Spread. The following proposition shows how the expected quoted spread affects an algorithm's ability to learn to manipulate the order book.

Proposition 4 *Let Assumption A hold and let* $p^a_{SH} < p^a_N < p^a_{BH}$ *and* $p^b_{SH} > p^b_N > p^b_{BH}$ *hold. If the expected quoted spread* $\vartheta \to 0$*, then the algorithm will not learn to manipulate the order book for any state* $\mathbf{s} = (\omega, q)$ *where* $q \neq 0$*.*

The result follows because $\bar{\alpha}_1(\omega, q) \to 0$ for all $\omega \in \Omega$ and $q \neq 0$ as $\vartheta \to 0$. This ensures that for $\alpha > 0$, we have $\alpha \notin I'(\mathbf{s})$ for all states $\mathbf{s} = (\omega, q)$ where $q \neq 0$, so the algorithm will not learn a manipulative strategy.

Theory shows that the quoted bid-ask spread will not be zero even if the tick size is zero (i.e., $\varphi = 0$) because market makers must recover inventory costs (e.g., Stoll, 1978; Ho and Stoll, 1981), and losses due to asymmetric information (e.g., Copeland and Galai, 1983; Glosten and Milgrom, 1985; Glosten, 1994). Nonetheless, the proposition demonstrates the relationship between the expected quoted spread and quote-based manipulation. As the expected quoted spread decreases, the range of values of the inventory aversion parameter for which manipulation is optimal decreases. Conversely, as the expected quoted spread increases, the range of values of the inventory aversion parameter for which manipulation is optimal increases.

The result is also intuitive when one analyzes the factors that make manipulation optimal. First, as the expected quoted spread decreases, the gains from using limit orders and the costs from using market orders become negligible, so it is more efficient to revert to the preferred inventory position using market orders because they guarantee execution. Therefore, the uncertainty of a limit order execution and the possibility that a manipulative order is caught out makes a manipulative strategy suboptimal. Finally, as the expected quoted spread decreases, the expected profit from round-trip trades also decreases. The decrease in the profits does not outweigh the costs required to manage the inventory risk, and hence a manipulative strategy is suboptimal.

6. Empirical Estimation

This section uses Nasdaq data to test the conditions derived in Section 4. We discuss the estimation procedure for the parameters of our model, and we use these estimates to determine if market conditions from Nasdaq are conducive for an algorithm to learn to manipulate the order book. We use the dataset from Section 2.

Table 8: Summary statistics for April 2023.

Ticker	Decision interval Δt	Ave. spread (ticks)	Ave. queue size best bid	Ave. queue size best ask	Ave. volume traded per Δt	Ave. volume imbalance
AAPL	5 seconds	1.168	583	600	1217	0.008
	1 second	1.167	583	600	243	0.006
	0.5 seconds	1.168	584	601	122	0.006
AMZN	5 second	1.205	532	572	1146	-0.027
	1 seconds	1.206	532	571	229	-0.028
	0.5 seconds	1.206	533	571	115	-0.027
CSCO	5 seconds	1.005	2088	2046	488	0.012
	1 second	1.006	2100	2060	98	0.012
	0.5 seconds	1.011	2120	2078	49	0.012
INTC	5 seconds	1.005	3378	3517	975	-0.014
	1 second	1.005	3385	3530	195	-0.016
	0.5 seconds	1.006	3405	3557	97	-0.017
MSFT	5 seconds	1.783	114	119	746	-0.021
	1 second	1.784	114	119	149	-0.022
	0.5 seconds	1.788	114	119	75	-0.022
TSLA	5 seconds	2.231	190	200	2089	-0.01
	1 second	2.235	195	198	418	-0.01
	0.5 seconds	2.232	194	198	209	-0.009

Table 8 provides summary statistics for seven assets at three different decision intervals Δt: 5 seconds, 1 second, and 0.5 seconds. We estimate the statistics by sampling the relevant features at every decision interval.

6.1. Estimation Procedure

In our model, there are two sets of model parameters to estimate: the transition probabilities of the volume imbalance regime $p_{\omega'|\omega}$ for all $\omega, \omega' \in \Omega$, and the fill probabilities in each volume imbalance regime p^a_ω and p^b_ω for all $\omega \in \Omega$.

Transition probabilities. Let $n_{\omega,\omega'}$ denote the number of times that volume imbalance moved from state ω to ω'. Then, from standard results, we have that

$$\hat{p}_{\omega'|\omega} = \frac{n_{\omega,\omega'}}{\sum_{\omega''} n_{\omega,\omega''}},$$

where transitions from the end of one trading day to the start of the next trading day are excluded from the count. Table 12 of Appendix B provides the estimates of the transition probability matrix for the assets at the three different decision intervals.[16]

Fill probabilities. We estimate the fill probabilities with counterfactual analysis. Following Arroyo et al. (2023), we submit "hypothetical" limit orders (with unit volume) at the end of the best bid and best ask queues at time t, and we track if the hypothetical order was filled between t and $t+1$ following price-time priority. These hypothetical orders account for all the change in behavior of market participants described in Section 2.

6.2. *Spoofing Conditions*

Table 9 uses the estimates from Tables 7 and 12 to check if conditions in Theorems 1 and 2 are satisfied. The entry NA indicates that the conditions in Theorems 4.1 and 4.2 are not applicable because condition (C3) does not hold.

Table 9: Testable conditions from Theorems 1 and 2.

Ticker	5 seconds			1 second			0.5 seconds		
	(C1), (C2)	(C3)	Side	(C1), (C2)	(C3)	Side	(C1), (C2)	(C3)	Side
AAPL	✓	✓	$q > 0$	✓	✓	$q > 0$	✓	✓	$q > 0$
AMZN	✓	✓	$q < 0$	✓	✓	$q > 0$	✓	✓	$q > 0$
CSCO	✓	✓	$q < 0$	✓	✓	$q < 0$	✓	✓	$q > 0$
INTC	✓	✓	$q > 0$	✓	✓	$q > 0$	✓	✓	$q > 0$
MSFT	✓	✓	$q < 0$	✓	✓	$q > 0$	✓	✗	NA
TSLA	✓	✓	$q < 0$	✓	✓	$q < 0$	✓	✓	$q < 0$

For all assets and decision intervals considered, conditions (C1) and (C2) are satisfied. Therefore, there are values of the inventory aversion parameter α where a manipulative strategy is optimal for $\mathbf{s} = (SH, q > 0)$, $\mathbf{s} = (BH, q < 0)$, and $\mathbf{s} = (N, q)$ for either $q > 0$ or $q < 0$. On the

[16]The rows of the estimates of the transition probability matrix do not always sum to unity because we round the estimates to the second decimal point.

other hand, condition (C3) allows one to determine if manipulation occurs in the neutral regime when inventory is long or short.

Figure 7 in Appendix B plots $\bar{\alpha}_1(\omega, q)$ as a function of inventory, so the area under the curves describes the intervals $I'(\mathbf{s})$, where $\alpha \in I'(\mathbf{s})$ is a necessary condition for the algorithm to learn to manipulate the order book. Therefore, for all assets and decision intervals considered, there exists a range of values of the inventory aversion parameter α below the curve where an algorithm will learn to manipulate the book.

7. Extensions

This section analyzes extensions to our model, including additional testable conditions to determine when algorithms will learn to manipulate the order book.

7.1. Transition Probability

In our model, when a large limit order is submitted, the volume imbalance regime moves to the buy-heavy (the sell-heavy) regime with probability one at the next time step, i.e., $p(BH \,|\, \omega, LLB) = 1$ and $p(SH \,|\, \omega, LLS) = 1$ for all $\omega \in \Omega$. Here, we relax the assumption so that a large buy (sell) limit order moves the volume imbalance regime to buy-heavy (sell-heavy) with probability $1 - \kappa$, and the volume imbalance moves to the "wrong" regime with probability $\kappa/2$, where $\kappa \in [0, 1)$. Formally, we have $p(BH \,|\, \omega, LLB) = 1 - \kappa$, $p(N \,|\, \omega, LLB) = \kappa/2$, and $p(SH \,|\, \omega, LLB) = \kappa/2$ when a large buy limit order is submitted; and $p(SH \,|\, \omega, LLS) = 1 - \kappa$, $p(N \,|\, \omega, LLS) = \kappa/2$, and $p(BH \,|\, \omega, LLS) = \kappa/2$ when a large sell limit order is submitted. We recover the original model when $\kappa = 0$.

The following theorem shows that our testable conditions from Theorem 1 continue to hold if the transition probabilities from large limit orders (i) tilt the volume imbalance regime into the appropriate heavy regime with a higher probability than the baseline transition dynamics in Figure 1b, and (ii) tilt the volume imbalance regime into the wrong regime with a lower probability than the baseline transition dynamics in Figure 1b.

Theorem 3 *Let $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$,* (C1)*, and* (C2) *hold. If the transition probabilities associated with large limit orders are such that*

$$\begin{aligned} p(BH \,|\, \omega, LLB) = 1 - \kappa > p_{BH|\omega}\,,\ p(N \,|\, \omega, LLB) = \frac{\kappa}{2} < p_{N|\omega}\,,\ p(SH \,|\, \omega, LLB) = \frac{\kappa}{2} < p_{SH|\omega}\,, \\ p(SH \,|\, \omega, LLS) = 1 - \kappa > p_{SH|\omega}\,,\ p(N \,|\, \omega, LLS) = \frac{\kappa}{2} < p_{N|\omega}\,,\ p(BH \,|\, \omega, LLS) = \frac{\kappa}{2} < p_{BH|\omega}\,, \end{aligned} \tag{C4}$$

hold for all $\omega \in \Omega$. Then $I_1(\mathbf{s}) \neq \emptyset$ and $I_2(\mathbf{s}) \neq \emptyset$ for all $\mathbf{s} \in \mathcal{S}$ such that (i) $\mathbf{s} = (SH, q > 0)$, (ii) $\mathbf{s} = (BH, q < 0)$, and (iii) $\mathbf{s} = (N, q)$ for either $q > 0$ or $q < 0$.

The conditions on the fill probabilities play the same role as before. However, condition (C4) on the transition probabilities ensures large limit orders meaningfully affect the transition probabilities so that one can exploit the benefits created by a manipulative order, i.e., to facilitate the market maker in reverting her inventory position, or to exploit a round-trip trade.

For most assets in Table 12 of Appendix B, we see that condition (C4) holds for values of κ ranging from 0.2 to 0.5 depending on the stock. Therefore, in certain instruments, if a manipulative order allows one to transition to the appropriate heavy regime half of the time, then a manipulative strategy can become dynamically optimal. On the other hand, assets such as CSCO and INTC only hold for small values of κ ranging from 0.04 to 0.2 depending on the decision interval. In these cases, a manipulative strategy may no longer be dynamically optimal if a manipulative order does not allow one to transition to the appropriate heavy regime almost all of the time.

7.2. Finite Trading Horizon

Our analysis above focused on the tractable case with an infinite trading horizon because most learning algorithms are designed for the infinite horizon setting. However, a finite horizon model best captures intraday trading because many market makers close inventories before the end of the trading day. To capture this, we assume the one-step utility at time T corresponds to DN. With a finite trading horizon, theoretical results guarantee only that there exists an optimal non-stationary pure Markov strategy (see for example Proposition 4.4.3 in Puterman, 1994), so the space of strategies to search over significantly increases. This is intuitive because the optimal action with a few minutes before the end of the trading horizon differs from the optimal action with a few hours before closing. Nonetheless, the problem can be readily solved through backwards induction.

Intuitively, for a sufficiently long horizon T, the trading behaviour at the start should resemble that from an infinite horizon problem. Although we do not have testable conditions for a finite trading horizon, we use dynamic programming to solve numerically for the optimal non-stationary pure Markov strategy. Figures 8 and 9 of Appendix B use the empirical estimates from Tables 7 and 12, and discount factor $\delta = 1$, to plot the optimal actions for each state and at each point in time $t = 0, 1, 2, ..., T = 30$. From the figure, we can string together the optimal action from one time step to another, and we show that quote-based manipulation can occur at every time point t, and that manipulation occurs for inventory levels closer to zero.

8. Multiple Market Makers

Our analysis thus far focused on quote-based manipulation by a single market maker. An extension is to study multiple market makers who delegate their decision making processes to

learning algorithms. How does the introduction of another market maker (who also uses a learning algorithm) affect a single algorithm's ability to manipulate the order book?

We ignore competition between limit orders from the different algorithms to simplify the analysis. Instead, we analyze the effect of multiple algorithms attempting to control the volume imbalance regime through manipulative orders. To formalize the new transition dynamics of the volume imbalance regime, we focus on two market makers where $\mathbf{a} = (a^1, a^2)$ is the action profile. The transition dynamics now depend on the action profile $p(\omega' \mid \omega, \mathbf{a})$ for all $\omega, \omega' \in \Omega$, which we summarize below:

- If both market makers submit a small or no order, then the volume imbalance evolves according to its baseline dynamics in Figure 1b.
- If one market maker submits a large limit order and the other market maker submits a small or no order, then the volume imbalance regime moves to buy-heavy or sell-heavy with probability one (depending on which side the large limit order is placed).
- If both market makers submit large limit orders on the same side of the book, then the volume imbalance regime moves to buy-heavy or sell-heavy with probability one (depending on which side the large limit orders are placed).
- If both market makers submit large limit orders on opposing sides of the book, then the volume imbalance evolves according its baseline dynamics in Figure 1b.

The notion of optimality with multiple market makers is based on equilibrium solution concepts. In turn, the solution concepts depend on the game and the strategies used by the learning algorithms. Generic learning algorithms search for an optimal strategy in the space of stationary Markov strategies; hence, an algorithm conditions its behavior on the set of states encoded in the algorithm. In our setting, each algorithm conditions their behavior on the volume imbalance regime which is publicly observable, but they also condition their behavior on their own level of inventory which is private information.

With private information, the most appropriate equilibrium solution concept in this setting is a perfect Bayesian equilibrium, where the belief determines the opponent's level of inventory, and the optimal strategy should be optimal with respect to the belief. However, generic learning algorithms use stationary Markov strategies and do not account for an opponent's level of inventory, so a perfect Bayesian equilibrium is not appropriate for generic algorithms.[17] Nevertheless, we can

[17] A Berk–Nash equilibrium (see Esponda and Pouzo, 2016) is the most suitable equilibrium solution concept for this misspecified setting, but the analysis is beyond the scope of the paper, and also not applicable to generic learning algorithms.

analyze the effect of this misspecification on an algorithm's ability to manipulate the order book.

Table 10: Average number of manipulative sequences over 50 trading intervals.

Ticker	Setup	Decision Interval Δt	Zero inventory		Same inventory		Opposing inventory	
			Agent 1 $q = 0$	Agent 2 $q = 0$	Agent 1 $q = 4$	Agent 2 $q = 4$	Agent 1 $q = 4$	Agent 2 $q = -4$
AMZN	Baseline	5 seconds	24.87	20.87	20.79	25.93	21.97	22.11
		1 second	25.22	14.92	14.78	29.25	18.52	18.77
		0.5 seconds	27.03	14.51	14.52	32.42	17.37	14.45
	Offline	5 seconds	24.92	26.29	21.01	22.65	20.85	22.52
		1 second	27.01	29.62	17.12	19.02	17.46	19.40
		0.5 seconds	30.71	32.76	16.20	18.32	22.05	18.27
	Online	5 seconds	24.40	25.89	20.47	22.12	20.41	22.04
		1 second	22.49	29.16	12.69	19.26	11.98	18.16
		0.5 seconds	21.27	32.13	1.21	15.20	1.12	14.43
CSCO	Baseline	5 seconds	25.56	15.25	15.12	29.19	18.29	18.82
		1 second	32.98	13.88	13.70	36.48	11.59	10.40
		0.5 seconds	37.27	9.54	9.55	40.27	7.73	7.89
	Offline	5 seconds	29.72	29.65	20.50	18.91	21.67	19.99
		1 second	37.13	37.10	14.73	12.62	21.25	18.53
		0.5 seconds	40.90	41.04	9.40	8.66	21.07	19.69
	Online	5 seconds	22.16	29.33	12.75	17.94	11.69	18.78
		1 second	20.13	36.46	0.0	13.00	0.0	10.79
		0.5 seconds	32.59	39.78	0.0	14.91	0.0	14.49

To study the effect of introducing a second learning algorithm, we first establish a baseline with one market maker who uses an algorithm to learn the optimal trading strategy. The market makers solve for the optimal strategy with the policy iteration algorithm using the empirical estimates from Tables 7 and 12, discount factor $\delta = 0.95$, and inventory aversion $\alpha = 10^{-4}$ for market maker one and $\alpha = 10^{-5}$ for market maker two. Although we have two market makers for the baseline, we do not study their interaction for the baseline setting. For each market maker, we simulate their optimal strategy over 50 time steps 10,000 times when the market maker starts with different values of the initial level of inventory. The initial volume imbalance regime is sampled with equal probability. Table 10 reports the average number of manipulative sequences for each market maker. We count a manipulative sequence as LLB at time t followed by LS or LLS at time $t + 1$ for $q \geq 0$, or LLS at time t followed by LB or LLB at time $t + 1$ for $q \leq 0$. The table

reports the results of two representative assets AMZN and CSCO. The results of the other assets are reported in Tables 13–15 of Appendix B.

8.1. Offline Learning

Here, both market makers train their algorithms offline with the misspecified model (the original model from Section 3) which ignores the strategic behavior of other algorithms, and we analyze the outcome of the interaction between the algorithms in the market. With the same setup as that of the baseline, we simulate the interaction of the market makers over 50 time steps 10,000 times when (i) both market makers start with zero inventory, (ii) both market makers start with the same level of inventory ($q = 4$), and (iii) the market makers start with opposing levels of inventory ($q = 4$ and $q = -4$).

When comparing the result with the baseline, the introduction of another market maker increases the number of times manipulation occurs in the market (i.e., market makers ride the manipulative sequences of each other) when the market makers start with zero inventory and with opposing levels of inventory. On the other hand, the introduction of another market maker decreases the number of times manipulation occurs in the market when the market makers start with the same level of inventory.

Table 11: Average manipulation statistics.

Ticker	Setup	Δt	Mismatching manipulative orders			Single manipulative order		
			Zero inv.	Same inv.	Opposing inv.	Zero inv.	Same inv.	Opposing inv.
AMZN	Offline	5s	0.1554%	0.2388%	0.4408%	13.46	18.42	18.75
		1s	0.1215%	1.3516%	0.0054%	22.47	22.01	29.52
		0.5s	0%	0%	0%	19.07	18.71	34.65
	Online	5s	0.2256%	0.3738%	0.4949%	19.39	21.58	21.92
		1s	0.4190%	1.8937%	2.3348%	24.25	26.66	23.93
		0.5s	0.7635%	0%	0%	25.25	25.96	25.69
CSCO	Offline	5s	0.3008%	0.0591%	1.0232%	20.13	19.82	34.29
		1s	0.0115%	1.7566%	4.6456%	12.63	7.66	38.89
		0.5s	0.0109%	0%	4.5198%	8.82	2.98	41.07
	Online	5s	0.6417%	2.2775%	4.6541%	24.77	26.35	25.06
		1s	1.4149%	0%	0%	25.85	24.88	20.30
		0.5s	2.2263%	0%	0%	16.17	30.77	29.38

To analyze the impact of introducing another market maker, Table 11 reports the percentage of times when competing market makers submit large orders that cancel each other out divided by

the number of times the market makers used a large order (mismatching manipulative orders), and the number of times where only one out of the two market makers submits a large order (single manipulative order). The results of the other assets are reported in Tables 16 and 17 of Appendix B.

When starting with zero inventory or opposing levels of inventory, the additional market maker does not lead to many instances where the large orders cancel each other out, but it does lead to more instances where the order book moves to a heavy regime. This allows the market makers to exploit the manipulative sequences of each other so that we have more manipulative sequences than would otherwise occur with only one market maker. When starting with the same level of inventory, the additional market maker does not lead to many instances where the large orders cancel each other out, but there are fewer instances where only one market maker submits a manipulative order. We see that market maker one disrupts market maker two because market maker one manipulates as often as the baseline setting but market maker two has significantly fewer manipulative sequences. In the offline learning setting, the algorithms either coordinate or mis-coordinate depending on their initial inventory.

8.2. Online Learning

Next, we assume that both market makers pre-train their algorithms offline with the misspecified model and use the results to initialize an online learning algorithm. With the same setup as that of the baseline, the market makers pre-train with the policy iteration algorithm and then use Q-learning to learn online (see Sutton and Barto, 2018; Calvano et al., 2020, for a basic explanation of Q-learning).[18]

For the online learning, we follow the experimental setup of Calvano et al. (2020). The Q-learning algorithms learn have an ε-greedy choice rule with a time-declining exploration rate given by $\varepsilon_t = \exp(-\tau\, t)$, where the parameter $\tau > 0$ controls the rate of decay of exploration. The ε-greedy choice rule picks the (current) optimal action with probability $1 - \varepsilon$, and a random action is chosen with probability ε. The learning rate of the algorithms is 0.125 and the exploration parameter is $\tau = 10^{-5}$. Similar to Calvano et al. (2020), we say that the online learning converged if the optimal strategy for each player does not change for 100,000 consecutive periods.

To analyze the effect of online learning, we simulate the learning process until convergence 1,000 times. Once each learning process converges, we use the learned strategies to simulate the

[18]Training an algorithm online by interacting with the market is costly because the algorithm needs to experiment frequently to learn to behave optimally. In most situations with multiple algorithms, learning algorithms are longer guaranteed to behave optimally. Realistically, market makers train their algorithms offline or partially train their algorithms offline (with some online experimentation) to minimize the cost of experimentation. We use the policy iteration algorithm to solve for the optimal continuation value v^* in the misspecified model, which we use to compute the optimal action values to initialize the Q-values for Q-learning.

interaction of the market makers over 50 time steps 10 times when (i) both market makers start with zero inventory, (ii) both market makers start with the same level of inventory ($q = 4$), and (iii) the market makers start with opposing levels of inventory ($q = 4$ and $q = -4$). This produces a total of 10,000 interactions over 50 time steps as in the case of the baseline.

When comparing the number of manipulative sequences to the baseline, we see that online learning often leads to a reduction in manipulation by market maker one, but an increase in manipulation by market maker two when the market makers start with zero inventory and with opposing levels of inventory. When the market makers start with the same level of inventory, there is a reduction in manipulation from both market makers. When comparing the manipulation statistics to offline learning, we see that online learning leads to more instances where only one market maker sends a manipulative order when the market makers start with zero inventory and with the same level of inventory. When the market makers start with opposing levels of inventory, there are fewer instances where only one market maker sends a manipulative order.

In the online learning setting, we see that the market makers learn to coordinate. How they coordinate depends on their initial inventory. If the market makers start with zero inventory, then they coordinate by riding the sequences of each other to increase market manipulation. On the other hand, if the market makers start with the same level of inventory or with opposing levels of inventory, then they coordinate by allowing market maker one to ride market maker two's sequences to avoid their large limit orders cancelling each other out.

9. Discussion

Our analysis focuses on when an algorithm learns to create misleading information to buy or to sell an asset with a higher probability than was otherwise likely to occur, and spoofing is a special case when the preference is for the manipulative order not to be filled. In both types of manipulation, the manipulative step is to submit a large quantity of limit orders to mislead other market participants who react to the misleading signal, so the manipulator benefits from this manipulation. Indeed, this manipulative step is consistent with what is considered illegal in Article 12(2)(c) of Regulation (EU) No 596/2014 and Section 9(a)(2) of the Securities Exchange Act of 1934.[19,20] Therefore, our results can help identify limit order books in both the European Union and US securities exchanges where quote-based manipulation is likely to occur.

[19] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02014R0596-20210101 and https://www.govinfo.gov/content/pkg/COMPS-1885/pdf/COMPS-1885.pdf, respectively.

[20] It is worth noting that Section 9(a)(2) of the Securities Exchange Act of 1934 refers to trade-based manipulation, whereas spoofing is a form of quote-based manipulation. However, existing case law roughly accomplishes the goal of making spoofing illegal, but it lacks clarity and makes errors in its reach.

In this paper, we define spoofing as a manipulative sequence in Definition 1 with the preference for the manipulative order not to be filled. Another definition of spoofing is given by the Dodd-Frank (2010) Act, which defines spoofing as bidding or offering with the intent to cancel the bid or offer before execution. Our definition encapsulates the spirit of the Dodd–Frank definition by including the preference not to get caught out with the manipulative order, while also having greater reach.[21] For example, manipulative orders with a time-in-force achieve the same effect as cancellations, but the Dodd–Frank definition will not cover this case because there are no cancellations. Another shortfall of the Dodd–Frank definition is its specific focus on spoofing, which is only a particular case of quote-based manipulation. This narrow focus fails to prohibit other forms of quote-based manipulation.

Our model predicts simple manipulative sequences that are easy to identify. However, in practice, identifying and detecting these sequences is not straightforward. For example, market fragmentation allows for cross market manipulation, so our sequences need not appear within the same order book. Moreover, the interaction of multiple market makers makes it more difficult to detect these sequences because the market makers can coordinate and ride the manipulative sequences of each other. Nonetheless, a straightforward mechanical artifact of quote-based manipulation is that the volatility of the microprice (volume weighted midprice) increases as quote-based manipulation increases; how one establishes a counterfactual baseline without quote-based manipulation remains unclear. Finally, although we focused on quote-based manipulation, our approach can be used to understand analytically other forms of unintended market manipulation from learning algorithms such as layering or quote stuffing.

References

ABADA, IBRAHIM AND XAVIER LAMBIN (2023): "Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided?" *Management Science*, 0.

AFM (2023): "Machine Learning in Trading Algorithms: Application by Dutch Proprietary Trading Firms and Possible Risks," Avaialble at:
www.afm.nl/en/sector/actueel/2023/maart/her-machine-learning.

ALLEN, FRANKLIN AND DOUGLAS GALE (1992): "Stock-Price Manipulation," *The Review of Financial Studies*, 5, 503–529.

ALLEN, FRANKLIN AND GARY GORTON (1992): "Stock Price Manipulation, Market Microstructure and Asymmetric Information," *European Economic Review*, 36, 624–630.

[21] Our requirement to have the preference not to get caught with the manipulative order is restrictive, but it is no more restrictive than proving intent, which has its own set of issues. For example, it is the main reason behind why tacit collusion is legal (see pp. 339–340 of Harrington, 2018). Nevertheless, showing a preference is not required to outlaw spoofing in securities exchanges.

AMIHUD, YAKOV AND HAIM MENDELSON (1986): "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 17, 223–249.

AQUILINA, MATTEO, ERIC BUDISH, AND PETER O'NEILL (2021): "Quantifying the High-Frequency Trading "Arms Race"," *The Quarterly Journal of Economics*, 137, 493–564.

ARROYO, ÁLVARO, ÁLVARO CARTEA, FERNANDO MORENO-PINO, AND STEFAN ZOHREN (2023): "Deep Attentive Survival Analysis in Limit Order Books: Estimating Fill Probabilities with Convolutional-Transformers," *Available at SSRN 4432087*.

BAGNOLI, MARK AND BARTON L. LIPMAN (1996): "Stock Price Manipulation Through Takeover Bids," *The RAND Journal of Economics*, 27, 124–147.

CALVANO, EMILIO, GIACOMO CALZOLARI, VINCENZO DENICOLÓ, AND SERGIO PASTORELLO (2020): "Artificial Intelligence, Algorithmic Pricing, and Collusion," *American Economic Review*, 110, 3267–97.

——— (2021): "Algorithmic Collusion with Imperfect Monitoring," *International Journal of Industrial Organization*, 79, 102712.

CAO, CHARLES, OLIVER HANSCH, AND XIAOXIN WANG (2009): "The Information Content of an Open Limit-Order Book," *Journal of Futures Markets*, 29, 16–41.

CARTEA, ÁLVARO, PATRICK CHANG, MATEUSZ MROCZKA, AND ROEL OOMEN (2022a): "AI-Driven Liquidity Provision in OTC Financial Markets," *Quantitative Finance*, 22, 2171–2204.

CARTEA, ÁLVARO, PATRICK CHANG, AND JOSÉ PENALVA (2022b): "Algorithmic Collusion in Electronic Markets: The Impact of Tick Size," *Available at SSRN 4105954*.

CARTEA, ÁLVARO, PATRICK CHANG, JOSÉ PENALVA, AND HARRISON WALDON (2023): "Algorithms can Learn to Collude: A Folk Theorem from Learning with Bounded Rationality," *Available at SSRN 4293831*.

CARTEA, ÁLVARO, RYAN DONNELLY, AND SEBASTIAN JAIMUNGAL (2017): "Algorithmic Trading with Model Uncertainty," *SIAM Journal on Financial Mathematics*, 8, 635–671.

CARTEA, ÁLVARO, SEBASTIAN JAIMUNGAL, AND YIXUAN WANG (2020): "Spoofing and Price Manipulation in Order-Driven Markets," *Applied Mathematical Finance*, 27, 67–98.

CHAKRABORTY, ARCHISHMAN AND BILGE YILMAZ (2004a): "Informed Manipulation," *Journal of Economic Theory*, 114, 132–152.

——— (2004b): "Manipulation in Market Order Models," *Journal of Financial Markets*, 7, 187–206.

COLLIARD, JEAN-EDOUARD, THIERRY FOUCAULT, AND STEFANO LOVO (2022): "Algorithmic Pricing and Liquidity in Securities Markets," *HEC Paris Research Paper*.

COPELAND, THOMAS E. AND DAN GALAI (1983): "Information Effects on the Bid-Ask Spread," *The Journal of Finance*, 38, 1457–1469.

DODD-FRANK (2010): "Dodd–Frank Wall Street Reform and Consumer Protection Act," Public Law 111–203. Availalble at: https://www.govinfo.gov/app/details/PLAW-111publ203.

DOU, WINSTON WEI, ITAY GOLDSTEIN, AND YAN JI (2023): "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency," *Available at SSRN 4452704*.

ESPONDA, IGNACIO AND DEMIAN POUZO (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 84, 1093–1130.

FOX, MERRITT B., LAWRENCE R. GLOSTEN, AND SUE S. GUAN (2021): "Spoofing and its Regulation," *Colum. Bus. L. Rev.*, 1244.

GLOSTEN, LAWRENCE R. (1994): "Is the Electronic Open Limit Order Book Inevitable?" *The Journal of Finance*,

49, 1127–1161.

GLOSTEN, LAWRENCE R. AND PAUL R. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

HAMBLY, BEN, RENYUAN XU, AND HUINING YANG (2023): "Recent Advances in Reinforcement Learning in Finance," *Mathematical Finance*, 33, 437–503.

HANSEN, LARS PETER AND THOMAS J. SARGENT (2007): *Robustness*, Princeton University Press.

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

HARRIS, LAWRENCE E. AND VENKATESH PANCHAPAGESAN (2005): "The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions," *Journal of Financial Markets*, 8, 25–67.

HASBROUCK, JOEL AND GEORGE SOFIANOS (1993): "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists," *The Journal of Finance*, 48, 1565–1593.

HO, THOMAS AND HANS R. STOLL (1981): "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9, 47–73.

LEE, EUN JUNG, KYONG SHIK EOM, AND KYUNG SUH PARK (2013): "Microstructure-Based Manipulation: Strategic Behavior and Performance of Spoofing Traders," *Journal of Financial Markets*, 16, 227–252.

MADHAVAN, ANANTH AND SEYMOUR SMIDT (1993): "An Analysis of Changes in Specialist Inventories and Quotations," *The Journal of Finance*, 48, 1595–1628.

NING, BRIAN, FRANCO HO TING LIN, AND SEBASTIAN JAIMUNGAL (2021): "Double Deep Q-Learning for Optimal Execution," *Applied Mathematical Finance*, 28, 361–380.

OECD (2021): "Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers," Available at: https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm.

O'HARA, MAUREEN AND GEORGE S. OLDFIELD (1986): "The Microeconomics of Market Making," *The Journal of Financial and Quantitative Analysis*, 21, 361–376.

PARLOUR, CHRISTINE A. (1998): "Price Dynamics in Limit Order Markets," *The Review of Financial Studies*, 11, 789–816.

PUTERMAN, MARTIN L. (1994): *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, USA: John Wiley & Sons, Inc., 1st ed.

SINGH, SATINDER, TOMMI JAAKKOLA, MICHAEL L. LITTMAN, AND CSABA SZEPESVÁRI (2000): "Convergence Results for Single-Step On-Policy Reinforcement-Learning Algorithms," *Machine Learning*, 38, 287–308.

SPOONER, THOMAS, JOHN FEARNLEY, RAHUL SAVANI, AND ANDREAS KOUKORINIS (2018): "Market Making via Reinforcement Learning," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, AAMAS '18, 434–442.

STOLL, HANS R. (1978): "The Supply of Dealer Services in Securities Markets," *The Journal of Finance*, 33, 1133–1151.

SUTTON, RICHARD S. AND ANDREW G. BARTO (2018): *Reinforcement Learning: An Introduction*, The MIT Press, second ed.

SZEPESVÁRI, CSABA (2010): *Algorithms for Reinforcement Learning*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Springer Cham.

VAN BOMMEL, JOS (2003): "Rumors," *The Journal of Finance*, 58, 1499–1519.

VILA, JEAN-LUC (1989): "Simple Games of Market Manipulation," *Economics Letters*, 29, 21–26.

WANG, YUN-YI (2019): "Strategic Spoofing Order Trading by Different Types of Investors in Taiwan Index Futures Market," *Journal of Financial Studies*, 27, 65–103.

WATKINS, CHRISTOPHER J. C. H. AND PETER DAYAN (1992): "Q-Learning," *Machine Learning*, 8, 279–292.

WILLIAMS, BASIL AND ANDRZEJ SKRZYPACZ (2021): "Spoofing in Equilibrium," *Available at SSRN 3742327*.

Appendix

A. Proofs

Define the action value as

$$\begin{aligned} \upsilon_{\mathbf{s}}(a) &= \mathbb{E}_{\sigma^*}\left[\sum_{t=0}^{\infty}\delta^t\, u(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) \,\middle|\, \mathbf{s}_0 = \mathbf{s}, a_0 = a\right] \\ &= \bar{u}(\mathbf{s}, a) + \delta \sum_{\omega' \in \Omega} p(\omega'|\omega, a) \sum_{q' \in \mathcal{Q}} p(q'|q, a)\, \upsilon^*_{\omega', q'}\,, \end{aligned} \tag{8}$$

which is the expected stream of discounted payoffs from playing action a in state $\mathbf{s}$ then playing optimally thereafter according to an optimal strategy σ^*.

Proof of Lemma 1 Order the set $\mathcal{Q}$ so that $\mathcal{Q} = \{-\bar{q}, ..., -2, -1, 0, 1, 2, ..., \bar{q}\}$. To keep notation simple, let $\mathcal{A}_{\mathbf{s}} = \mathcal{A}'$ denote the set of all actions for all $\mathbf{s} \in \mathcal{S}$. The result continues to hold without this simplification (see Puterman, 1994, pp. 108).[22] We have the following observations:

(i) For $q \geq 0$, $\bar{u}(\omega, q, a)$ is non-increasing in q; i.e., $\bar{u}(\omega, q, a) \leq \bar{u}(\omega, q', a)$ for all $\omega \in \Omega$ and for all $a \in \mathcal{A}'$ when $0 \leq q' \leq q$;

(ii) For $q \leq 0$, $\bar{u}(\omega, q, a)$ is non-decreasing in q; i.e., $\bar{u}(\omega, q', a) \leq \bar{u}(\omega, q, a)$ for all $\omega \in \Omega$ and for all $a \in \mathcal{A}'$ when $q' \leq q \leq 0$; and

(iii) $\sum_{j=k}^{\infty} p(j|q, a)$ is non-decreasing in q for all $k \in \mathcal{Q}$ and for all $a \in \mathcal{A}'$, where $p(j|q, a)$ is the transition probability from inventory level q to inventory level j under action a.[23]

Consider the finite-horizon version of our model up to period N. Let $\upsilon^*_{\mathbf{s}}(t)$ denote the optimal continuation value in state $\mathbf{s}$ at period t and adopt the convention that the terminal payoff $\bar{u}_N(\omega, q) = \bar{u}(\omega, q, DN)$.

[22] Indeed, the three conditions laid out on pp. 108 are satisfied with an appropriate adjustment to the ordering of the set $\mathcal{Q}$ and the proof follows with some minor adjustments.

[23] We adopt the convention that $p(j|q, a) = 0$ when $j \notin \mathcal{Q}$.

Claim 1 *For $t = 0, 1, 2, ..., N$, we have that*

$$v^*_{\omega,q}(t) = \max_{a \in \mathcal{A}'} \left\{ \bar{u}(\omega, q, a) + \delta \sum_{\omega'} p(\omega'|\omega, a) \sum_{j=0}^{\infty} p(j|q, a)\, v^*_{\omega',j}(t+1) \right\}$$

is non-increasing in q when $q \geq 0$, so $v^*_{\omega,q}(t) \leq v^*_{\omega,q'}(t)$ for all $\omega \in \Omega$ and for all $t = 0, 1, 2, ..., N$ when $0 \leq q' \leq q$. Similarly, when $q \leq 0$, $v^*_{\omega,q}(t)$ is non-decreasing in q, i.e., $v^*_{\omega,q'}(t) \leq v^*_{\omega,q}(t)$ for all $\omega \in \Omega$ and for all $t = 0, 1, 2, ..., N$ when $q' \leq q \leq 0$.

Proof of Claim 1 The claim follows from a straightforward modification of Proposition 4.7.3 in Puterman (1994) using backwards induction. Consider the case $q \geq 0$. First, the result holds for $t = N$ from observation (i) because $v^*_{\omega,q}(N) = \bar{u}_N(\omega, q)$. Next, assume (for induction) that when $0 \leq q' \leq q$, we have $v^*_{\omega,q}(t) \leq v^*_{\omega,q'}(t)$ for all $\omega \in \Omega$ and for all $t = n+1, ..., N$. By Proposition 4.4.3 in Puterman (1994), there exists $a^* \in \mathcal{A}'$ so that

$$v^*_{\omega,q}(t) = \bar{u}(\omega, q, a^*) + \delta \sum_{\omega'} p(\omega'|\omega, a^*) \sum_{j=0}^{\infty} p(j|q, a^*)\, v^*_{\omega',j}(t+1)\,.$$

Let $0 \leq q' \leq q$. Use observations (i) and (iii), the induction hypothesis, and Lemma 4.7.2 in Puterman (1994) to write

$$\begin{aligned} v^*_{\omega,q}(t) &\leq \bar{u}(\omega, q', a^*) + \delta \sum_{\omega'} p(\omega'|\omega, a^*) \sum_{j=0}^{\infty} p(j|q', a^*)\, v^*_{\omega',j}(t+1) \\ &\leq \max_{a \in \mathcal{A}'} \left\{ \bar{u}(\omega, q', a) + \delta \sum_{\omega'} p(\omega'|\omega, a) \sum_{j=0}^{\infty} p(j|q', a)\, v^*_{\omega',j}(t+1) \right\} \\ &= v^*_{\omega,q'}(t)\,. \end{aligned}$$

When $q \leq 0$, similar calculations hold with observations (ii) and (iii). Thus, the claim follows. ■

Finally, the pointwise limit (as $N \to \infty$) of non-increasing functions is non-increasing, and the pointwise limit (as $N \to \infty$) of non-decreasing functions is non-decreasing. Hence, $v^*_{\omega,q}(t)$ is non-increasing in q for all t when $q \geq 0$, and $v^*_{\omega,q}(t)$ is non-decreasing in q for all t when $q \leq 0$, so the lemma follows. □

Proof of Lemma 2 The optimal action a^* in state $\mathbf{s} = (\omega, q)$ maximizes the action value in (8). To show that an action a is not optimal in state $\mathbf{s}$, we show that the action is strictly dominated by another action a', i.e., $v_{\mathbf{s}}(a) < v_{\mathbf{s}}(a')$ for state $\mathbf{s}$.

For $q > 0$, we first show that submitting a sell limit order LS always dominates do nothing DN. Specifically, we have

$$\begin{aligned}
\upsilon_{\mathbf{s}}(LS) &= p^a_\omega \left[\frac{\vartheta}{2} - \alpha\,(q-1)^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q-1}\right] + (1 - p^a_\omega)\left[-\alpha\, q^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q}\right] \\
&> -\alpha\, q^2 + p^a_\omega\, \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q-1} + (1 - p^a_\omega)\, \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q} \\
&\geq -\alpha\, q^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q} = \upsilon_{\mathbf{s}}(DN)\,,
\end{aligned}$$

where the last inequality follows from Lemma 1. Next, do nothing DN always dominates a buy market order MB because

$$\begin{aligned}
\upsilon_{\mathbf{s}}(DN) &= -\alpha\, q^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q} \\
&> -\frac{\vartheta}{2} - \alpha\,(q+1)^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\, \upsilon^*_{\omega,q+1} = \upsilon_{\mathbf{s}}(MB)\,,
\end{aligned}$$

where the inequality follows from Lemma 1. Thus, LS dominates DN, which dominates MB. Therefore, both DN and MB are not optimal for $q > 0$.

The same reasoning and calculations hold for $q < 0$. Thus, the result follows because a buy limit order always dominates do nothing, and do nothing always dominates a sell market order. □

Proof of Proposition 1 For $q > 0$, Lemma 2 ensures that the optimal action is never DN or MB. We consider the ten pairwise comparisons from the set of actions $\{MS, LS, LLS, LLB, LB\}$. For each (ω, q), there exists a unique value $\alpha_{a,a'}(\omega, q)$ where two action values $\upsilon_{\omega,q}(a)$ and $\upsilon_{\omega,q}(a')$ as a function of α intersect because the action values are linear in α.

To compare the pairwise actions excluding LS versus LLS and LB versus LLB, we choose a large value of α so that it is optimal to revert to zero inventory as fast as possible and then stop making markets at zero inventory. That is, MS is the optimal action for all pairs $(\omega, q > 0)$ and DN is the optimal action for all pairs $(\omega, q = 0)$, so it is easy to see that for $q > 0$ and all $\omega \in \Omega$:

1. There exists $\alpha_{LS,MS}(\omega, q)$ such that MS is preferred to LS if and only if $\alpha > \alpha_{LS,MS}(\omega, q)$.
2. There exists $\alpha_{LLS,MS}(\omega, q)$ such that MS is preferred to LLS if and only if $\alpha > \alpha_{LLS,MS}(\omega, q)$.
3. There exists $\alpha_{LLB,MS}(\omega, q)$ such that MS is preferred to LLB if and only if $\alpha > \alpha_{LLB,MS}(\omega, q)$.
4. There exists $\alpha_{LB,MS}(\omega, q)$ such that MS is preferred to LB if and only if $\alpha > \alpha_{LB,MS}(\omega, q)$.

5. There exists $\alpha_{LLB,LS}(\omega, q)$ such that LS is preferred to LLB if and only if $\alpha > \alpha_{LLB,LS}(\omega, q)$.

6. There exists $\alpha_{LB,LS}(\omega, q)$ such that LS is preferred to LB if and only if $\alpha > \alpha_{LB,LS}(\omega, q)$.

7. There exists $\alpha_{LLB,LLS}(\omega, q)$ such that LLS is preferred to LLB if and only if $\alpha > \alpha_{LLB,LLS}(\omega, q)$.

8. There exists $\alpha_{LB,LLS}(\omega, q)$ such that LLS is preferred to LB if and only if $\alpha > \alpha_{LB,LLS}(\omega, q)$.

For the final two comparisons, we consider a reduced action set without MS. We choose a large value of α so that it is optimal to revert to zero inventory as fast as possible and then stop making markets at zero inventory, so for $q > 0$ and all $\omega \in \Omega$:

9. If we play LLS instead of LS, the tilt of the book becomes sell-heavy where the probability of selling a limit order is the lowest, which will take longer to revert to zero inventory. Therefore, there exists $\alpha_{LLS,LS}(\omega, q)$ such that LS is preferred to LLS if and only if $\alpha > \alpha_{LLS,LS}(\omega, q)$.

10. If we play LLB instead of LB, the tilt of the book becomes buy-heavy where the probability of selling a limit order is the highest, and it will take less time to revert to zero inventory. Therefore, there exists $\alpha_{LB,LLB}(\omega, q)$ such that LLB is preferred to LB if and only if $\alpha > \alpha_{LB,LLB}(\omega, q)$.

With the ten pairwise preferences for $q > 0$, we have that Figure 2 is the only (non-contradictory) ordering of action preferences and cutoffs. The same reasoning holds for $q < 0$. □

Lemma 4 *Let $0 < q' \leq q$. We have that $v^*_{\omega,q} - v^*_{\omega,q'}$ converges to zero from below as $\alpha \to 0$ for all $\omega \in \Omega$. Similarly, $v^*_{\omega,q'} - v^*_{\omega,q}$ converges to zero from above as $\alpha \to 0$.*

Proof of Lemma 4 The result follows from adapting the proof of Lemma 1 to show that if $\alpha = 0$ then $v^*_{\omega,q} - v^*_{\omega,q'} = 0$. First, observe that if $\alpha = 0$, then $\bar{u}(\omega, q, a)$ is constant in q when $q \geq 0$ for all $\omega \in \Omega$ and for all $a \in \mathcal{A}'$. Following the proof of Lemma 1, we have that the optimal continuation value $v^*_{\omega,q}$ is both non-increasing and non-decreasing in q when $q \geq 0$. Therefore, $v^*_{\omega,q} - v^*_{\omega,q'} = 0$ for all $q, q' \geq 0$, and hence, the result follows as a consequence of Lemma 1. □

Lemma 5 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. We have $\alpha_3(\omega, q) > 0$ for all $\omega \in \Omega$ and $q \neq 0$.*

Proof of Lemma 5 We focus on $q > 0$. We prove this result by showing that the following claim holds.

Claim 2 *For $q > 0$, there exists $\alpha > 0$ such that $MS \prec LS$.*

Proof of Claim 2 The action values of LS and MS are given by

$$\upsilon_{\mathbf{s}}(LS) = p^a_\omega \left[\frac{\vartheta}{2} - \alpha\,(q-1)^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\,\upsilon^*_{\omega,q-1}\right] + (1-p^a_\omega)\left[-\alpha\,q^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\,\upsilon^*_{\omega,q}\right],$$
$$\upsilon_{\mathbf{s}}(MS) = -\frac{\vartheta}{2} - \alpha\,(q-1)^2 + \delta \sum_{\omega'} p_{\omega'|\omega}\,\upsilon^*_{\omega,q-1},$$

and their difference is

$$\upsilon_{\mathbf{s}}(LS) - \upsilon_{\mathbf{s}}(MS) = (1+p^a_\omega)\,\vartheta/2 - \alpha\,(1-p^a_\omega)\,(2q-1) + \delta\,(1-p^a_\omega) \sum_{\omega'} p_{\omega'|\omega}\left(\upsilon^*_{\omega,q} - \upsilon^*_{\omega,q-1}\right).$$

From Lemma 4, we have that $\upsilon_{\mathbf{s}}(LS) - \upsilon_{\mathbf{s}}(MS) \to (1+p^a_\omega)\,\vartheta/2 > 0$ as $\alpha \to 0$. Thus, the claim follows because there are values of $\alpha > 0$ for which $MS \prec LS$. ■

The remainder of the lemma follows by contradiction. Suppose $\alpha_3(\omega, q) \leq 0$. Then MS is optimal for all $\alpha > 0$, which contradicts Claim 2, so the result follows. The same reasoning holds for $q < 0$. □

Lemma 6 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. If condition (C1) holds, then $\alpha_1(BH, q) < 0$ for all $q > 0$. Similarly, if condition (C2) holds, then $\alpha_1(SH, q) < 0$ for all $q < 0$.*

Proof of Lemma 6 We focus on $q > 0$. The action value of LLB is

$$\upsilon_{\mathbf{s}}(LLB) = p^b_\omega\left[\vartheta/2 - \alpha\,(q+1)^2 + \delta\,\upsilon^*_{BH,q+1}\right] + (1-p^b_\omega)\left[-\alpha\,q^2 + \delta\,\upsilon^*_{BH,q}\right].$$

To prove the result, we show that if condition (C1) holds, then LLB is never optimal in $\mathbf{s} = (BH, q)$ where $q > 0$. We proceed by contradiction. Suppose LLB is optimal in $\mathbf{s} = (BH, q)$ where $q > 0$. This implies the following claim is true.

Claim 3 *If LLB is optimal in $\mathbf{s} = (BH, q)$ where $q > 0$, then $\vartheta/2 - \alpha\,(q+1)^2 + \delta\,\upsilon^*_{BH,q+1} \geq -\alpha\,q^2 + \delta\,\upsilon^*_{BH,q}$.*

Proof of Claim 3 We prove this claim by contradiction. Suppose that $-\alpha\,q^2 + \delta\,\upsilon^*_{BH,q} > \vartheta/2 - \alpha\,(q+1)^2 + \delta\,\upsilon^*_{BH,q+1}$, then we have

$$\upsilon_{BH,q}(LLB) = p^b_{BH}\left[\vartheta/2 - \alpha\,(q+1)^2 + \delta\,\upsilon^*_{BH,q+1}\right] + (1-p^b_{BH})\left[-\alpha\,q^2 + \delta\,\upsilon^*_{BH,q}\right]$$

$$\leq p^b_{BH}\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]+(1-p^b_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]=-\alpha\,q^2+\delta\,v^*_{BH,q}\,,$$

which follows from Lemma 1. However, $v^*_{BH,q} = v_{BH,q}(LLB)$ because LLB is optimal by assumption. Therefore, the inequality above becomes $v^*_{BH,q} \leq -\alpha\,q^2+\delta\,v^*_{BH,q}$, which implies

$$v^*_{BH,q} \leq -\frac{\alpha\,q^2}{1-\delta}\,.$$

Now, consider a suboptimal strategy σ that does nothing in all $\omega \in \Omega$ at inventory level q. The value of this strategy in $\mathbf{s} = (BH, q)$ where $q > 0$ is given by

$$v_{BH,q}(\sigma) = -\frac{\alpha\,q^2}{1-\delta}\,.$$

Therefore, $v_{BH,q}(\sigma) \geq v^*_{BH,q}$ is a contradiction because the strategy σ is suboptimal as a consequence of Lemma 2. Hence, the claim follows. ■

Next, the claim implies that $v_{SH,q}(LLB) \geq v_{BH,q}(LLB)$ because

$$\begin{aligned}
v_{SH,q}(LLB) &= p^b_{SH}\,\left[\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1}\right]+(1-p^b_{SH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&\geq p^b_{BH}\,\left[\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1}\right]+(1-p^b_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&= v_{BH,q}(LLB)\,,
\end{aligned}$$

where the inequality follows as a result of $p^b_{SH} > p^b_{BH}$.

Use $1-p^b_{BH} = (1-p^a_{BH})+(p^a_{BH}-p^b_{BH})$ in the action value of LLB in $\mathbf{s} = (BH, q)$ to obtain

$$\begin{aligned}
v_{BH,q}(LLB) &= p^b_{BH}\,\left[\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1}\right]+(1-p^b_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&= p^b_{BH}\,\left[\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1}\right]+(p^a_{BH}-p^b_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&\quad+(1-p^a_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&\leq p^b_{BH}\max\left\{\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1},-\alpha\,q^2+\delta\,v^*_{BH,q}\right\}\\
&\quad+(p^a_{BH}-p^b_{BH})\max\left\{\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1},-\alpha\,q^2+\delta\,v^*_{BH,q}\right\}\\
&\quad+(1-p^a_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\\
&= p^a_{BH}\max\left\{\vartheta/2-\alpha\,(q+1)^2+\delta\,v^*_{BH,q+1},-\alpha\,q^2+\delta\,v^*_{BH,q}\right\}\\
&\quad+(1-p^a_{BH})\,\left[-\alpha\,q^2+\delta\,v^*_{BH,q}\right]\,,
\end{aligned}$$

where the inequality above follows as a result of $p^b_{BH} < p^a_{BH}$ from condition (C1) and because $x \leq \max\{x, y\}$ and $y \leq \max\{x, y\}$ for all $x, y \in \mathbb{R}$.

Next, observe that both $\vartheta/2 - \alpha\,(q+1)^2 + \delta\,v^*_{BH,q+1}$ and $-\alpha\,q^2 + \delta\,v^*_{BH,q}$ are less than $\vartheta/2 - \alpha\,(q-1)^2 + \delta\,v^*_{SH,q-1}$ because $v^*_{SH,q-1} \geq v^*_{SH,q} \geq v_{SH,q}(LLB) \geq v_{BH,q}(LLB) = v^*_{BH,q} \geq v^*_{BH,q+1}$. Therefore,

$$\begin{aligned}
v_{BH,q}(LLB) &\leq p^a_{BH} \max\left\{\vartheta/2 - \alpha\,(q+1)^2 + \delta\,v^*_{BH,q+1}, -\alpha\,q^2 + \delta\,v^*_{BH,q}\right\} \\
&\quad + (1 - p^a_{BH})\left[-\alpha\,q^2 + \delta\,v^*_{BH,q}\right] \\
&\leq p^a_{BH}\left[\vartheta/2 - \alpha\,(q-1)^2 + \delta\,v^*_{SH,q-1}\right] + (1 - p^a_{BH})\left[-\alpha\,q^2 + \delta\,v^*_{BH,q}\right] \\
&\leq p^a_{BH}\left[\vartheta/2 - \alpha\,(q-1)^2 + \delta\,v^*_{SH,q-1}\right] + (1 - p^a_{BH})\left[-\alpha\,q^2 + \delta\,v^*_{SH,q}\right] \\
&= v_{BH,q}(LLS)\,,
\end{aligned}$$

where the last inequality follows from $v^*_{SH,q} \geq v_{SH,q}(LLB) \geq v_{BH,q}(LLB) = v^*_{BH,q}$. This implies that $v_{BH,q}(LLB) \leq v_{BH,q}(LLS)$. Hence, we have a contradiction as LLB cannot be optimal in $\mathbf{s} = (BH, q)$ where $q > 0$.

Finally, if LLB is never optimal in $\mathbf{s} = (BH, q)$, then $\alpha_1(BH, q) < 0$. Thus, the lemma follows. The same reasoning holds for $q < 0$. □

Lemma 7 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. If condition (C2) holds, then $\alpha_1(\omega, q) > 0$ for all $q > 0$ and $\omega = SH$. Similarly, if condition (C1) holds, then $\alpha_1(\omega, q) > 0$ for all $q < 0$ and $\omega = BH$.*

Proof of Lemma 7 We focus on $q > 0$. We require the following claim.

Claim 4 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. If condition (C2) holds, then there exists $\alpha > 0$ such that $LS \prec LB$ for all $q > 0$ and $\omega = SH$. Therefore, for this $\alpha > 0$, LS is not optimal.*

Proof of Claim 4 The action values of LS and LB are given by

$$v_{\mathbf{s}}(LS) = p^a_\omega\,\vartheta/2 + \delta \sum_{\omega'} p_{\omega'|\omega}\,v^*_{\omega,q} + \delta\,p^a_\omega \sum_{\omega'} p_{\omega'|\omega}\left(v^*_{\omega,q-1} - v^*_{\omega,q}\right) - \alpha\,p^a_\omega(q-1)^2 - \alpha(1-p^a_\omega)q^2$$

$$v_{\mathbf{s}}(LB) = p^b_\omega\,\vartheta/2 + \delta \sum_{\omega'} p_{\omega'|\omega}\,v^*_{\omega,q} + \delta\,p^b_\omega \sum_{\omega'} p_{\omega'|\omega}\left(v^*_{\omega,q+1} - v^*_{\omega,q}\right) - \alpha\,p^b_\omega(q+1)^2 - \alpha(1-p^b_\omega)q^2\,.$$

As $\alpha \to 0$, $v_{\mathbf{s}}(LB) - v_{\mathbf{s}}(LS) \to \left(p^b_\omega - p^a_\omega\right)\vartheta/2$ as a consequence of Lemma 4. Therefore, $LS \prec LB$ in $\omega = SH$ because of $p^a_{SH} < p^b_{SH}$ from condition (C2). Hence, LS cannot be optimal because $LS \prec LB$. ■

From Claim 4, there exists $\alpha > 0$ such that $LS \prec LB$ so that LS is not optimal. We set the value of α so that MS is not optimal. For this value of α, either LB, LLB or LLS is the optimal action. If LB or LLB is the optimal action, then we have $\alpha_1(\omega, q) > 0$. To complete the proof we see that LLS cannot be the optimal action for the value of α. We proceed by contradiction. Assume LLS is optimal and take α close to zero so that the effect of inventory is negligible when considering the optimal continuation value. We denote $\upsilon^*_\omega = \upsilon^*_{\omega,q} = \upsilon^*_{\omega,q'}$ to obtain

$$\begin{aligned}\upsilon_{SH}(LLB) &= p^b_{SH}(\vartheta/2 + \delta\,\upsilon^*_{BH}) + (1 - p^b_{SH})\,\delta\,\upsilon^*_{BH} = p^b_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{BH} \\ &\geq p^b_{SH}\,\vartheta/2 + \delta\,\upsilon_{BH}(LLS) = p^b_{SH}\,\vartheta/2 + \delta\,(p^a_{BH}\,\vartheta/2 + \delta\,\upsilon^*_{SH}) \\ &> p^b_{SH}\,\vartheta/2 + \delta\,(p^a_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{SH})\,,\end{aligned}$$

where the last inequality follows from $p^a_{BH} > p^a_{SH}$.

Observe that

$$p^b_{SH}\,\vartheta/2 + \delta\,(p^a_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{SH}) = p^b_{SH}\,\vartheta/2 + \delta\,\upsilon_{SH}(LLS) = p^b_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{SH}\,,$$

where the last equality follows from assuming that LLS is optimal at SH. Therefore, we have

$$\upsilon_{SH}(LLB) > p^b_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{SH} > p^a_{SH}\,\vartheta/2 + \delta\,\upsilon^*_{SH} = \upsilon^*_{SH}\,,$$

which is a contradiction. Therefore, LLS is not optimal. The same reasoning holds for $q < 0$. □

Lemma 8 *Let $p^a_{SH} < p^a_N < p^a_{BH}$ and $p^b_{SH} > p^b_N > p^b_{BH}$ hold. For $\omega = N$, we have $\alpha_1(\omega, q) > 0$ for either $q > 0$ or $q < 0$.*

Proof of Lemma 8 From Lemma 4, we have

$$\upsilon_N(LLS) = p^a_N\,\vartheta/2 + \delta\,\upsilon^*_{SH} \quad \text{and} \quad \upsilon_N(LLB) = p^b_N\,\vartheta/2 + \delta\,\upsilon^*_{BH}\,,$$

for a small enough value of α so that if

$$p^a_N\,\vartheta/2 + \delta\,\upsilon^*_{SH} > p^b_N\,\vartheta/2 + \delta\,\upsilon^*_{BH}\,,$$

then $\alpha_1(\omega, q) > 0$ for $q < 0$. Similarly, if

$$p^a_N\,\vartheta/2 + \delta\,\upsilon^*_{SH} < p^b_N\,\vartheta/2 + \delta\,\upsilon^*_{BH}\,,$$

then $\alpha_1(\omega, q) > 0$ for $q > 0$. □

Proof of Theorem 1 Lemma 5 ensures that $\alpha_3(\omega, q) > 0$ for all $\omega \in \Omega$ and $q > 0$, while Proposition 1 ensures that $\alpha_0(\omega, q) < \alpha_1(\omega, q)$ for all $\omega \in \Omega$ and $q > 0$. The result is immediate from Lemmas 6, 7, and 8. □

Proof of Corollary 1 The result follows as an immediate consequence of Theorem 1. □

Lemma 9 *Let* $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$, *and* (C3) *hold. If* $(p^b_N - p^a_N) > \frac{\delta}{1+\delta}(p^b_{SH} - p^a_{BH})$, *then* $\alpha_1(\omega, q) > 0$ *for* $q > 0$ *and* $\omega = N$. *Similarly, if* $(p^a_N - p^b_N) > \frac{\delta}{1+\delta}(p^a_{BH} - p^b_{SH})$, *then* $\alpha_1(\omega, q) > 0$ *for* $q < 0$ *and* $\omega = N$.

Proof of Lemma 9 To prove the lemma, we first establish the following claim.

Claim 5 *If* $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$, *and* (C3) *hold, then for a small enough value of* α, $LS \prec LLS$ *for* $\omega = BH$, $LB \prec LLB$ *for* $\omega = SH$, *and* $LB \prec LLB$ *and* $LS \prec LLS$ *for* $\omega = N$.

Proof of Claim 5 If (C3) holds, then the following hold:

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{p_{N|BH}}{p_{BH|BH}}, \quad p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{p_{N|SH}}{p_{SH|SH}},$$

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{p_{N|N}}{p_{BH|N}}, \quad p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{p_{N|N}}{p_{SH|N}}.$$

We first focus on $\omega = BH$. For a small enough value of α, the optimal action is LS or LLS. Moreover, LLS is preferred to LS if and only if

$$v^*_{SH} > p_{BH|BH}\, v^*_{BH} + p_{N|BH}\, v^*_N + p_{SH|BH}\, v^*_{SH} \iff$$

$$p_{BH|BH}\, v^*_{SH} + p_{N|BH}\, v^*_{SH} + p_{SH|BH}\, v^*_{SH} > p_{BH|BH}\, v^*_{BH} + p_{N|BH}\, v^*_N + p_{SH|BH}\, v^*_{SH}.$$

If $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{SH}$, then the last inequality trivially holds.[24] On the other hand, if $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then the last inequality holds if and only if

$$p_{N|BH}\,(v^*_{SH} - v^*_N) > p_{BH|BH}\,(v^*_{BH} - v^*_{SH}) \iff v^*_{BH} - v^*_{SH} < (v^*_{SH} - v^*_N)\,\frac{p_{N|BH}}{p_{BH|BH}}.$$

[24] It is not possible for both $v^*_N > v^*_{SH}$ and $v^*_N > v^*_{BH}$ to hold because $p^a_N < p^a_{BH}$ and $p^b_N < p^b_{SH}$. This is enough to exclude the case $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_N$.

If $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then $v^*_{BH} < p^a_{BH}\,\vartheta/2 + \delta\, v^*_{BH}$ so that

$$v^*_{BH} - v^*_{SH} < p^a_{BH}\,\vartheta/2 + \delta\, v^*_{BH} - p^b_{SH}\,\vartheta/2 - \delta\, v^*_{BH} = (p^a_{BH} - p^b_{SH})\,\vartheta/2\,.$$

Now, the optimal action in $\omega = SH$ is LLB because $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$. Hence, since $v^*_N < \max\{p^a_N, p^b_N\}\vartheta/2 + \delta v^*_{BH}$

$$v^*_{SH} - v^*_N = v_{SH}(LLB) - v^*_N > p^b_{SH}\,\frac{\vartheta}{2} + \delta\, v^*_{BH} - \max\{p^a_N, p^b_N\}\,\frac{\vartheta}{2} - \delta\, v^*_{BH} = (p^b_{SH} - \max\{p^a_N, p^b_N\})\,\frac{\vartheta}{2}\,.$$

Therefore, if

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\})\,\frac{p_{N|BH}}{p_{BH|BH}}\,,$$

then LLS is preferred to LS because

$$(v^*_{SH} - v^*_N)\,\frac{p_{N|BH}}{p_{BH|BH}} > (p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{\vartheta}{2}\,\frac{p_{N|BH}}{p_{BH|BH}} > (p^a_{BH} - p^b_{SH})\,\frac{\vartheta}{2} > v^*_{BH} - v^*_{SH}\,.$$

For $\omega = SH$, we follow a similar reasoning so that if

$$p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\})\,\frac{p_{N|SH}}{p_{SH|SH}}\,,$$

then LLB is preferred to LB in $\omega = SH$.

For $\omega = N$, we first compare LLS with LS. As before, LLS is preferred to LS if and only if

$$v^*_{SH} > p_{BH|N}\,v^*_{BH} + p_{N|N}\,v^*_N + p_{SH|N}\,v^*_{SH}\,.$$

If $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{SH}$, then the former inequality holds. On the other hand, if $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then the last inequality holds if and only if

$$p_{BH|N}\,v^*_{SH} + p_{N|N}\,v^*_{SH} + p_{SH|N}\,v^*_{SH} > p_{BH|N}\,v^*_{BH} + p_{N|N}\,v^*_N + p_{SH|N}\,v^*_{SH} \iff$$

$$p_{N|N}\,(v^*_{SH} - v^*_N) > p_{BH|N}\,(v^*_{BH} - v^*_{SH}) \iff v^*_{BH} - v^*_{SH} < \frac{p_{N|N}}{p_{BH|N}}\,(v^*_{SH} - v^*_N)\,.$$

The remainder follows the same reasoning as that in the case with $\omega = BH$ so that if

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\})\,\frac{p_{N|N}}{p_{BH|N}}\,,$$

then LLS is preferred to LS in $\omega = N$.

Finally, using a similar reasoning, we have that if

$$p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\}) \frac{p_{N|N}}{p_{SH|N}},$$

then LLB is preferred to LB in $\omega = N$. ■

For $\omega = N$, if the value of α is sufficiently small, then we have the following action values

$$v_N(LLB) = p^b_N (\frac{\vartheta}{2} + \delta\, v^*_{BH}) + (1 - p^b_N)\, \delta\, v^*_{BH} = p^b_N \frac{\vartheta}{2} + \delta\, v^*_{BH},$$
$$v_N(LLS) = p^a_N (\frac{\vartheta}{2} + \delta\, v^*_{SH}) + (1 - p^a_N)\, \delta\, v^*_{SH} = p^a_N \frac{\vartheta}{2} + \delta\, v^*_{SH}.$$

For a sufficiently small value of α, the optimal action is either LLS or LS in $\omega = BH$, whereas the optimal action is either LLB or LB at $\omega = SH$. From Claim 5, we have $LS \prec LLS$ for $\omega = BH$ and $LB \prec LLB$ for $\omega = SH$. Therefore,

$$v_N(LLB) = p^b_N \frac{\vartheta}{2} + p^a_{BH} \frac{\vartheta}{2} (\delta + \delta^3 + ...) + p^b_{SH} \frac{\vartheta}{2} (\delta^2 + \delta^4 + ...) = p^b_N \frac{\vartheta}{2} + \frac{\delta\, p^a_{BH} \frac{\vartheta}{2} + \delta^2\, p^a_{BH} \frac{\vartheta}{2}}{1 - \delta^2},$$

and similarly

$$v_N(LLS) = p^a_N \frac{\vartheta}{2} + p^b_{SH} \frac{\vartheta}{2} (\delta + \delta^3 + ...) + p^a_{BH} \frac{\vartheta}{2} (\delta^2 + \delta^4 + ...) = p^a_N \frac{\vartheta}{2} + \frac{\delta\, p^b_{SH} \frac{\vartheta}{2} + \delta^2\, p^a_{BH} \frac{\vartheta}{2}}{1 - \delta^2}.$$

Hence, LLB is preferred to LLS if and only if

$$p^b_N \frac{\vartheta}{2} (1 - \delta^2) + \delta\, p^a_{BH} \frac{\vartheta}{2} + \delta^2\, p^b_{SH} \frac{\vartheta}{2} > p^a_N \frac{\vartheta}{2} (1 - \delta^2) + \delta\, p^b_{SH} \frac{\vartheta}{2} + \delta^2\, p^a_{BH} \frac{\vartheta}{2} \iff$$

$$(p^b_N - p^a_N)(1 - \delta^2) > (\delta - \delta^2)(p^b_{SH} - p^a_{BH}) \iff (p^b_N - p^a_N) > \frac{\delta}{1 + \delta} (p^b_{SH} - p^a_{BH}).$$

Similarly, LLS is preferred to LLB if and only if $(p^a_N - p^b_N) > \frac{\delta}{1+\delta} (p^a_{BH} - p^b_{SH})$. □

Proof of Theorem 2 The result is immediate from Lemmas 5, 6, and 9. □

Proof of Lemma 3 We compare the value of two strategies. Specifically, we consider an optimal stationary pure Markov strategy σ^* and a strategy σ that is suboptimal.

We run both the suboptimal strategy and the optimal strategy until the states match, and then the suboptimal strategy plays according to the optimal strategy. Here, the target state continues to change as a consequence of running the optimal strategy; so the suboptimal strategy is defined to follow the inventory level of the optimal strategy. Specifically, if the optimal action leads to the inventory staying at the same level, then the suboptimal strategy does nothing; if the optimal action leads to the inventory increasing by one unit, then the suboptimal strategy submits a buy market order; finally, if the optimal action leads to the inventory decreasing by one unit, then the suboptimal strategy submits a sell market order. Therefore, the states of the two chains are always at the same level of inventory, so the difference in the payoff received (at each step) is less than ϑ.

Now, for the states to match, the volume imbalance regime ω also needs to be the same. Observe that at each time step, the probability that the two chains meet at the same volume imbalance regime (after one step) is greater than m, where m is the minimum element of the transition probability matrix for the Markov chain given in Figure 1b. Hence, the hitting time is dominated by a geometric random variable with success probability m; thus, the expectation of the hitting time is less than $1/m$. Therefore, we have

$$v^*_{\omega,q} - v^*_{\omega',q} \leq v^*_{\omega,q} - v_{\omega',q}(\sigma) \leq \vartheta/m$$

because the discount parameter $\delta < 1$. □

Proof of Proposition 2 From the action values, $LLB \prec LLS$ if and only if

$$\alpha > \frac{\left(p^b_\omega - p^a_\omega\right)\vartheta/2}{p^a_\omega(2q-1) + p^b_\omega(2q+1)} + \delta \frac{p^b_\omega\left(v^*_{BH,q+1} - v^*_{BH,q}\right) + p^a_\omega\left(v^*_{SH,q} - v^*_{SH,q-1}\right) + \left(v^*_{BH,q} - v^*_{SH,q}\right)}{p^a_\omega(2q-1) + p^b_\omega(2q+1)}.$$

We use the upper bound from Lemma 3 and the upper bound $v^*_{\omega,q} - v^*_{\omega,q-1} \leq 0$ from Lemma 1, to obtain

$$\alpha_1(\omega, q) \leq \frac{\left(p^b_\omega - p^a_\omega\right)\vartheta/2 + \vartheta/m}{p^a_\omega(2q-1) + p^b_\omega(2q+1)} = \bar{\alpha}_1(\omega, q)$$

as an upper bound for $\alpha_1(\omega, q)$ that is strictly positive.

From Lemma 5, we have $\alpha_3(\omega, q) > 0$ for all $\omega \in \Omega$ and $q > 0$. Therefore, the result follows because $\alpha_1(\omega, q) \wedge \alpha_3(BH, q) \leq \bar{\alpha}_1(\omega, q)$ and $\alpha_1(\omega, q) \wedge \alpha_3(BH, q+1) \leq \bar{\alpha}_1(\omega, q)$. □

Proof of Proposition 3 For a fixed volume imbalance regime ω, the function $\bar{\alpha}_1(\omega, q)$ monotonically decreases as the absolute value of q increases. The choice of α ensures that $\alpha \notin I'(\mathbf{s})$ for all

states $\mathbf{s} = (\omega, q)$ where $q \neq 0$, so the result follows. □

Proof of Proposition 4 If $\vartheta \to 0$, then $\bar{\alpha}_1(\omega, q) \to 0$ for all $\omega \in \Omega$ and $q \neq 0$. Therefore, the result follows as a consequence of Proposition 3. □

Lemma 10 *If condition (C4) holds, then Proposition 1 continues to hold.*

Proof of Lemma 10 The first 8 comparisons follow from the same reasoning in Proposition 1. Therefore, we consider the comparisons LS vs LLS and LB vs LLB. We consider a reduced action set without MS. For a large enough value of α, we have $v^*_{BH,q} > v^*_{N,q} > v^*_{SH,q}$ and $v^*_{BH,q-1} > v^*_{N,q-1} > v^*_{SH,q-1}$ because the probability of selling a limit order is highest (lowest) in BH (SH). The action values of LLS and LS (excluding the one-step utility) are

$$v_{\omega,q}(LLS) = p^a_\omega \left((1-\kappa)\, v^*_{SH,q-1} + \frac{\kappa}{2}\, v^*_{N,q-1} + \frac{\kappa}{2}\, v^*_{BH,q-1}\right) + (1-p^a_\omega) \left((1-\kappa)\, v^*_{SH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{BH,q}\right)$$

$$v_{\omega,q}(LS) = p^a_\omega \left(p_{SH|\omega}\, v^*_{SH,q-1} + p_{N|\omega}\, v^*_{N,q-1} + p_{BH|\omega}\, v^*_{BH,q-1}\right) + (1-p^a_\omega) \left(p_{SH|\omega}\, v^*_{SH,q} + p_{N|\omega}\, v^*_{N,q} + p_{BH|\omega}\, v^*_{BH,q}\right),$$

respectively. Therefore, $v_{\omega,q}(LLS) < v_{\omega,q}(LS)$ if and only if

$$\begin{aligned}
&(1-\kappa-p_{SH|\omega}) \left(p^a_\omega\, v^*_{SH,q-1} + (1-p^a_\omega)\, v^*_{SH,q}\right) < \\
&p^a_\omega \left((p_{N|\omega} - \frac{\kappa}{2})\, v^*_{N,q-1} + (p_{BH|\omega} - \frac{\kappa}{2})\, v^*_{BH,q-1}\right) + (1-p^a_\omega) \left((p_{N|\omega} - \frac{\kappa}{2})\, v^*_{N,q} + (p_{BH|\omega} - \frac{\kappa}{2})\, v^*_{BH,q}\right),
\end{aligned}$$

which follows because

$$\begin{aligned}
&\left(p_{N|\omega} - \frac{\kappa}{2}\right) v^*_{N,q-1} + \left(p_{BH|\omega} - \frac{\kappa}{2}\right) v^*_{BH,q-1} > (p_{N|\omega} + p_{BH|\omega} - \kappa)\, v^*_{N,q-1} \\
&= (1-\kappa-p_{SH|\omega})\, v^*_{N,q-1} > (1-\kappa-p_{SH|\omega})\, v^*_{SH,q-1}
\end{aligned}$$

and

$$\begin{aligned}
&\left(p_{N|\omega} - \frac{\kappa}{2}\right) v^*_{N,q} + \left(p_{BH|\omega} - \frac{\kappa}{2}\right) v^*_{BH,q} > (p_{N|\omega} + p_{BH|\omega} - \kappa)\, v^*_{N,q} \\
&= (1-\kappa-p_{SH|\omega})\, v^*_{N,q} > (1-\kappa-p_{SH|\omega})\, v^*_{SH,q}\,.
\end{aligned}$$

Hence, LS is preferred to LLS for a large enough value of α. Next, we compare LLB and LB. The action values of LLB and LB (excluding the one-step utility) are

$$v_{\omega,q}(LLB) = p^b_\omega \left((1-\kappa)\, v^*_{BH,q+1} + \frac{\kappa}{2}\, v^*_{N,q+1} + \frac{\kappa}{2}\, v^*_{SH,q+1}\right) + (1-p^b_\omega) \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q}\right)$$

$$v_{\omega,q}(LB) = p^b_\omega \left(p_{BH|\omega}\, v^*_{BH,q+1} + p_{N|\omega}\, v^*_{N,q+1} + p_{SH|\omega}\, v^*_{SH,q+1}\right) + (1-p^b_\omega)\left(p_{BH|\omega}\, v^*_{BH,q} + p_{N|\omega}\, v^*_{N,q} + p_{SH|\omega}\, v^*_{SH,q}\right),$$

respectively. Thus, $v_{\omega,q}(LLB) > v_{\omega,q}(LB)$ if and only if

$$\begin{aligned}&(1-\kappa-p_{BH|\omega})\left(p^b_\omega\, v^*_{BH,q+1} + (1-p^b_\omega)\, v^*_{BH,q}\right) > \\ &p^b_\omega \left((p_{N|\omega}-\frac{\kappa}{2})\, v^*_{N,q+1} + (p_{SH|\omega}-\frac{\kappa}{2})\, v^*_{SH,q+1}\right) + (1-p^b_\omega)\left((p_{N|\omega}-\frac{\kappa}{2})\, v^*_{N,q} + (p_{SH|\omega}-\frac{\kappa}{2})\, v^*_{SH,q}\right),\end{aligned}$$

which follows because

$$\begin{aligned}&\left(p_{N|\omega}-\frac{\kappa}{2}\right) v^*_{N,q+1} + \left(p_{SH|\omega}-\frac{\kappa}{2}\right) v^*_{SH,q+1} < (p_{N|\omega}+p_{SH|\omega}-\kappa)\, v^*_{N,q+1} \\ &= (1-\kappa-p_{BH|\omega})\, v^*_{N,q+1} < (1-\kappa-p_{BH|\omega})\, v^*_{BH,q+1}\end{aligned}$$

and

$$\begin{aligned}&\left(p_{N|\omega}-\frac{\kappa}{2}\right) v^*_{N,q} + \left(p_{SH|\omega}-\frac{\kappa}{2}\right) v^*_{SH,q} < (p_{N|\omega}+p_{SH|\omega}-\kappa)\, v^*_{N,q} \\ &= (1-\kappa-p_{BH|\omega})\, v^*_{N,q} < (1-\kappa-p_{BH|\omega})\, v^*_{BH,q}\,.\end{aligned}$$

The same reasoning holds for $q < 0$. □

Lemma 11 *If condition (C4) holds, then Lemma 6 continues to hold.*

Proof of Lemma 11 Suppose that LLB is optimal at $(BH, q > 0)$, then

$$\frac{\vartheta}{2} - \alpha\,(q+1)^2 + \delta\left((1-\kappa)\, v^*_{BH,q+1} + \frac{\kappa}{2}\, v^*_{N,q+1} + \frac{\kappa}{2}\, v^*_{SH,q+1}\right) > -\alpha\, q^2 + \delta\left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q}\right).$$

Suppose the inequality above is not true. We know $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{BH,q}$ because if the maximum was $v^*_{SH,q}$, then $v_{BH,q}(LLS) > v_{BH,q}(LLB)$, which cannot be true assuming LLB is optimal at $(BH, q > 0)$. Therefore, if both $v^*_{BH,q}$ is the maximum and LLB is optimal at (BH, q), then staying at $(BH, q > 0)$ forever would be optimal. However, this would give the same payoff as doing nothing forever, which is never optimal. Therefore the previous inequality holds. From the previous inequality, we have $v_{SH,q}(LLB) > v_{BH,q}(LLB)$ because $p^b_{SH} > p^b_{BH}$. Therefore, $v^*_{BH,q} < v^*_{SH,q}$. Now, observe that

$$v_{BH,q}(LLB) = p^b_{BH}\left(\frac{\vartheta}{2} - \alpha\,(q+1)^2 + \delta\left((1-\kappa)\, v^*_{BH,q+1} + \frac{\kappa}{2}\, v^*_{N,q+1} + \frac{\kappa}{2}\, v^*_{SH,q+1}\right)\right)$$

$$
\begin{aligned}
&+ (1 - p^b_{BH}) \left(-\alpha\, q^2 + \delta \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q} \right) \right) \\
&= p^b_{BH} \left(\frac{\vartheta}{2} - \alpha\, (q+1)^2 + \delta \left((1-\kappa)\, v^*_{BH,q+1} + \frac{\kappa}{2}\, v^*_{N,q+1} + \frac{\kappa}{2}\, v^*_{SH,q+1} \right) \right) \\
&\quad + (p^a_{BH} - p^b_{BH}) \left(-\alpha\, q^2 + \delta \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q} \right) \right) \\
&\quad + (1 - p^b_{BH}) \left(-\alpha\, q^2 + \delta \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q} \right) \right) \\
&\leq p^a_{BH} \left(\frac{\vartheta}{2} - \alpha\, (q+1)^2 + \delta \left((1-\kappa)\, v^*_{BH,q+1} + \frac{\kappa}{2}\, v^*_{N,q+1} + \frac{\kappa}{2}\, v^*_{SH,q+1} \right) \right) \\
&\quad + (1 - p^a_{BH}) \left(-\alpha\, q^2 + \delta \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q} \right) \right) \\
&\leq p^a_{BH} \left(\frac{\vartheta}{2} - \alpha\, (q-1)^2 + \delta \left((1-\kappa)\, v^*_{BH,q-1} + \frac{\kappa}{2}\, v^*_{N,q-1} + \frac{\kappa}{2}\, v^*_{SH,q-1} \right) \right) \\
&\quad + (1 - p^a_{BH}) \left(-\alpha\, q^2 + \delta \left((1-\kappa)\, v^*_{BH,q} + \frac{\kappa}{2}\, v^*_{N,q} + \frac{\kappa}{2}\, v^*_{SH,q} \right) \right) = v_{BH,q}(LLS)\,.
\end{aligned}
$$

Hence, a contradiction, so LLB is never optimal at $(BH, q > 0)$. The same reasoning holds for $q < 0$. □

Lemma 12 *If condition (C4) holds, then Lemma 7 continues to hold.*

Proof of Lemma 12 It is straightforward to see that Claim 4 still holds because the transition probabilities of large limit orders do no play a role in its proof. To show that Lemma 7 continues to hold, we prove that LLS is not optimal in $(SH, q > 0)$ for a small enough value of α. We proceed by contradiction. Assuming that LLS is optimal in $(SH, q > 0)$, for a small enough value of α, we have

$$
\begin{aligned}
v_{SH}(LLB) &= p^b_{SH} \left(\frac{\vartheta}{2} + \delta \left((1-\kappa)\, v^*_{BH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{SH} \right) \right. \\
&\quad \left. + (1 - p^b_{SH})\, \delta \left((1-\kappa)\, v^*_{BH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{SH} \right) \right) \\
&= p^b_{SH}\, \frac{\vartheta}{2} + \delta \left((1-\kappa)\, v^*_{BH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{SH} \right),
\end{aligned}
$$

$$
\begin{aligned}
v_{SH}(LLS) &= p^a_{SH} \left(\frac{\vartheta}{2} + \delta \left((1-\kappa)\, v^*_{SH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{BH} \right) \right. \\
&\quad \left. + (1 - p^a_{SH})\, \delta \left((1-\kappa)\, v^*_{SH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{BH} \right) \right) \\
&= p^a_{SH}\, \frac{\vartheta}{2} + \delta \left((1-\kappa)\, v^*_{SH} + \frac{\kappa}{2}\, v^*_{N} + \frac{\kappa}{2}\, v^*_{BH} \right)
\end{aligned}
$$

We have that $v^*_{BH} > v^*_{SH}$ because $p^a_{BH} > p^a_{SH}$ and there is less penalty at $q-1$ than at q and because we assume that LLS is optimal at $(SH, q > 0)$. Therefore, use $p^b_{SH} > p^a_{SH}$, $v^*_{BH} > v^*_{SH}$, and $(1-\kappa) > \frac{\kappa}{2}$ to obtain

$$p^b_{SH}\frac{\vartheta}{2} + \delta\left((1-\kappa)\,v^*_{BH} + \frac{\kappa}{2}\,v^*_N + \frac{\kappa}{2}\,v^*_{SH}\right) > p^a_{SH}\frac{\vartheta}{2} + \delta\left((1-\kappa)\,v^*_{SH} + \frac{\kappa}{2}\,v^*_N + \frac{\kappa}{2}\,v^*_{BH}\right)$$

so that $v_{SH}(LLB) > v_{SH}(LLS) = v^*_{SH}$, and therefore a contradiction. The same reasoning holds for $q < 0$. □

Remark 1 *If condition* (C4) *holds, then Lemma 8 continues to hold because the transition probabilities of the large limit orders do not play a role in the proof of Lemma 8.*

Proof of Theorem 3 The result is immediate from Lemmas 10, 11 and 12 and Remark 1. □

B. Additional Tables and Figures

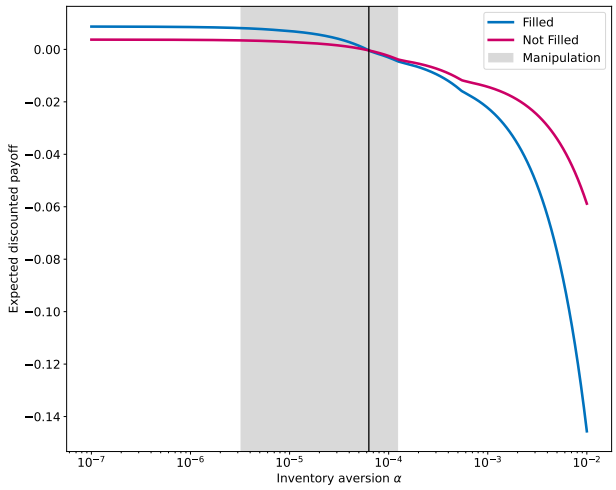
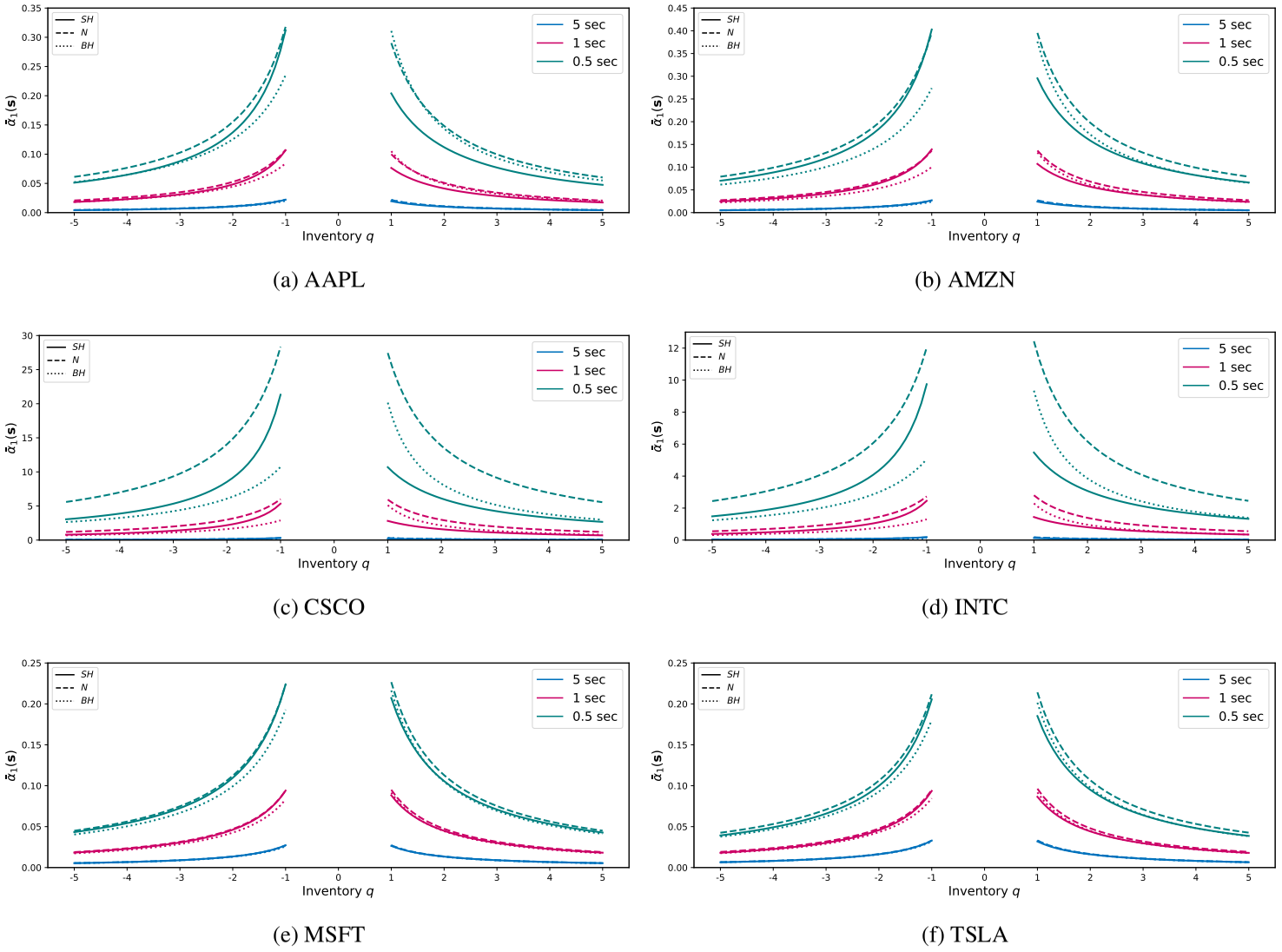


Figure 6: Expected stream of discounted payoffs when the manipulative order is filled or not for CSCO with 1 second decision intervals in $\mathbf{s} = (N, q = 2)$. For values of the inventory aversion parameter α below the shaded region, manipulation is no longer optimal because the optimal action is LB as a consequence of Proposition 1 and Lemma 6.



(a) AAPL (b) AMZN

(c) CSCO (d) INTC

(e) MSFT (f) TSLA

Figure 7: Intervals $I'(\mathbf{s})$.

Table 12: Transition probability matrix.

(a) AAPL: 5 seconds

	SH	N	BH
SH	0.32	0.41	0.28
N	0.28	0.43	0.29
BH	0.27	0.41	0.32

(b) AAPL: 1 second

	SH	N	BH
SH	0.42	0.37	0.21
N	0.26	0.48	0.26
BH	0.20	0.37	0.43

(c) AAPL: 0.5 seconds

	SH	N	BH
SH	0.52	0.32	0.16
N	0.22	0.55	0.23
BH	0.15	0.33	0.52

(d) AMZN: 5 seconds

	SH	N	BH
SH	0.35	0.41	0.24
N	0.3	0.44	0.26
BH	0.27	0.41	0.32

(e) AMZN: 1 second

	SH	N	BH
SH	0.47	0.35	0.18
N	0.26	0.51	0.23
BH	0.20	0.36	0.43

(f) AAPL: 0.5 seconds

	SH	N	BH
SH	0.56	0.31	0.13
N	0.23	0.57	0.21
BH	0.15	0.32	0.52

(g) CSCO: 5 seconds

	SH	N	BH
SH	0.52	0.39	0.09
N	0.15	0.7	0.15
BH	0.08	0.38	0.54

(h) CSCO: 1 second

	SH	N	BH
SH	0.76	0.21	0.03
N	0.08	0.84	0.08
BH	0.03	0.21	0.76

(i) CSCO: 0.5 seconds

	SH	N	BH
SH	0.82	0.16	0.02
N	0.06	0.87	0.07
BH	0.01	0.16	0.83

(j) INTC: 5 seconds

	SH	N	BH
SH	0.48	0.41	0.11
N	0.17	0.67	0.16
BH	0.11	0.43	0.45

(k) INTC: 1 second

	SH	N	BH
SH	0.71	0.24	0.04
N	0.10	0.81	0.09
BH	0.04	0.26	0.70

(l) INTC: 0.5 seconds

	SH	N	BH
SH	0.79	0.18	0.03
N	0.08	0.85	0.07
BH	0.02	0.20	0.78

(m) MSFT: 5 seconds

	SH	N	BH
SH	0.38	0.34	0.28
N	0.33	0.36	0.31
BH	0.31	0.34	0.35

(n) MSFT: 1 second

	SH	N	BH
SH	0.46	0.32	0.22
N	0.31	0.40	0.29
BH	0.24	0.33	0.43

(o) MSFT: 0.5 seconds

	SH	N	BH
SH	0.52	0.30	0.18
N	0.29	0.44	0.27
BH	0.20	0.30	0.49

(p) TSLA: 5 seconds

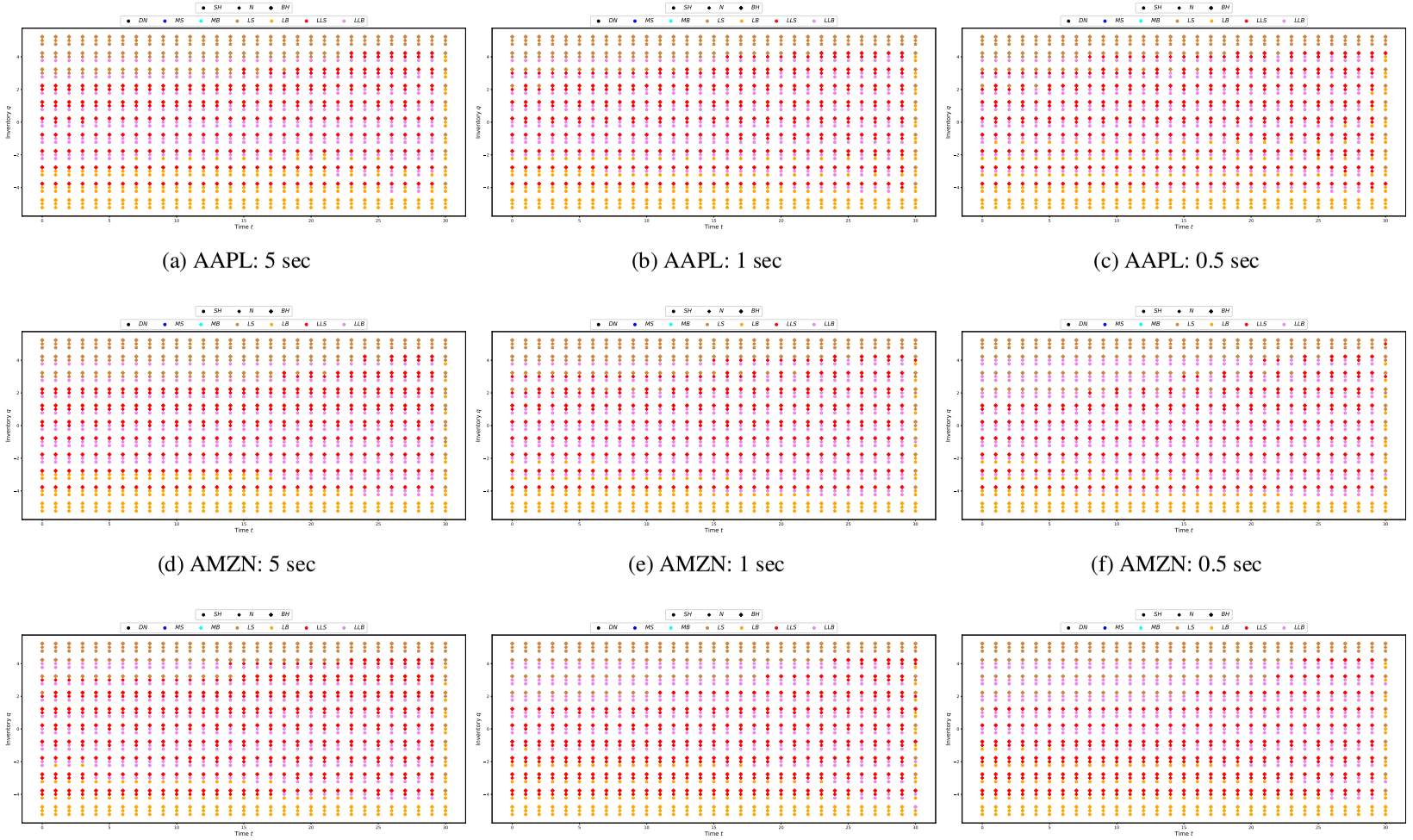
	SH	N	BH
SH	0.34	0.39	0.27
N	0.30	0.41	0.29
BH	0.28	0.39	0.32

(q) TSLA: 1 second

	SH	N	BH
SH	0.42	0.37	0.22
N	0.28	0.46	0.27
BH	0.23	0.37	0.41

(r) TSLA: 0.5 seconds

	SH	N	BH
SH	0.49	0.34	0.17
N	0.25	0.50	0.25
BH	0.18	0.33	0.48

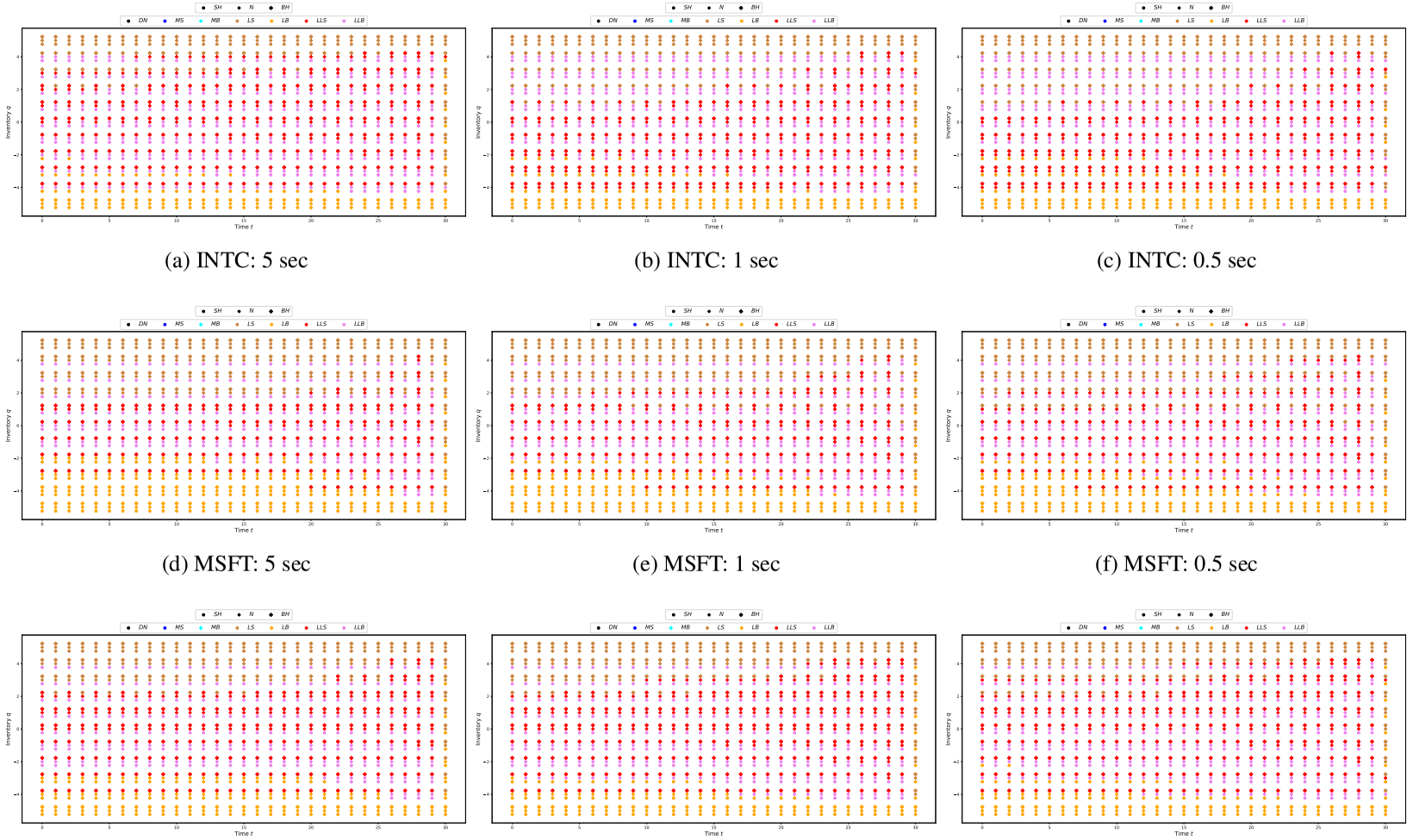


(a) AAPL: 5 sec (b) AAPL: 1 sec (c) AAPL: 0.5 sec

(d) AMZN: 5 sec (e) AMZN: 1 sec (f) AMZN: 0.5 sec

(g) CSCO: 5 sec (h) CSCO: 1 sec (i) CSCO: 0.5 sec

Figure 8: Optimal action for a finite trading horizon with $\alpha = \times 10^{-5}$.



(a) INTC: 5 sec (b) INTC: 1 sec (c) INTC: 0.5 sec

(d) MSFT: 5 sec (e) MSFT: 1 sec (f) MSFT: 0.5 sec

(g) TSLA: 5 sec (h) TSLA: 1 sec (i) TSLA: 0.5 sec

Figure 9: Optimal action for a finite trading horizon with $\alpha = \times 10^{-5}$.

Table 13: Baseline: Average number of manipulation sequences over 50 trading intervals.

Ticker	Decision Interval Δt	Zero inventory		Same inventory		Opposing inventory	
		Agent 1 $q=0$	Agent 2 $q=0$	Agent 1 $q=4$	Agent 2 $q=4$	Agent 1 $q=4$	Agent 2 $q=-4$
AAPL	5 seconds	24.77	20.78	20.80	25.72	21.80	22.45
	1 second	25.04	13.88	14.09	28.72	18.56	19.10
	0.5 seconds	26.23	11.56	11.68	31.48	16.36	15.08
INTC	5 seconds	25.28	17.27	17.27	28.75	20.27	18.88
	1 second	32.00	17.00	17.05	34.83	12.59	13.14
	0.5 seconds	35.22	12.04	12.03	39.12	9.40	9.92
MSFT	5 seconds	25.07	21.68	21.70	25.43	22.20	21.97
	1 second	25.55	16.87	16.87	27.33	19.15	18.86
	0.5 seconds	26.31	10.32	10.30	29.82	15.30	15.95
TSLA	5 seconds	24.97	21.85	21.87	25.41	22.41	22.31
	1 second	25.32	18.50	18.55	27.38	20.65	21.01
	0.5 seconds	26.13	14.29	14.17	29.25	18.64	19.48

Table 14: Offline learning: Average number of manipulation sequences over 50 trading intervals.

Ticker	Decision Interval Δt	Zero inventory		Same inventory		Opposing inventory	
		Agent 1 $q=0$	Agent 2 $q=0$	Agent 1 $q=4$	Agent 2 $q=4$	Agent 1 $q=4$	Agent 2 $q=-4$
AAPL	5 seconds	24.86	26.09	20.96	22.20	20.79	23.00
	1 second	26.31	29.13	15.75	19.00	14.82	20.25
	0.5 seconds	28.92	31.90	14.34	16.79	13.54	17.38
INTC	5 seconds	27.79	28.89	21.76	20.62	22.16	19.44
	1 second	35.28	35.48	17.76	14.53	21.76	17.39
	0.5 seconds	39.53	39.49	12.18	10.66	21.17	18.93
MSFT	5 seconds	25.03	25.77	21.55	22.54	21.72	22.38
	1 second	25.94	28.10	17.56	20.08	17.42	19.88
	0.5 seconds	27.02	30.51	11.30	16.06	11.24	17.27
TSLA	5 seconds	25.07	25.91	21.98	22.97	21.90	22.86
	1 second	25.53	27.66	18.92	21.24	18.73	21.36
	0.5 seconds	26.32	29.22	14.83	19.35	14.91	20.09

Lemma 13 *Let $q > 0$. If $LLB \prec LB$ at (SH,q), then $LS \prec LLS$ at (BH,q).*

Table 15: Online learning: Average number of manipulation sequences over 50 trading intervals.

Ticker	Decision Interval Δt	Zero inventory		Same inventory		Opposing inventory	
		Agent 1 $q=0$	Agent 2 $q=0$	Agent 1 $q=4$	Agent 2 $q=4$	Agent 1 $q=4$	Agent 2 $q=-4$
AAPL	5 seconds	24.68	25.76	20.80	21.81	20.63	22.40
	1 second	21.80	28.62	12.57	18.25	11.84	19.39
	0.5 seconds	20.42	31.51	1.49	15.05	1.41	15.10
INTC	5 seconds	22.89	28.53	14.26	19.96	13.75	18.8
	1 second	20.26	34.81	0.0	12.36	0.0	12.39
	0.5 seconds	27.04	38.39	0.0	14.77	0.0	16.13
MSFT	5 seconds	24.65	25.38	21.27	22.05	21.32	21.85
	1 second	23.84	27.33	15.48	18.88	15.41	18.67
	0.5 seconds	23.03	29.36	8.53	14.31	8.51	14.37
TSLA	5 seconds	24.56	25.44	21.44	22.31	21.47	22.30
	1 second	24.30	27.23	17.67	20.79	17.65	20.86
	0.5 seconds	24.18	29.19	12.78	18.91	12.67	19.02

Table 16: Offline learning: Average manipulation statistics.

(a) Percentage of large orders on opposite sides over 50 trading intervals.

Ticker	Δt	Zero inv.	Same inv.	Opposing inv.
AAPL	5s	0.1002%	0.1008%	0.4369%
	1s	0.1187%	1.2180%	0.0044%
	0.5s	0.0250%	1.0324%	0%
INTC	5s	0.1235%	1.1789%	0.6517%
	1s	0.0609%	0.0153%	4.9916%
	0.5s	0%	0.0005%	4.1832%
MSFT	5s	0.6766%	1.6972%	0.1863%
	1s	0.2604%	1.0269%	0.0401%
	0.5s	0.1993%	0.7633%	0.0095%
TSLA	5s	0.1195%	0.1283%	0.5394%
	1s	0.3506%	0.2716%	1.9706%
	0.5s	1.7434%	1.6606%	0.9519%

(b) Number of times where only one market maker submits a large order over 50 trading intervals.

Ticker	Δt	Zero inv.	Same inv.	Opposing inv.
AAPL	5s	13.28	18.26	18.47
	1s	23.10	25.81	27.52
	0.5s	20.75	24.17	28.98
INTC	5s	21.58	20.71	29.97
	1s	14.64	8.99	35.89
	0.5s	10.10	4.29	39.34
MSFT	5s	25.92	25.22	27.84
	1s	24.00	22.06	26.58
	0.5s	22.07	17.58	26.12
TSLA	5s	13.48	17.35	18.69
	1s	24.01	23.99	27.89
	0.5s	22.65	24.07	29.37

Table 17: Online learning: Average manipulation statistics.

(a) Percentage of large orders on opposite sides over 50 trading intervals.

Ticker	Δt	Zero inv.	Same inv.	Opposing inv.
AAPL	5s	0.2159%	0.4622%	0.5406%
	1s	0.3898%	1.3460%	3.5092%
	0.5s	0.6900%	0%	0%
INTC	5s	0.3182%	1.6649%	1.7663%
	1s	1.7169%	0%	0%
	0.5s	1.8299%	8.2368%	0%
MSFT	5s	0.0366%	0.2648%	0.5282%
	1s	0.1880%	2.0196%	2.4215%
	0.5s	0.3387%	0%	6.8309%
TSLA	5s	0.0415%	0.2238%	0.3103%
	1s	0.3340%	1.3380%	1.8276%
	0.5s	0.4559%	3.1344%	4.5598%

(b) Number of times where only one market maker submits a large order over 50 trading intervals.

Ticker	Δt	Zero inv.	Same inv.	Opposing inv.
AAPL	5s	19.24	23.47	19.97
	1s	25.29	26.97	24.09
	0.5s	26.78	27.36	24.69
INTC	5s	24.01	26.20	25.41
	1s	25.89	20.79	23.26
	0.5s	19.58	29.51	32.50
MSFT	5s	21.35	22.15	22.31
	1s	23.32	23.67	23.77
	0.5s	23.81	22.90	23.22
TSLA	5s	18.51	20.67	19.76
	1s	21.67	24.39	23.59
	0.5s	22.15	25.94	24.22

Proof Suppose not. Then $LLB \prec LB$ at (SH, q) and $LLS \prec LS$ at (BH, q). We note that $LLS \prec LS$ at $\omega = BH \iff$

$$p^a_\omega(m_{BH|\omega}v_{BH,q-1}+m_{N|\omega}v_{N,q-1}+m_{SH|\omega}v_{SH,q-1})+(1-p^a_\omega)(m_{BH|\omega}v_{BH,q}+m_{N|\omega}v_{N,q}+m_{SH|\omega}v_{SH,q}) >$$

$$p^a_\omega v_{SH,q-1} + (1 - p^a_\omega)v_{SH,q}$$

so that v_{SH} cannot be maximum at both $q-1$ and q (that is, it cannot be true that $\max\{v^*_{BH,q-1}, v^*_{N,q-1}, v^*_{SH,q-1}\} = v^*_{SH,q-1}$ and $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{SH,q}$). But note that if $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{SH,q}$, then also $\max\{v^*_{BH,q-1}, v^*_{N,q-1}, v^*_{SH,q-1}\} = v^*_{SH,q-1}$, so that at level q we cannot have v_{SH} as maximum, that is, we cannot have $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{SH,q}$.
Similarly, $LLB \prec LB$ at $\omega = SH \iff$

$$p^b_\omega(m_{BH|\omega}v_{BH,q+1}+m_{N|\omega}v_{N,q+1}+m_{SH|\omega}v_{SH,q+1})+(1-p^b_\omega)(m_{BH|\omega}v_{BH,q}+m_{N|\omega}v_{N,q}+m_{SH|\omega}v_{SH,q}) >$$

$$p^b_\omega v_{BH,q+1} + (1 - p^b_\omega)v_{BH,q}$$

so that v_{BH} cannot be maximum at both q and $q+1$ (that is, it cannot be true that $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{BH,q}$ and $\max\{v^*_{BH,q+1}, v^*_{N,q+1}, v^*_{SH,q+1}\} = v^*_{BH,q+1}$). But note that if $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} =$

$v^*_{BH,q}$ then we also have $\max\{v^*_{BH,q+1}, v^*_{N,q+1}, v^*_{SH,q+1}\} = v^*_{BH,q+1}$, so that at level q we cannot have v_{BH} as maximum (hence we cannot have $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{BH,q}$).
Therefore, we have seen that if both $LLB \prec LB$ at (SH, q) and $LLS \prec LS$ at (BH, q), then v_{BH} and v_{SH} are not maximum a level q, so that $\max\{v^*_{BH,q}, v^*_{N,q}, v^*_{SH,q}\} = v^*_{N,q}$, which is impossible due to $p^a_{BH} > p^a_N > p^a_{SH}$ and $p^b_{BH} < p^b_N < p^b_{SH}$. We then get a contradiction and consequently the lemma holds.

C. Non martingale case

In the original scenario we had that

$$E\left[Y(s,a,s')\right] = \begin{cases} p^b_\omega\,\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q + p^b_\omega\,\varphi) & \text{for } a = \{LB, LLB\}\,, \\ p^a_\omega\,\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q - p^a_\omega\,\varphi) & \text{for } a = \{LS, LLS\}\,, \\ -\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q + \varphi) & \text{for } a = MB\,, \\ -\vartheta/2 + (2\,\beta - 1)\,(\varphi\,q - \varphi) & \text{for } a = MS\,, \\ (2\,\beta - 1)\,\varphi\,q & \text{for } a = DN\,, \end{cases}$$

which simplified since $2\beta - 1 = 0$. In the new case we assume β depends on the regime. Moreover, by symmetry, we will assume $\beta(BH) = 1 - \beta(SH)$ and $\beta(N) = \frac{1}{2}$. This captures the idea that when the book is buy-heavy (sell-heavy) the price of the asset tends to increase (decrease). In this new framework $E\left[Y(s,a,s')\right]$ depends not just on the action a but also on the current state s (since the transition probabilities regarding the book regime depend on the current state). This does not allow us to immediately get rid of MB and DN as we did in the previous case, since they may be optimal for some range of α's in this new scenario.
We proceed as follows. Let $q > 0$. For a given book regime ω, the previous $2\beta - 1$ becomes $2\beta(BH) - 1$ if the action is LLB, $2\beta(SH) - 1$ if the action is LLS and for the rest of actions it becomes:

$$m_{BH|\omega}(2\beta(BH) - 1) + m_{N|\omega}(2\beta(N) - 1) + m_{SH|\omega}(2\beta(SH) - 1) =$$

$$m_{BH|\omega}(2\beta(BH) - 1) + m_{SH|\omega}(1 - 2\beta(BH)) = (m_{BH|\omega} - m_{SH|\omega})(2\beta(BH) - 1).$$

As $\beta(BH) > \frac{1}{2}$ so that $2\beta(BH) - 1 > 0$, the sign of the previous expression depends on the sign of $m_{BH|\omega} - m_{SH|\omega}$. But this is positive for $\omega = BH$ and negative for $\omega = SH$ (reasonable assumption seems to be). Given that, we establish the following results:

Lemma 14 *Let $\omega = SH$ and $q > 0$. The action DN is never optimal.*

Proof We claim DN is dominated by LS when $q > 0$ and $\omega = SH$. The continuation values are the same for both actions, so they do not play a role in the comparison. Observe that the LS is better than DN both in terms of penalty (since $q > 0$) and immediate reward not related to change in the price of the asset (since with LS we gain $p^a_{SH}\frac{\theta}{2} > 0$ and 0 is what we gain with DN). So far this is the reasoning used for the martingale case. Here, we finally focus in the immediate reward related to the change in the price of the asset. By playing LS we get:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi(q - p^a_{SH})$$

whereas by playing DN we get:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi q$$

But since both expressions are negative for any positive q, and also $q > q - p^a_{SH}$, this immediate reward related to the change in the price favours LS in comparison to DN as well. Therefore the lemma follows.

Lemma 15 *Let $\omega = SH$ and $q > 0$. The action that gets more extra gain in the non-martingale case compared to the martingale case is LLB.*

Proof The extra gains in the non-martingale case, given $\omega = SH$ and $q > 0$, are the following:

For $a = LLB$:

$$(2\beta(BH) - 1)\phi(q + p^b_{BH})$$

For $a = LLS$:

$$(2\beta(SH) - 1)\phi(q - p^a_{BH})$$

For $a = LB$:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi(q + p^b_{BH})$$

For $a = LS$:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi(q - p^a_{BH})$$

For $a = MB$:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi(q + 1)$$

For $a = MS$:

$$(m_{BH|SH} - m_{SH|SH})(2\beta(BH) - 1)\phi(q - 1)$$

The only positive expression of the former 6 is the one corresponding to LLB since $(2\beta(SH) - 1) < 0$, $(2\beta(BH) - 1) > 0$ and $(m_{BH|SH} - m_{SH|SH}) < 0$ so that the lemma follows.

Lemma 16 *We still have $\alpha_1(SH, q) > 0$ in the non-martingale case*

Proof Due to Lemma 1.1 we have DN is never optimal at SH, which allows us to have the same order that in the martingale case (DN would be problematic regarding penalty, since it is the action that lies between LLB and LLS regarding the penalty cutoffs). Note that MB is not causing problems regarding the order, since it is the one that gives less penalty. Moreover, as Lemma 1.2 says that LLB is even more attractive in the non-martingale case compared to the martingale case (when we compare it with any other action), the range of values of α where LLB is optimal in the martingale case is included in the range of values of α where LLB is optimal in the non-martingale case. Therefore, since we already had $\alpha_1(SH, q) > 0$ in the martingale case, we must have $\alpha_1(SH, q) > 0$ in the non-martingale case.

We did not consider the action DN in this previous lemma since we have already proved it is never optimal.

Lemma 17 *Assume $\theta = k\phi$ for $k \geq 2$. Then for $\omega = BH$, the action DN is never optimal.*

Proof We again have that in terms of penalty LS is better than DN in the sense that receives less expected penalty. We again do not need to deal with continuation values since they are the same for both actions. Given our extra assumption relating spread and tick we will see that the immediate reward of DN is never higher than the immediate reward of LS, and so the claim follows. To see that observe that the immediate reward of playing LS is

$$p^a_{BH}\frac{\theta}{2} + (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)\phi(q - p^a_{BH})$$

and the immediate reward of playing DN is

$$(m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)\phi q$$

Subtracting both expressions we get

$$p^a_{BH}\frac{\theta}{2} + (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)\phi(-p^a_{BH}) \geq$$

$$p^a_{BH}\frac{\theta}{2} + \phi(-p^a_{BH}) = p^a_{BH}\frac{\theta}{2} + \frac{\theta}{k}(-p^a_{BH}) \geq 0$$

where the first inequality comes from

$$0 \leq (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1) \leq 1$$

and the last one holds since $k \geq 2$. Therefore the lemma follows.

Lemma 18 *If LLB is optimal at (BH, q) and one prefers to go to $(BH, q+1)$ instead of staying at (BH, q) when playing LLB, then $LLS \prec LLB$ at (SH, q).*

Proof As one prefers to end up at $(BH, q+1)$ instead of at (BH, q) and $p^b_{SH} > p^b_{BH}$, then $v^{LLB}_{SH,q} > v^{LLB}_{BH,q}$.
Suppose now $v^{LLS}_{SH,q} > v^{LLB}_{SH,q}$. Notice that when playing LLS at (ω, q), one always wants to sell (and hence go to $(SH, q-1)$ instead of staying at (SH, q)). The reason is that when playing LLS, one prefers ending up at $(SH, q-1)$ over (SH, q) both in terms of penalty (there is less penalty at $q-1$) and in terms of price change (since by playing LLS one reduces the value of the inventory due to the term $(2\beta(SH) - 1)$, which is negative). This term multiplies inventory level, so that the reduction is lower with less inventory. Also by ending at $(SH, q-1)$ we have gained an immediate reward of $\frac{\theta}{2}$. Consequently, as $p^a_{BH} > p^a_{SH}$, we have $v^{LLS}_{BH} > v^{LLS}_{SH}$. Therefore we have the following chain of inequalities:

$$v^{LLS}_{BH} > v^{LLS}_{SH} > v^{LLB}_{SH} > v^{LLB}_{BH} = v^*_{BH}$$

so that in particular one gets the contradiction $v^{LLS}_{BH} > v^*_{BH}$. The contradiction comes from having assumed $v^{LLS}_{SH,q} > v^{LLB}_{SH,q}$. Therefore the lemma holds.

Lemma 19 *Assume LLB is optimal at (BH, q) and one prefers to go to $(BH, q+1)$ instead of staying at (BH, q) when playing LLB. Then the cutoff α_1 such that $v^{LLB}_{BH}(\alpha_1) = v^{LLS}_{BH}(\alpha_1)$ and the cutoff α_2 such that $v^{LLB}_{SH}(\alpha_2) = v^{LLS}_{SH}(\alpha_2)$ cannot be equal.*

Proof Suppose they are equal. Then, we have a value α^* such that

$$v^{LLB}_{BH}(\alpha^*) = v^{LLS}_{BH}(\alpha^*)$$

and

$$v^{LLB}_{SH}(\alpha^*) = v^{LLS}_{SH}(\alpha^*)$$

As one prefers to end up at $(BH, q+1)$ instead of (BH, q) when playing LLB, then $v_{SH}^{LLB}(\alpha^*) > v_{BH}^{LLB}(\alpha^*)$. Moreover, as seen in the previous Lemma, we always have $v_{BH}^{LLS}(\alpha^*) > v_{SH}^{LLS}(\alpha^*)$. Therefore, we have the following chain of inequalities:

$$v_{SH}^{LLB}(\alpha^*) > v_{BH}^{LLB}(\alpha^*) = v_{BH}^{LLS}(\alpha^*) > v_{SH}^{LLS}(\alpha^*)$$

which contradicts $v_{SH}^{LLB}(\alpha^*) = v_{SH}^{LLS}(\alpha^*)$. Therefore $\alpha_1 \neq \alpha_2$ and the lemma holds.

Lemma 20 *Assume LLB is optimal at (BH, q) and one prefers to stay at (BH, q) instead of moving to $(BH, q+1)$ when playing LLB. Then $LLS \prec LLB$ at (SH, q).*

Proof Observe that when $\alpha \to 0$, if one plays LLB, it is better to move to $(BH, q+1)$ instead of staying at (BH, q), due to the combination of having negligible penalty and having more inventory, so that the total increase in its value is higher. Therefore, if LLB is optimal at (BH, q) and one prefers to stay at (BH, q) instead of moving to $(BH, q+1)$ when playing LLB, then $\alpha \in [\alpha_1, \alpha_2]$, with $\alpha_1 > 0$. Take α in this interval $[\alpha_1, \alpha_2]$. Then we both have

$$v_{BH}^{*}(\alpha) = v_{BH}^{LLB}(\alpha) > v_{SH}^{LLB}(\alpha)$$

where the inequality holds since we prefer to stay at (BH, q) when playing LLB and $p_{BH}^{b} < p_{SH}^{b}$, and

$$v_{BH}^{*}(\alpha) = v_{BH}^{LLB}(\alpha) > v_{BH}^{LLS}(\alpha) > v_{SH}^{LLS}(\alpha).$$

where the first inequality is due to optimality and the second is always true as we saw in Lemma 1.5.
Suppose now that $v_{SH}^{LLS}(\alpha) > v_{SH}^{LLB}(\alpha)$. Then we have the following chain of inequalities:

$$v_{BH}^{LLB}(\alpha) > v_{BH}^{LLS}(\alpha) > v_{SH}^{LLS}(\alpha) > v_{SH}^{LLB}(\alpha).$$

But we know there exists α^* such that LLB is optimal at BH and one prefers to move to $(BH, q+1)$. Therefore $v_{SH}^{LLB}(\alpha^*) > v_{BH}^{LLB}(\alpha^*)$. Hence, due to continuity of the value functions with respect to α, there exists a value α^{**} such that $v_{BH}^{LLB}(\alpha^{**}) = v_{BH}^{LLS}(\alpha^{**}) = v_{SH}^{LLS}(\alpha^{**}) = v_{SH}^{LLB}(\alpha^{**})$, which is impossible since we know that $v_{BH}^{LLS}(\alpha^{**}) > v_{SH}^{LLS}(\alpha^{**})$, again using the same reasoning we did in Lemma 1.5.

Lemma 21 *Assume LLB is optimal at (BH, q) and one prefers to stay at (BH, q) instead of moving to $(BH, q+1)$ when playing LLB. Then the cutoff α_1 such that $v_{BH}^{LLB}(\alpha_1) = v_{BH}^{LLS}(\alpha_1)$*

and the cutoff α_2 *such that* $v_{SH}^{LLB}(\alpha_2) = v_{SH}^{LLS}(\alpha_2)$ *are not equal, with possibly the exception of a set of measure* 0.

Proof Note we cannot apply the reasoning we did in Lemma 1.6 since now as one prefers to end up at (BH, q) instead of $(BH, q+1)$ when playing LLB, then $v_{BH}^{LLB}(\alpha^*) > v_{SH}^{LLB}(\alpha^*)$. However, note that the value functions are polynomial functions on each of the p_ω^a's and p_ω^b's. Therefore for a fixed α and a fixed set of all but one of the p_ω^a's and p_ω^b's, the remaining one can have at most a countable number of values for which both

$$v_{BH}^{LLB}(\alpha^*) = v_{BH}^{LLS}(\alpha^*)$$

and

$$v_{SH}^{LLB}(\alpha^*) = v_{SH}^{LLS}(\alpha^*)$$

hold. Therefore the lemma holds.

Lemma 22 *The following statements hold:*
If $m_{BH|N} < m_{SH|N}$ *then* DN *is dominated by* LS *at regime* $\omega = N$ *for* $q > 0$.
If $m_{BH|N} > m_{SH|N}$ *then* DN *is dominated by* LB *at regime* $\omega = N$ *for* $q < 0$.
If $m_{BH|N} = m_{SH|N}$ *then* DN *is dominated by* LS *at regime* $\omega = N$ *for* $q > 0$ *and by* LB *at regime* $\omega = N$ *for* $q < 0$.
Hence, the following statements hold:
If $m_{BH|N} < m_{SH|N}$, *then we have manipulation starting at regime* $\omega = N$ *for* $q > 0$.
If $m_{BH|N} > m_{SH|N}$, *then we have manipulation starting at regime* $\omega = N$ *for* $q < 0$.
If $m_{BH|N} = m_{SH|N}$, *then we have manipulation starting at regime* $\omega = N$ *for both* $q > 0$ *and* $q < 0$.

Proof We just consider the case $m_{BH|N} < m_{SH|N}$ since the others follow a similar reasoning. Note that by playing LS at regime N one gets

$$p_N^a\,\vartheta/2 + (m_{BH|N} - m_{SH|N})(2\beta(BH) - 1)\,(\varphi\,q - p_N^a\,\varphi) - \alpha\,p_N^a\,(q-1)^2 - \alpha\,(1 - p_N^a)\,q^2$$

whereas by playing DN at regime N one gets

$$(m_{BH|N} - m_{SH|N})(2\,\beta(BH) - 1)\,(\varphi\,q) - \alpha(q)^2$$

Therefore, if $m_{BH|N} - m_{SH|N} \leq 0$ we have that LS dominates DN, and therefore we can proceed as we did for $\omega = SH$ at $q > 0$, so that we have manipulation starting at regime N for positive inventory.
The other cases follow an analogous reasoning.

Lemma 23 *Let $q > 0$. If $LLB \prec LB$ at (SH,q), then $LS \prec LLS$ at (BH,q).*

Proof We first show that the cutoff between LS and MS at BH regime increases with respect to their cutoff in the original scenario. The reason is that the value of playing MS in this price impact case (denoted MS_{new}) is the value of playing MS in the original scenario (denoted MS_{or}) plus adding an extra term, and a similar reasoning applies to the action LS. In particular, forgetting about the continuation values we have:

$$MS_{new} = MS_{or} + (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)(\varphi q - \varphi)$$

and

$$LS_{new} = LS_{or} + (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)(\varphi q - p^a_{BH}\varphi)$$

Note that the increase is higher for LS since

$$(m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)(\varphi q - p^a_{BH}\varphi) > (m_{BH|BH} - m_{SH|BH})(2\beta(BH) - 1)(\varphi q - \varphi).$$

Moreover, the increase in continuation values is also higher for LS since in the original case, due to monotonicity, we have $v_{\omega,q-1} > v_{\omega,q}$. However in the new case $v_{\omega,q}$ has a higher increase in value compared to $v_{\omega,q-1}$. Therefore, LS becomes relatively more attractive for any value of α compared to MS and therefore the cutoff increases.
Second, the cutoff between LB and LLB at SH decreases, which happens since LLB becomes relatively more attractive in the new scenario. Combining both facts and using that the cutoff between LLS and LS is lower than the cutoff between LS and MS we get the result.

Theorem 4 *Let $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$,* (C1)*,* (C2) *and* (C4) *hold, and let*

$$p^a_{BH} - p^b_{SH} < \min\left\{(p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{(p_{N|BH} - \frac{\kappa}{2})}{(p_{BH|BH} - \frac{\kappa}{2})}\,,\ (p^b_{SH} - \max\{p^a_N, p^b_N\})\frac{(p_{N|N} - \frac{\kappa}{2})}{(p_{BH|N} - \frac{\kappa}{2})}\right\}$$
$$p^b_{SH} - p^a_{BH} < \min\left\{(p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{p_{N|SH} - \frac{\kappa}{2}}{(p_{SH|SH} - \frac{\kappa}{2})}\,,\ (p^a_{BH} - \max\{p^a_N, p^b_N\})\frac{(p_{N|N} - \frac{\kappa}{2})}{(p_{SH|N} - \frac{\kappa}{2})}\right\} \quad \text{(C3')}$$

hold.

4.1 If $p^b_N - p^a_N > (p^a_{BH} - p^b_{SH})\dfrac{\delta(\frac{\kappa}{2}+\kappa-1)}{1+\delta(1-\kappa-\frac{\kappa}{2})}$ *holds, then* $I_1(\mathbf{s}) \neq \emptyset$ *and* $I_2(\mathbf{s}) \neq \emptyset$ *for all states* $\mathbf{s} = (N, q > 0)$.

4.2 If $p^a_N - p^b_N > (p^b_{BH} - p^a_{SH})\dfrac{\delta(\frac{\kappa}{2}+\kappa-1)}{1+\delta(1-\kappa-\frac{\kappa}{2})}$ *holds, then* $I_1(\mathbf{s}) \neq \emptyset$ *and* $I_2(\mathbf{s}) \neq \emptyset$ *for all states* $\mathbf{s} = (N, q < 0)$.

Proof First we establish the following claim:

Claim 6 *If* $p^a_{SH} < p^a_N < p^a_{BH}$, $p^b_{SH} > p^b_N > p^b_{BH}$, *and* (C3') *hold, then for a small enough value of* α, $LS \prec LLS$ *for* $\omega = BH$, $LB \prec LLB$ *for* $\omega = SH$, *and* $LB \prec LLB$ *and* $LS \prec LLS$ *for* $\omega = N$.

For non-deterministic transition probabilities we have that at $\omega = BH$, LLS is preferred to LS if and only if

$$(1-\kappa)v^*_{SH} + \frac{\kappa}{2}v^*_N + \frac{\kappa}{2}v^*_{BH} > p_{BH|BH}\, v^*_{BH} + p_{N|BH}\, v^*_N + p_{SH|BH}\, v^*_{SH} \iff$$

$$(1-\kappa)v^*_{SH} > (p_{BH|BH} - \frac{\kappa}{2})\, v^*_{BH} + (p_{N|BH} - \frac{\kappa}{2})\, v^*_N + p_{SH|BH}\, v^*_{SH} \iff$$

$$(p_{BH|BH}-\frac{\kappa}{2})v^*_{SH}+(p_{N|BH}-\frac{\kappa}{2})\, v^*_{SH}+p_{SH|BH}\, v^*_{SH} > (p_{BH|BH}-\frac{\kappa}{2})\, v^*_{BH}+(p_{N|BH}-\frac{\kappa}{2})\, v^*_N+p_{SH|BH}\, v^*_{SH}\,.$$

This inequality trivially holds if $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{SH}$. On the other hand, if $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then the last inequality holds if and only if

$$(p_{N|BH} - \frac{\kappa}{2})\,(v^*_{SH} - v^*_N) > (p_{BH|BH} - \frac{\kappa}{2})\,(v^*_{BH} - v^*_{SH}) \iff v^*_{BH} - v^*_{SH} < (v^*_{SH} - v^*_N)\,\frac{(p_{N|BH} - \frac{\kappa}{2})}{(p_{BH|BH} - \frac{\kappa}{2})}\,.$$

If $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then $v^*_{BH} = p^a_{BH}\,\vartheta/2 + \delta\,((1-\kappa)v^*_{SH} + \frac{\kappa}{2}v^*_N + \frac{\kappa}{2}v^*_{BH}) < p^a_{BH}\,\vartheta/2 + \delta\,((1-\kappa)v^*_{BH} + \frac{\kappa}{2}v^*_N + \frac{\kappa}{2}v^*_{SH})$ so that

$$v^*_{BH} - v^*_{SH} < (p^a_{BH} - p^b_{SH})\,\vartheta/2\,.$$

Now, the optimal action in $\omega = SH$ is LLB because $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$. Hence, since $v^*_N < \max\{p^a_N, p^b_N\}\vartheta/2 + \delta((1-\kappa)v^*_{BH} + \frac{\kappa}{2}v^*_N + \frac{\kappa}{2}v^*_{SH})$

$$v^*_{SH} - v^*_N = v_{SH}(LLB) - v^*_N > (p^b_{SH} - \max\{p^a_N, p^b_N\})\,\frac{\vartheta}{2}\,.$$

Therefore, if

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\}) \frac{(p_{N|BH} - \frac{\kappa}{2})}{(p_{BH|BH} - \frac{\kappa}{2})},$$

then LLS is preferred to LS because

$$(v^*_{SH} - v^*_N) \frac{(p_{N|BH} - \frac{\kappa}{2})}{(p_{BH|BH} - \frac{\kappa}{2})} > (p^b_{SH} - \max\{p^a_N, p^b_N\}) \frac{\vartheta}{2} \frac{(p_{N|BH} - \frac{\kappa}{2})}{(p_{BH|BH} - \frac{\kappa}{2})} > (p^a_{BH} - p^b_{SH}) \frac{\vartheta}{2} > v^*_{BH} - v^*_{SH}.$$

For $\omega = SH$, we follow a similar reasoning so that if

$$p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\}) \frac{(p_{N|SH} - \frac{\kappa}{2})}{(p_{SH|SH} - \frac{\kappa}{2})},$$

then LLB is preferred to LB in $\omega = SH$.

For $\omega = N$, we first compare LLS with LS. As before, LLS is preferred to LS if and only if

$$(1-\kappa) v^*_{SH} + \frac{\kappa}{2} v^*_N + \frac{\kappa}{2} v^*_{BH} > p_{BH|N}\, v^*_{BH} + p_{N|N}\, v^*_N + p_{SH|N}\, v^*_{SH}.$$

If $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{SH}$, then the former inequality holds. On the other hand, if $\max\{v^*_{SH}, v^*_N, v^*_{BH}\} = v^*_{BH}$, then the last inequality holds if and only if

$$v^*_{BH} - v^*_{SH} < \frac{(p_{N|N} - \frac{\kappa}{2})}{(p_{BH|N} - \frac{\kappa}{2})} (v^*_{SH} - v^*_N).$$

The remainder follows the same reasoning as that in the case with $\omega = BH$ so that if

$$p^a_{BH} - p^b_{SH} < (p^b_{SH} - \max\{p^a_N, p^b_N\}) \frac{(p_{N|N} - \frac{\kappa}{2})}{(p_{BH|N} - \frac{\kappa}{2})},$$

then LLS is preferred to LS in $\omega = N$.

Finally, using a similar reasoning, we have that if

$$p^b_{SH} - p^a_{BH} < (p^a_{BH} - \max\{p^a_N, p^b_N\}) \frac{(p_{N|N} - \frac{\kappa}{2})}{(p_{SH|N} - \frac{\kappa}{2})},$$

then LLB is preferred to LB in $\omega = N$.

For $\omega = N$, if the value of α is sufficiently small, then we have the following action values

$$v_N(LLB) = p_N^b \frac{\vartheta}{2} + \delta\left((1-\kappa)v_{BH}^* + \frac{\kappa}{2}v_N^* + \frac{\kappa}{2}v_{SH}^*\right),$$
$$v_N(LLS) = p_N^a \frac{\vartheta}{2} + \delta\left((1-\kappa)v_{SH}^* + \frac{\kappa}{2}v_N^* + \frac{\kappa}{2}v_{BH}^*\right),$$

Therefore, again, for a sufficiently small value of α, the optimal action is either LLS or LS in $\omega = BH$, whereas the optimal action is either LLB or LB at $\omega = SH$. From Claim 6, we have $LS \prec LLS$ for $\omega = BH$ and $LB \prec LLB$ for $\omega = SH$. Consequently,

$$v_N(LLB) = p_N^b \frac{\vartheta}{2} + \delta(1-\kappa)v_{BH}^* + \delta\frac{\kappa}{2}v_N^* + \delta\frac{\kappa}{2}v_{SH}^*,$$

and similarly

$$v_N(LLS) = p_N^a \frac{\vartheta}{2} + \delta(1-\kappa)v_{SH}^* + \delta\frac{\kappa}{2}v_N^* + \delta\frac{\kappa}{2}v_{BH}^*, .$$

so that

$$v_N(LLB) \geq v_N(LLS) \iff p_N^b \frac{\vartheta}{2} + \delta(1-\kappa)v_{BH}^* + \delta\frac{\kappa}{2}v_{SH}^* \geq p_N^a \frac{\vartheta}{2} + \delta(1-\kappa)v_{SH}^* + \delta\frac{\kappa}{2}v_{BH}^*, .$$

Noting that

$$v_{BH}^* = p_{BH}^a \frac{\vartheta}{2} + \delta((1-\kappa)v_{SH}^* + \frac{\kappa}{2}v_N^* + \frac{\kappa}{2}v_{BH}^*$$

and

$$v_{SH}^* = p_{SH}^b \frac{\vartheta}{2} + \delta((1-\kappa)v_{BH}^* + \frac{\kappa}{2}v_N^* + \frac{\kappa}{2}v_{SH}^*$$

we get that

$$v_{BH}^* - v_{SH}^* = (p_{BH}^a - p_{SH}^b)\frac{\vartheta}{2} + \delta((1-\kappa)(v_{SH}^* - v_{BH}^*) + \delta\frac{\kappa}{2}(v_{BH}^* - v_{SH}^*)$$

so that

$$v_{BH}^* - v_{SH}^* = \frac{(p_{BH}^a - p_{SH}^b)\frac{\vartheta}{2}}{1 + \delta(1 - \kappa - \frac{\kappa}{2})}$$

Hence, $v_N(LLB)$ is preferred to $v_N(LLS)$ if and only if

$$p_N^b \frac{\vartheta}{2} + \delta(1-\kappa)(\upsilon_{SH}^* + \frac{(p_{BH}^a - p_{SH}^b)\frac{\vartheta}{2}}{1+\delta(1-\kappa-\frac{\kappa}{2})}) + \delta\frac{\kappa}{2}\upsilon_{SH}^* \geq$$

$$p_N^a \frac{\vartheta}{2} + \delta(1-\kappa)\upsilon_{SH}^* + \delta\frac{\kappa}{2}(\upsilon_{SH}^* + \frac{(p_{BH}^a - p_{SH}^b)\frac{\vartheta}{2}}{1+\delta(1-\kappa-\frac{\kappa}{2})})$$

which happens if and only if

$$p_N^b - p_N^a > (p_{BH}^a - p_{SH}^b)\frac{\delta(\frac{\kappa}{2}+\kappa-1)}{1+\delta(1-\kappa-\frac{\kappa}{2})}$$

Similarly, $v_N(LLS)$ is preferred to $v_N(LLB)$ if and only if

$$p_N^a - p_N^b > (p_{BH}^b - p_{SH}^a)\frac{\delta(\frac{\kappa}{2}+\kappa-1)}{1+\delta(1-\kappa-\frac{\kappa}{2})} + \partial$$

□