12-20-2023

Market Making in Limit Order Books with Latency and Running Inventory Control

Chang Liu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Statistics and Data Science

Dissertation Examination Committee:
José E. Figueroa-López, Chair
Jonathan Chávez-Casillas
Jimin Ding
Renato Feres
Mladen Victor Wickerhauser

Market Making in Limit Order Books with Latency and Running Inventory Control

by

Chang Liu

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

Dec 2023
St. Louis, Missouri

Table of Contents

List of Figures . . . . . **iv**

List of Tables . . . . . **vi**

Acknowledgments . . . . . **viii**

Abstract . . . . . **x**

Chapter 1: Introduction **1**
1.1 Limit Order Book and Market Making 1
1.1.1 Limit Order Book 1
1.1.2 Market Making 5
1.2 Latency in Algorithmic Trading 6
1.3 Reinforcement Learning in Market Making 7
1.3.1 Overview of Reinforcement Learning 7
1.3.2 Reinforcement Learning on Market Marking 21

Chapter 2: Market Making with Latency **24**
2.1 Model Setup and Assumptions 24
2.1.1 Market Making with Latency 24
2.1.2 State and Action Spaces and System Dynamics 26
2.2 Characterization of the Optimal Market Making Strategy 34
2.2.1 Value of the Order 36
2.2.2 Value Functions 40
2.3 Variations in the Market Making Model: Special Cases Analysis 42
2.3.1 Simplified Model with Latency 43
2.3.2 Market Making without Latency 46
2.3.3 Simplified Market Making without Latency 52
2.4 Implementation and Empirical Performance 53
2.4.1 Parameter Estimation 54
2.4.2 Inventory and Optimal Spread Process 59
2.4.3 Results 62
2.4.4 Computational Challenge 70

Chapter 3: Reinforcement Learning in Market Making with Latency . 72
3.1 Simulation Study 72
3.1.1 RL Settings 73
3.1.2 Simulation Results 78

Chapter 4: Market Making with Running Inventory Penalty 80
4.1 Model Setup 80
4.2 Analytical Optimal Control 86
4.2.1 Optimal Strategy Under a Martingale Fundamental Price Process 87
4.2.2 Optimal Strategy with a General Adapted Fundamental Price Process 92
4.3 Properties of the Optimal Placement Strategies 96
4.3.1 Sensitivity to Simultaneous Arrival Probability 96
4.3.2 Sensitivity to Inventory Level 99
4.3.3 Sensitivity to Terminal Inventory Penalty 105

Chapter 5: Conclusion and Future Work 112

References 114

Appendix A Proofs and Simulation Configuration of Chapter 2 . . . 118
A.1 Proof of Proposition 2.2.1 118
A.2 Proof of Theorem 2.2.2 121
A.3 Proof of Proposition 2.3.1 125
A.4 Proof of Theorem 2.3.2 128
A.5 Parameters under 0.20 seconds Latency 131

Appendix B Proofs of Chapter 4 132
B.1 Proof of Proposition 4.2.1 and Theorem 4.2.4 132
B.2 Proof of Lemma 4.2.2 140
B.3 Proof of Proposition 4.2.5 146
B.4 Proof of Corollary 4.3.1 148
B.5 Proof of Corollary 4.3.2 150
B.6 Proof of Corollary 4.3.3 151

List of Figures

Figure 1.1: Matching mechanism in LOB when a new LO is placed. 3

Figure 1.2: Matching mechanism in LOB when a sell MO is submitted. 4

Figure 1.3: The agent–environment interaction in a discrete time MDP. 10

Figure 2.1: Average demand vs. estimated linear demand during a 1-second trading interval on January 2, 2019, for AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second. 58

Figure 4.1: Paths of α_{t_k} for different values of running inventory penalty ϕ under symmetric market conditions. The agent acts every one second from 0 to 19800 seconds. Plots are generated with the following parameters: $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1\times10^4$, $\mu_{c^2p}^{\pm} = 5\times10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$ and $\pi_{t_k}(1,1) \equiv 0.1$. 90

Figure 4.2: Optimal bid-ask spread for different values of running inventory penalty ϕ under symmetric market conditions. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, we set $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1)$ ranges from 0 to $\pi^{\pm}$. 100

Figure 4.3: Optimal spreads in the last 500-seconds with various inventory levels for different values of ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, we set $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} = 0.2$, $\pi(1,1) = 0$ and $\Delta_{t_k} = 0$ for $k = 0, \ldots, N$. 103

Figure 4.4: Optimal trading strategies in the final 100 seconds: impact of inventory, terminal inventory penalty λ, and running inventory penalty ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1) = 0$ and $\Delta t_k = 0$ for $k = 0, \ldots, N$. 107

Figure 4.5: Optimal trading strategies in the final 100 seconds when $\pi_{t_k}(1,1) > 0$ for different values of ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. λ ranges from 0 to 1. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1) = 0.05$ and $\Delta t_k = 0$ for $k = 0, \ldots, N$. 111

List of Tables

Table 2.1: Default values of the parameters in simulation in Section 2.4. 54

Table 2.2: Average rate (per second) of price changes over 252 trading days in 2019. 55

Table 2.3: Average arrival probability of market orders over 252 trading days in 2019. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second. 55

Table 2.4: Average values of $\hat{\mu}^{\pm}_{\{c,p\}}$ over 252 trading days for 2019 AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second. 57

Table 2.5: Parameter values for c and p in simulation for 2019 AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second. 60

Table 2.6: Average arrival probability of market orders during half-hour intervals at different times of day averaged over 252 trading days in 2019 for AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second. 63

Table 2.7: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$. 64

Table 2.8: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.20$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$. 66

Table 2.9: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.1$. 67

Table 2.10: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 1$. 67

Table 2.11: Terminal P&L for AAPL stock in 2019 across different trading times when placing new LOs only at all time intervals. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.1$. 68

Table 2.12: Terminal P&L for AAPL Stock in 2019 across different trading times when LOs can only be filled at full quantity. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$. 69

Table 3.1: Default values of the parameters in simulation in Section 3.1. 73

Table 3.2: Default parameters for RL algorithms in Section 3.1. 78

Table 3.3: Mean and standard deviation of objective values $W_T + S_T I_T - \lambda I_T^2$ (given in units of 10^4), and average terminal inventory (given in unit of 1 share). $\lambda = 0.0005$. 79

Table A.1: Average arrival probability of market orders over 252 trading days in 2019. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second. 131

Table A.2: Average values of $\hat{\mu}^{\pm}_{\{c,p\}}$ over 252 trading days for 2019 AAPL. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second. 131

Table A.3: Parameter values for c and p in simulation for 2019 AAPL. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second. 131

Acknowledgments

I would like to express my sincere gratitude and appreciation to my advisor, José E. Figueroa-López, for his exceptional mentorship, guidance, and unwavering support throughout the entirety of this research project. His profound knowledge in the field of statistics, coupled with his commitment to academic excellence, has significantly shaped the direction and quality of this thesis. His insightful feedback, attention to detail, and rigorous approach to research have been invaluable in refining my analytical skills and expanding my understanding of statistical methodologies. I am truly grateful for his dedication, patience, and belief in my capabilities, which have continuously motivated and inspired me to strive for excellence.

Furthermore, I would like to extend my heartfelt thanks to Jonathan Chávez Casillas from University of Rhode Island for his collaboration and contribution to this thesis. His expertise and willingness to share his knowledge have greatly enriched the research process, leading to valuable insights and innovative approaches.

In addition, I want to express my deepest gratitude to my friends and peers, Jiayi Fu, Cezareo Rodriguez and Chuyi Yu, for their unwavering support, intellectual discussions, and encouragement throughout this journey. Their diverse perspectives and constructive feedback have played a crucial role in shaping my ideas and refining my arguments.

I would like to acknowledge and extend my appreciation to my loving wife, Yiying Cheng, for her unwavering belief in me, endless encouragement, and understanding during the highs and lows of this research endeavor. Her presence, love, and support have provided

me with the necessary strength and motivation to overcome challenges and pursue my academic goals.

Last but certainly not least, I want to express my profound gratitude to my parents, Zhisheng Liu and Yanxia Wang, for their unconditional love, unwavering support, and lifelong encouragement. Their sacrifices, belief in my abilities, and unwavering confidence in me have been a constant source of inspiration and motivation. I am eternally grateful for their relentless support, guidance, and the values they instilled in me, which have shaped me into the person I am today.

Chang Liu

Washington University in St. Louis

Dec 2023

ABSTRACT OF THE DISSERTATION

Market Making in Limit Order Books with Latency and Running Inventory Control

by

Chang Liu

Doctor of Philosophy in Statistics

Washington University in St. Louis, 2023

Professor José E. Figueroa-López, Chair

In this thesis, we delve into the intricate optimization challenges of market making, the concurrent provision of buy and sell prices in financial assets. The focus is particularly on the complexities inherent in high-frequency trading scenarios, addressing optimal market making in the presence of latency and incorporating a running inventory penalty.

The initial exploration involves the formulation of a stochastic control model that aptly captures the actions of an electronic market maker navigating a trading environment influenced by latency. The main objective of the market maker lies in the maximization of expected terminal wealth. To systematically address and resolve this control problem, we recast it into a finite-horizon Markov Decision Process, subsequently amenable to numerical solutions.

A complementary avenue is explored by employing model-free Reinforcement Learning algorithms to tackle the intricacies of market making in the presence of latency. This approach signifies a departure from traditional model-centric methods, harnessing the power of RL to adapt and optimize market-making strategies.

We finally propose an extension of the approach introduced in [12]. This extension introduces a penalty mechanism linked to the running inventory across the entire trading horizon. The incorporation of this running inventory penalty framework significantly

enhances the market maker's risk management capabilities. By doing so, it establishes a more resilient and effective framework for the evaluation of diverse trading strategies. This augmentation is crucial for mitigating the market maker's exposure to inventory risk, thereby contributing to the robustness of the overall market-making process.

Chapter 1

Introduction

1.1 Limit Order Book and Market Making

1.1.1 Limit Order Book

The financial markets are complex and dynamic ecosystems where various participants engage in buying and selling securities. One crucial aspect of these markets is the trading mechanism, which determines how orders are executed and securities are priced. Over the years, technological advancements and the evolution of electronic trading platforms have revolutionized the way securities are traded, leading to the emergence of new trading strategies and instruments.

One such innovation that has gained significant importance in modern financial markets is the Limit Order Book (LOB). The LOB is a critical component of electronic trading platforms and serves as a transparent marketplace where market participants can submit and execute their buy and sell orders. It provides a comprehensive view of the supply and demand dynamics for a particular security at any given point in time.

The LOB consists of two main types of orders: limit orders (LOs) and market orders (MOs). A LO is an order to buy or sell a security at a specified price or better. It

remains in the LOB until it is executed, canceled, or expires. On the other hand, a MO is an order to buy or sell a security at the prevailing market price, executing immediately against the best available prices in the LOB.

More specifically, LOs contain critical information such as a specified price, associated volume, and a directional cue indicating either a buy or sell stance on the underlying asset. Once an LO is entered, it takes residence in the LOB, where it abides until encountering one of two outcomes: cancellation or fulfillment against a subsequent MO. The cancellation or execution of an LO may be comprehensive, involving the removal of the entire order volume, or partial, allowing a residual portion to linger within the LOB.

In instances where a subsequent LO is introduced at an identical price point, concurrent existence within the LOB ensues, governed by the matching rules intrinsic to the exchange. The order of execution adheres to these rules, which we expound upon hereafter, embodying a price-time priority schema. This schema engenders a queue formation for LOs at each price level, with execution sequence determined by their position in this queue—a precept colloquially recognized as the First-In-First-Out (FIFO) rule.

In the coexistence of diverse LO types (buy and sell) within the LOB, two distinctive price levels emerge. The highest price among all buy LOs designates the *best bid*, while the lowest price among all sell LOs designates the *best ask*. The structure of the LOB is hierarchical, with buy orders sorted in descending order of price and sell orders sorted in ascending order of price. This allows market participants to easily observe the best bid and best ask in the LOB. Collectively, these levels delineate the bid-ask spread, defined as the difference between the best bid and ask prices, represents the transaction cost for market participants and serves as an important measure of liquidity. The arithmetic mean of the best bid and best ask defines the midprice of the asset. The proposed graphical

representation in Figure 1.1 [1] illuminates the state of the LOB at a specific temporal juncture, followed by its transformation subsequent to the introduction of a singular LO.

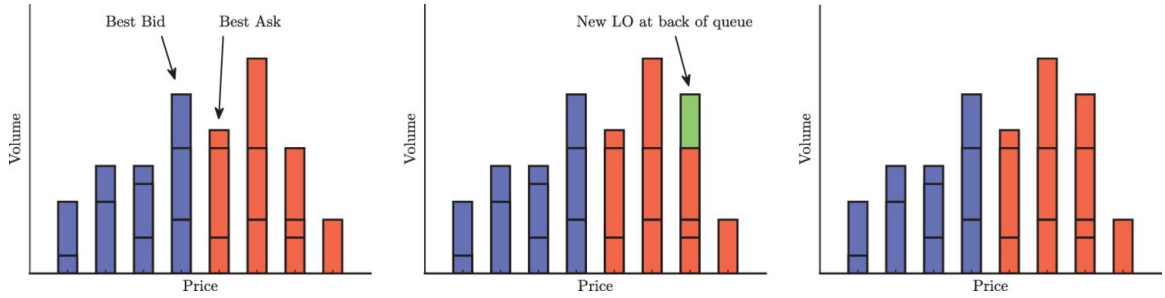


Figure 1.1: Matching mechanism in LOB when a new LO is placed.

In Figure 1.1, the graphical representation delineates critical aspects of the LOB. Blue bars symbolize buy LOs, and red bars represent sell LOs. In the left panel, the highest price among all buy LOs and the lowest price among all sell LOs are denoted as the best bid and best ask, respectively. This snapshot captures the prevailing market conditions before any subsequent changes. Moving to the middle panel, a pivotal event unfolds as a new LO is introduced at a price where orders already exist, positioning it at the back of the queue. This scenario mirrors the dynamics of the LOB when new orders are placed amid existing ones. The right panel extends this narrative by showcasing an adjustment in volume at this specific price, highlighting the malleability of the LOB. It's noteworthy that analogous modifications could arise from order cancellations, underscoring the dynamic nature of the LOB over time, even in the absence of actual trades.

A MO also contains a specified volume and an indicator signaling either the intent to buy or sell the asset. Upon submission, an MO undergoes a matching process against resting LOs of the opposing buy/sell type within the LOB. The price at which the transaction occurs is contingent upon the cumulative price levels of all filled LOs, extending up to the volume specified in the MO.

[1] Figure 1.1 and 1.2 are adapted from [23].

Figure 1.2 illustrates the sequential filling of an MO against several LOs, providing a depiction of how the MO's volume, along with the volumes and prices of the individual LOs, collectively determine the overarching transaction price of the trade. The left panel provides a snapshot of the LOB immediately preceding the submission of a MO. In the middle panel, a sell MO is introduced with a volume surpassing that of the best bid, necessitating the MO to traverse through the order book, a phenomenon commonly referred to as "walking the book", until it finds the next best available price. The right panel illustrates the aftermath of this MO execution, showing the depletion of volume at the relevant prices due to the filled LOs.

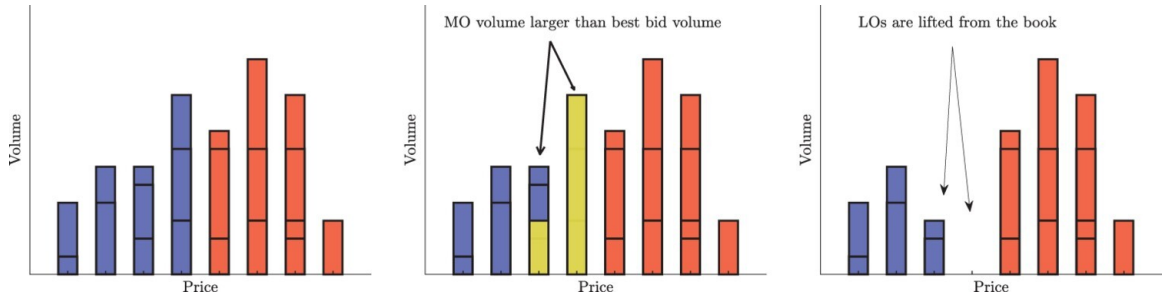


Figure 1.2: Matching mechanism in LOB when a sell MO is submitted.

For a comprehensive exploration of LOB dynamics, readers are encouraged to delve into the following: [3, 5, 9, 13, 15, 21, 22, 35, 42, 46]. These studies contribute to a nuanced understanding of LOB behaviors, spanning empirical analyses, order flow dynamics, and statistical approaches. For a more detailed examination of the operational intricacies of LOBs within financial markets, including advanced order types, additional insights can be gleaned from the works of [15, 27].

1.1.2 Market Making

Market making has undergone significant development over the years, evolving into a crucial aspect of financial markets. Market making, at its core, involves providing liquidity to financial markets by continuously quoting bid and ask prices for securities. It serves to bridge the gap between buyers and sellers, facilitating smooth transactions and enhancing market efficiency. The practice of market making has witnessed substantial progress and refinement through both theoretical research and practical implementation. Its foundations can be traced back to the early work of [31]. Their influential paper introduced a theoretical framework for analyzing market making activities and their impact on the bid-ask spread, while also showing that the bid-ask spreads from the market makers were influenced by factors such as inventory costs, adverse selection risk, and competition. In subsequent studies, for example, in [26], the dynamics of bid-ask spreads generated by the market makers were further explored as well as the impact of asymmetric information on market making.

As technology advanced, electronic trading platforms emerged, revolutionizing market making practices. The development of automated trading systems and algorithmic trading algorithms played a pivotal role in enhancing market liquidity and execution efficiency. The rise of high-frequency trading (HFT) brought further advancements in market making. Following the theory developed by [31], [4] proposed a novel model for market making in an LOB where an agent simultaneously places buy and sell LOs and attempts to maximize profits. They introduced the idea of optimal inventory control models, aiming to maximize profitability while mitigating risk. Their research contributed to the understanding of how market makers could optimize their positions and adapt to changing market conditions. [30] and [10] explored the impact of HFT on market liquidity and

price discovery, finding that HFT contributes to more efficient price formation and improved liquidity provision. In addition, the impact of MOs on the LOB and the associated adverse selection risk have been examined by [16]. Risk measures for HFT were investigated by [14]. Another relevant study by [29] explored a market making model that incorporated both LOs and MOs. For a comprehensive exploration of algorithmic and high-frequency trading, readers can refer to Cartea's book on the subject (cf. [15]).

1.2 Latency in Algorithmic Trading

The evolution of electronic trading platforms also brought new challenges and opportunities for market makers. One of the critical factors that emerged as a significant concern is *latency*, which, broadly, refers to "the time delay between an exchange sending market data to a trader, the trader processing information and deciding to trade, and the exchange receiving the order from the trader" (see [18]). Latency is inherent to electronic trading systems and can arise at various stages of the trading process, including data transmission, order routing, and order execution. Indeed, the role of latency in algorithmic trading and HFT has gained significant attention in the literature.

The impact of latency on trading strategies has been extensively explored in the context of optimal execution in the literature. [39] focused on quantifying the cost of latency in the context of HFT. They introduced a theoretical model that aimed to measure the impact of latency on the trading performance. By analyzing the optimal execution problem, they assessed the frictions arising from latency and its implications for trading strategies. [47] examined the role of low-latency trading algorithms in their study on reducing transaction costs. [17] addressed the optimal trade-off between missing trades and the costs associated with walking the order book in the presence of latency in the marketplace. [18] focused

specifically on liquidity taking orders in the foreign exchange market by examining how traders can adjust the limit price of marketable order to target a fill ratio. [40] and [11] explored impulse control problems in the case of deterministic delay, while [19] examined an impulse control problem involving stochastic latency, where the trader has control over the timing and price limits of marketable LOs submitted to the exchange.

While optimal execution and market making are closely related (see [27]), latency in market making has only recently received attention in the academic literature. [24] investigated the optimal strategy for a market maker providing liquidity to the LOB of large-tick assets. They considered a scenario where latency is fixed and deterministic throughout the trading period.

1.3 Reinforcement Learning in Market Making

1.3.1 Overview of Reinforcement Learning

Reinforcement learning (RL), a foundational concept within the domain of machine learning, draws inspiration from the human learning process. The RL paradigm centers on a trial-and-error framework where an agent interacts with an environment, discerns the ramifications of its actions, and adapts its behaviors based on received rewards. This learning approach, akin to human learning by exploration and feedback, aims to determine optimal actions that yield maximal cumulative rewards.

RL models, a powerful tool in machine learning, have been instrumental in various applications across diverse domains. Originating from the intersection of computational learning models and behavioral psychology, RL has exhibited significant potential for solving

complex problems, ranging from optimizing decision-making processes in dynamic environments to mastering strategic game play in artificial intelligence and robotics. RL's core principle involves learning a mapping from states to actions that maximize expected rewards, thereby optimizing decision-making strategies within a given context [48].

The fusion of RL with deep learning (DL) techniques has propelled significant advancements in recent years. This integration has facilitated the creation of more sophisticated and efficient learning models. While initially utilized in areas such as credit assignment and control problems [48], RL has since transcended these boundaries. For instance, RL has been pivotal in the domain of robotics, enabling the training of robotic agents to perform intricate tasks and navigate real world environments [33]. In recommendation systems, RL algorithms have significantly enhanced personalization and user engagement by optimizing content delivery based on user interactions [2]. Moreover, RL techniques have been pivotal in advancing natural language processing, fostering the development of conversational AI agents with enhanced dialogue generation and understanding capabilities [50]. Within the healthcare domain, RL models have contributed to optimizing treatment plans, clinical decision-making, and healthcare resource management [53]. Its adaptive nature and capacity to learn complex behaviors through interaction with environments make it a compelling tool for addressing intricate problems across different sectors.

Continuing our journey into the RL, we pivot to the core framework known as Markov Decision Process, which is instrumental in unraveling the complex interplay between an agent and its environment.

Markov Decision Process

Markov Decision Processes (MDPs) offer a fundamental framework for learning and decision-making within dynamic environments, encompassing the interactions between an agent and its surrounding environment [41]. Here, an *agent*, as the decision-maker, interacts with an external *environment*, engaging in a continual cycle of actions, responses, and the consequent acquisition of *rewards*. The dynamics entail the agent's actions triggering reactions from the environment, presenting new scenarios and influencing future states [7]. The goal of the agent within this structure is to make decisions that maximize the cumulative reward obtained over time, encapsulating the essence of learning from the interaction with the environment [48]. This section builds upon these established works to explore the mechanisms and implications of MDPs in the realm of decision-making and learning algorithms.

Precisely, according to [48], a MDP is defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where

- $\mathcal{S}$ denotes the set of states;
- $\mathcal{A}$ represents the set of actions;
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto [0, 1]$ characterizes the state transition probabilities;
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ indicates the reward function;
- $\gamma \in [0, 1]$ stands for the discount factor.

The agent and environment interact at discrete time steps, denoted as $t = 0, 1, 2, \ldots$. As shown in Figure 1.3, at each time step t, the agent observes the current state of the environment $S_t \in \mathcal{S}$ and, using this information, chooses an action $A_t \in \mathcal{A}$. Subsequently,

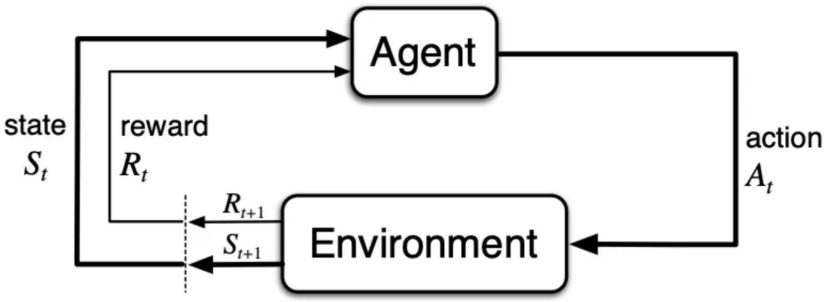


Figure 1.3: The agent–environment interaction in a discrete time MDP.

one time step later, the agent obtains a reward $R_{t+1} = R(S_t, A_t) \in \mathcal{R}$ contingent upon the action A_t and find itself in a subsequent state S_{t+1}. This iterative process generates a sequence $(S_0, A_0, R_1, \ldots, S_t, A_t, R_{t+1}, \ldots)$, which may extend infinitely. The environment encompasses all variables beyond the agent's direct influence. It's noteworthy that the action A_t typically does not uniquely determine the next state S_{t+1}, as the transition matrix can exhibit stochastic characteristics. Notably, the next state S_{t+1} depends solely on the current state S_t and action A_t, rendering it conditionally independent of past states and actions. This intrinsic property of the state transitions in a MDP aligns with the Markov property, providing a rationale for its nomenclature.

In a finite MDP, the sets of states, actions, and rewards ($\mathcal{S}$, $\mathcal{A}$ and $\mathcal{R}$) all consist of a finite number of elements. In this scenario, the random variables R_t and S_t exhibit well-defined discrete probability distributions that depend solely on the preceding state and action. Specifically, for specific values of these random variables, $s' \in \mathcal{S}$ and $r \in \mathcal{R}$, there exists a probability of those values occurring at time t, given specific values of the preceding state and action:

$$p(s', r \mid s, a) := \mathbb{P}(S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a),$$

for all $s', s \in \mathcal{S}$, $r \in \mathcal{R}$ and $a \in \mathcal{A}$. The function $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$ defines the *dynamics* of the MDP, and specifies a probability distribution for each s and a, that is,

$$\sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s', r \mid s, a) = 1, \text{ for all } s \in \mathcal{S}, a \in \mathcal{A}.$$

Let the sequence of rewards received after time step t be denoted by $R_{t+1}, R_{t+2}, R_{t+3}, \ldots$. The agent's objective is to maximize the expected return. The *return*, donated G_t, is formally expressed as follows

$$G_t := R_{t+1} + \gamma R_{t+2} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \tag{1.1}$$

where γ is a parameter, $0 \leq \gamma \leq 1$, called the *discount rate*, which determines the present value of future rewards. G_t in Eq. (1.1) is also referred to as the *discounted return*.

A *policy* of a MDP serves as the decision-making strategy of an agent. It's a mapping that guides the agent's choices based on the observed state of the environment. Policies can be stochastic $\pi : \mathcal{A} \times \mathcal{S} \mapsto [0, 1]$, if the agent is following policy π at time t, then $\pi(a \mid s)$ is the probability that $A_t = a$ if $S_t = s$; or deterministic $\pi : \mathcal{S} \mapsto \mathcal{A}$, which directly specifies a particular action for each state.

For a given policy π and a state s, the *state value function* serves as a pivotal metric, denoting the expected return when starting from state s and adhering to policy π. It is expressed as follows:

$$v_\pi(s) := \mathbb{E}_\pi[G_t \mid S_t = s],$$

where $s \in \mathcal{S}$.

Similarly, the *action value function* for a specific policy π, extends the evaluation to encompass the expected return starting from state s upon executing action a and subsequently following policy π. Its formulation is:

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t \mid S_t = s, A_t = a],$$

where $s \in \mathcal{S}, a \in \mathcal{A}$.

The objective of a MDP is to find a policy that maximizes the expected return:

$$\max_\pi \mathbb{E}_\pi[G_t].$$

The *optimal policy*, donated by π^*, is a policy that maximizes the expected return, i.e.,

$$\pi^*(s) = \arg\max_\pi \mathbb{E}_\pi[G_t].$$

Note that while an optimal policy always exists, it may not be unique.

An essential concept in MDP is the comparison of policies. A policy π is deemed at least as effective as another policy π', if and only if, for all states $s \in \mathcal{S}$,

$$v_\pi(s) \geq v_{\pi'}(s).$$

All optimal policies converge to a shared *optimal state-value function*, denoted as

$$v_*(s) = \max_\pi v_\pi(s),$$

for all $s \in \mathcal{S}$.

Optimal policies also share the same *optimal action-value function*, and is denoted as

$$q_*(s, a) := \max_\pi q_\pi(s, a),$$

for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$. Furthermore, an optimal policy, can be determined by selecting the action that maximizes the optimal action value function:

$$\pi^*(s) = \arg\max_a q_*(s, a).$$

It is evident that the value function serves as a crucial tool figuring out the best strategies. This concept enables the agent to assess the desirability of different states and actions, guiding the learning process towards strategic decision-making that maximizes cumulative rewards.

Dynamic Programming

While not conventionally classified as a RL technique, dynamic programming (DP), as outlined in [6], holds significant theoretical relevance in the RL paradigm and is therefore discussed here. DP methods are instrumental in the direct resolution of MDPs when the underlying dynamics, such as transition probability functions and reward functions, are known.

The state value function, a core component in MDP analysis, can be recursively expressed through the **Bellman equation**:

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s], \tag{1.2}$$

where $s \in \mathcal{S}$. This recursive formulation delineates the expected return for a given state s under policy π. The equation considers the immediate reward R_{t+1} and the discounted value of the subsequent state S_{t+1}.

Furthermore, for the optimal value function v^*, the *Bellman optimality equation* takes the form:

$$v_*(s) = \max_a \mathbb{E}[R_{t+1} + \gamma v_*(S_{t+1}) \mid S_t = s, A_t = a],$$

where $s \in \mathcal{S}, a \in \mathcal{A}$. This equation extends the recursive logic to the optimal context, where the state value function represents the maximum expected return under the optimal policy.

Similarly, *Bellman optimality equation* for q^* is

$$q^*(s, a) = \mathbb{E}[R_{t+1} + \gamma \max_{a'} q_*(S_{t+1}, a') \mid S_t = s, A_t = a],$$

for all $s, s' \in \mathcal{S}, a \in \mathcal{A}$.

DP methods offer a foundational framework for solving MDPs by breaking down complex problems through recursive transformations of Bellman equations. Among these, Value Iteration stands out as a well-explored DP algorithm [48]. This, alongside Policy Iteration, forms part of the broader family of Generalized Policy Iteration (GPI) approaches. Initiating with arbitrary policy values, Value Iteration systematically performs recursive backward updates based on the Bellman optimality equation until convergence is achieved. Notably, the convergence conditions align with those ensuring the existence of the optimal state value function v_*.

Despite their theoretical significance, DP methods encounter limited practical applicability due to the rarity of perfect knowledge regarding the model and their computational

infeasibility when dealing with extensive state spaces. Their computational cost proves prohibitive, hindering their use in practice. Nevertheless, DP retains profound theoretical importance, as many RL methods approximate the Bellman optimality equation without a complete understanding of underlying process dynamics. Instead, these RL methods rely solely on data acquired through sampling, highlighting the enduring theoretical relevance of DP in the broader landscape of RL research.

Value-Based Methods

Value-based methods, aiming to discover the optimal policy, often resort to approximating the optimal state–action value function $q_*(s, a)$. In this realm, Temporal-Difference (TD) methods play a vital role, leveraging bootstrapping from the current value function estimate.

A prominent TD control algorithm is **Q-learning** [51], known for its simplicity and effectiveness. Its one-step variant, defined by the following update rule:

$$q(S_t, A_t) \leftarrow q(S_t, A_t) + \alpha\big[R_{t+1} + \gamma \max_a q(S_{t+1}, a) - q(S_t, A_t)\big], \tag{1.3}$$

where $q(S_t, A_t)$ is the estimated state–action value, α is the learning rate, R_{t+1} is the immediate reward, γ is the discount factor, and S_t and S_{t+1} are the current and next states, respectively.

Another noteworthy TD algorithm is **SARSA** (State-Action-Reward-State-Action) [43], where the update rule becomes:

$$q(S_t, A_t) \leftarrow q(S_t, A_t) + \alpha\big[R_{t+1} + \gamma q(S_{t+1}, A_{t+1}) - q(S_t, A_t)\big]. \tag{1.4}$$

Here, A_{t+1} is the action taken in the next state. TD methods, being sample-based and computationally more feasible, are crucial in practice, especially when dealing with large state spaces.

In both SARSA and Q-learning, the algorithms are updated based on the temporal difference between the predicted value of a state-action pair and the actual observed value. This difference, multiplied by a learning rate, is used to adjust the estimate of the state-action value. The key distinction lies in how these methods handle the update.

- Q-learning (Off-policy): Q-learning is an off-policy TD control algorithm. It updates its Q-values based on the maximum estimated Q-value for the next state, regardless of the action actually taken. This characteristic allows Q-learning to learn an optimal policy while following a different, exploratory policy.
- SARSA (On-policy): SARSA, on the other hand, is an on-policy TD control algorithm. It updates its Q-values based on the actual action taken in the next state. This means that SARSA takes into account the policy it follows during exploration, making it more cautious and ensuring that the learned policy is the one being followed.

Alternative RL algorithms, such as Policy-Based Methods, take a distinct approach by directly parametrizing the policy π_θ with a parameter θ. One of the most significant advancements in this category is the Policy Gradient (PG) methods [49]. Another noteworthy approach is the Actor-Critic Methods (AC), strategically marrying the strengths of actor-only (policy-based) and critic-only (value-based) methods. Algorithms falling under this umbrella include Proximal Policy Optimization (PPO) [44] and the Deterministic Policy Gradient (DPG) algorithm [45], etc. For a more comprehensive understanding, interested readers are encouraged to delve into the references provided above.

Value Function Approximation

The conventional approach to representing value functions, $v(s)$ or $q(s, a)$, employs tabular structures where each state or state-action pair corresponds to a cell in the table. However, this method becomes computationally infeasible when dealing with continuous state spaces or a large number of states. Navigating all states or state-action pairs for value function updates becomes impractical with limited computational resources. Therefore, a scalable representation that offers an approximate yet effective solution is imperative. Value function approximation (VFA) is one such approach that aims to generalize learnings across states, enabling efficient learning in complex environments.

A widely used technique in VFA is **tile coding** [48], a method that discretizes the continuous state space into a set of overlapping tiles. Each tile corresponds to a binary feature, indicating whether a state falls within its boundaries. This method provides a structured representation of the state space, enabling the generalization of learned values across similar states. Tile coding is particularly advantageous in scenarios where state transitions are smooth, and it significantly reduces the dimensionality of the state space.

Here's how tile coding works:

1. Partitioning the State Space:
    - Divide the continuous state space into a set of tiles or grids.
    - Create multiple tilings with different offsets.
2. Binary Feature Activation:
    - For each tile, create a binary feature indicating whether the current state falls within that tile.

- Repeat this process for each tiling.

3. Overlapping Tilings:

   - Ensure that the tilings overlap to capture nuances in the state space.
   - Adjust tile offsets to create different but partially overlapping representations.

4. Feature Vector:

   - Combine the binary features from all tilings to create a feature vector that represents the state.

In the context of TD methods, specifically SARSA, the incorporation of VFA becomes crucial. The classic SARSA update rule is given by Eq. (1.4), while the tile-coded SARSA update rule involves updating the weights associated with the features (tiles) in the Q-function approximation. The update is based on the TD error, which is the difference between the observed reward and the predicted Q-value.

The update rule for the weights associated with the features is defined as follows:

$$w_i \leftarrow w_i + \alpha \delta x_i. \tag{1.5}$$

Here, w_i is the weight associated with the i-th feature (tile), α is the learning rate, δ is the TD error, x_i is the binary feature (0 or 1) indicating whether the state belongs to the i-th tile.

In the SARSA update rule, the TD error δ is given by

$$\delta = R_{t+1} + \gamma q(S_{t+1}, A_{t+1}, \mathbf{w}) - q(S_t, A_t, \mathbf{w}). \tag{1.6}$$

In this expression, $q(S_{t+1}, A_{t+1}, \mathbf{w})$ and $q(S_t, A_t, \mathbf{w})$ represent the Q-values for the next and current state-action pairs, respectively. These Q-values are the result of the linear combination of weights and features:

$$q(S_{t+1}, A_{t+1}, \mathbf{w}) = \sum_i w_i x_i(S_{t+1}, A_{t+1}),$$
$$q(S_t, A_t, \mathbf{w}) = \sum_i w_i x_i(S_t, A_t).$$

Here, the weights w are implicitly present in the computation of Q-values through the feature vectors, and the update rule adjusts these weights based on the TD error.

In Q-learning with VFA, the goal is to estimate the Q-values of state-action pairs using a function approximator. The update rule for Q-learning with VFA and tile coding is similar to SARSA, as shown in Eq. (1.5). The TD error δ for Q-learning with VFA is expressed as

$$\delta = R_{t+1} + \gamma \max_a q(S_{t+1}, a, \mathbf{w}) - q(S_t, A_t, \mathbf{w}). \tag{1.7}$$

Here, $\max_a q(S_{t+1}, a, \mathbf{w})$ represents the maximum Q-value over all possible actions in the next state. The Q-values are calculated using the weights and feature vectors as follows

$$q(S_{t+1}, a, \mathbf{w}) = \sum_i w_i x_i(S_{t+1}, a).$$

Similar to SARSA, the weights are adjusted based on the TD error, and tile coding facilitates a concise representation of the state space.

Exploration vs. Exploitation

The exploration-exploitation dilemma is a pivotal challenge faced by agents striving to maximize cumulative rewards. These two strategies, exploration and exploitation, encapsulate the agent's approach to decision-making within its environment.

Exploration refers to the agent's strategy of trying new actions or visiting new states that it has not yet extensively explored. The goal of exploration is to gather information about the environment, discover unknown aspects, and potentially find actions that lead to higher rewards. Exploration is inherently risky because the agent is venturing into unknown territory, and there's a chance that the chosen actions may lead to lower immediate rewards.

On the flip side, **exploitation** involves the agent selecting actions that are believed to yield the highest immediate reward based on the current knowledge. This strategy aims to capitalize on known actions or states to achieve the highest short-term gains. Exploitation is less risky than exploration since it relies on familiar territory, yet it carries the potential downside of settling for suboptimal long-term performance if the environment undergoes changes.

The delicate Exploration-Exploitation Dilemma arises in dynamic environments, where the agent must strike a balance between these two strategies. Early on, exploration takes precedence as the agent must traverse a breadth of actions and states to build a foundational understanding of the environment. Later in the learning process, a shift toward exploitation occurs, emphasizing the maximization of rewards based on acquired knowledge. However, too much reliance on exploitation without occasional exploration might lead to a suboptimal policy.

One widely used algorithm embodying this trade-off is the **epsilon-greedy algorithm** [48]. This approach introduces randomness by selecting a random action with a small probability ϵ, promoting exploration. With probability $1 - \epsilon$, it exploits the current knowledge by selecting the action with the highest estimated value.

In conclusion, the exploration-exploitation balance is a nuanced challenge in RL. Effective strategies must judiciously manage the tension between exploration and exploitation, adapting to the learning task's characteristics and the dynamic nature of the environment.

1.3.2 Reinforcement Learning on Market Marking

RL has emerged as a promising tool in the field of market making, revolutionizing trading strategies and reshaping the landscape of financial markets. Over the past few decades, RL has undergone a remarkable evolution, evolving from theoretical concepts to practical applications in various domains, including finance and trading. The application of RL in market making traces back to the late 20th century. A pivotal contribution to RL-based market making strategies was made by [20], presenting arguably the first model-free RL approach to optimal market making. In this seminal work, the agent dynamically sets bid/ask prices based on its inventory level, order imbalance, and market quality measures. Grounded in the Glosten–Milgrom information-based model, the approach considers distinct market participants: a monopolistic market maker, informed traders, and uninformed traders. The market maker, by adjusting bid/ask prices, implicitly tracks the underlying stock's true value through a stochastic process not directly accessible to it.

Building upon this foundation, [32] improved the RL-based market making approach by incorporating order flow dynamics modeled as an input–output hidden Markov model.

This introduced a more intricate, though noninformation-based, model by integrating the order book dynamics into consideration. Their approach involved running an RL algorithm based on likelihood ratios within the resulting partially observable environment. [37] extended the work of [20] by introducing a risk-sensitive RL approach, drawing on Mihatsch–Neuneier one-step temporal difference learning algorithms [38]. Their study emphasized the significance of employing a Boltzmann softmax action-selection rule for successful risk-averse market making, demonstrating that it leads to substantial profits while keeping inventory levels low.

The early incorporation of RL into market making revealed its potential in capturing complex market patterns. This technology empowered agents to learn from historical data, continuously refining their strategies to adeptly navigate the complexities of financial markets. RL-based market making systems demonstrated notable adaptability, showcasing the ability to learn from evolving market conditions and maintaining robustness in the face of uncertainty.

[36] proposed a straightforward tabular Q-learning-based approach, employing the constant absolute risk aversion (CARA) utility. The study delved into the influence of the CARA utility, concluding that the RL approach surpassed analytical Absolute Spread approximations and the zero tick offset benchmark concerning cumulative profit and inventory metrics. [8] crafted an market making agent utilizing temporal-difference RL. The study explored various state representations and reward function formulations, amalgamating the best-performing elements into a single agent exhibiting superior risk-adjusted performance. Notably, the optimal solution employed a linear combination of tile codings as a value function approximator and an asymmetrically dampened profit and loss function as a reward function. These endeavors underline the diverse ways RL can be

harnessed to enhance market-making strategies, providing valuable insights into optimal approaches and representations.

Advancements in RL techniques further enhanced the adaptive capabilities of market making algorithms. The evolution of deep reinforcement learning (DRL) introduced more sophisticated algorithms, leveraging deep neural networks to process vast amounts of data and extract complex features for decision-making. [34] engineered an market making agent grounded in a modified deep recurrent Q network (DRQN). This agent underwent training in a highly realistic simulator of the LOB and demonstrated its prowess by outperforming the benchmark proposed by [8]. [25] introduced a groundbreaking DRL framework for market making incorporating signals, emphasizing the interpretability of the learned controls. The DRL agent derived exhibited remarkable performance surpassing several benchmark strategies, including the approximations proposed by [28].

The adoption of RL in market making has not been without challenges. The complexity of financial markets, regulatory constraints, and the need to ensure the stability and fairness of market operations have presented formidable hurdles. Nonetheless, ongoing research and developments in RL continue to refine these algorithms, striving to address these challenges and improve the efficacy of market making strategies.

Chapter 2

Market Making with Latency

In this chapter, we adopt the frameworks in [12] and [24] to study a stochastic control model designed to capture the behavior of an electronic market maker navigating a trading environment in the presence of latency. The market maker's primary objective is to maximize her expected terminal wealth. To analyze and solve the control problem, we will reformulate it as a finite-horizon MDP, which ultimately can be solved numerically. In other words, the discrete time MDP serves as a framework to facilitate optimal decision-making at every stage of the trading period. It takes into account various factors such as market conditions, order book dynamics, latency effects, and inventory management.

2.1 Model Setup and Assumptions

2.1.1 Market Making with Latency

Latency, a characteristic inherent in electronic trading systems, arises due to various factors, including network communication time, system processing, and the time required for the order to traverse the trading infrastructure. As a result, the market maker's trading decisions, reflected in the submitted LOs, are not immediately reflected in the LOB's state. Instead, these LOs are only incorporated after a certain duration, corresponding

to the latency period. This latency-induced delay introduces an element of uncertainty and potential mismatch between the market maker's intended order placement and the actual placement of them. Market conditions can rapidly evolve during this latency period, leading to changes in the order book state, price movements, and the arrival of other market participants' orders. Consequently, the market maker faces challenges in accurately predicting the immediate market environment and optimizing her trading decisions. Understanding and accounting for latency effects are crucial for electronic market makers.

Given an asset, we assume that the market maker will implement her strategy from time 0 up to a fixed trading time horizon T, performing actions every Δt time unit on the LOB, specifically, at times $0 = t_0 < t_1 < \ldots < t_N$, where $t_N < T$. For simplicity, we adopt a regular time design:

$$t_k := k \cdot \Delta t, \text{ for } k = 0, 1, \ldots, N. \tag{2.1}$$

Throughout, we set $t_{N+1} = T$ and denote $\mathcal{T} = \{t_0, t_1, \ldots, t_{N+1}\}$. All variables introduced below are defined within the context of a probability space denoted as $(\Omega, \mathbb{P}, \mathcal{F})$, which is equipped with a filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$. This filtration captures the evolving information available to the market maker over time. Broadly, $\mathcal{F}_t$ consists of "all information" available to the market maker up to time t.

As in [24, 39], we assume the market maker experiences a constant latency of size $\Delta\tau \in [0, \Delta t]$. Specifically, we assume the action submitted at time t_k is registered by the LOB at time

$$t_{k+} := t_k + \Delta\tau, \text{ for } k = 0, 1, \ldots, N. \tag{2.2}$$

Of particular importance is the unwinding of the market maker's terminal inventory using a MO. As the trading period progresses, the market maker accumulates inventory that needs to be liquidated before the trading horizon concludes. To achieve this, the market maker employs a final MO at time t_N. The unwinding MO instruction arrives at the exchange at time $t_{N+} = t_N + \Delta\tau$.

As mentioned above, latency introduces a delay in the registration of ask and bid LOs into the LOB. Specifically, when a market maker submits an ask or bid LO at time t_k, it experiences a time lag until the order is actually recorded in the LOB, which we assume occurs at time $t_{k+} = t_k + \Delta\tau$. Once the market maker's ask and bid LOs are successfully registered in the LOB, they are subject to full or partial execution, which can occur in one of two ways:

(i) If the price of the submitted ask (bid) LO, when it is registered by the LOB, is less (more) than or equal to the current best bid (ask) price in the market, the LO will be fully executed instantly at the best bid (ask) price.

(ii) Otherwise, the order will be added to the LOB and may only be executed partially or in full if a MO matches its limit price at a subsequent time.

Further discussion about the number of filled shares will be provided in the following subsection.

2.1.2 State and Action Spaces and System Dynamics

At each time $t_k \in [0, T]$, we denote W_{t_k} and S_{t_k} the market maker's wealth and the fundamental price of the asset, respectively, and let I_{t_k} represent the market maker's

inventory. Throughout, best ask price is assumed to be $S_{t_k} + \frac{\text{tick size}}{2}$, while the best bid price is assumed to be $S_{t_k} - \frac{\text{tick size}}{2}$. In USA markets, tick size is usually 0.01. To control the running inventory risk, we constrain I_{t_k} to lie between a lower bound $\underline{I}$ and an upper bound $\overline{I}$. We denote the number of outstanding ask and bid orders at time t_k, which are left unfilled during $[t_{k-1}, t_k)$, by $Q^+_{out_{t_k}}$ and $Q^-_{out_{t_k}}$, respectively. We restrict $Q^+_{out_{t_k}}$ and $Q^-_{out_{t_k}}$ to be non-negative integers less than a constant $Q^{\pm}_{\max}$, which represents the volume of each LO submitted by the market maker. We assume the volume $Q^{\pm}_{\max}$ to be constant. Additionally, the corresponding quotes of the outstanding LOs at time t_k, relative to the fundamental price S_{t_k}, are denoted by $(r^+_{t_k}, r^-_{t_k})$. See Eqs. (2.13)-(2.14) below for precise formulas for $(Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k})$. We adopt the convention that whenever $r^{\pm}_{t_k} = \infty$, it means that there are no outstanding orders at time t_k. The state variables of the system then consist of $\Phi_{t_k} := (W_{t_k}, S_{t_k}, I_{t_k}, Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k})$, which take values in the state space

$$\Phi := \big\{(w, s, i, q^+, q^-, r^+, r^-) : w \in \mathbb{R}, s \in \mathbb{R}_+, i \in \mathbb{Z}, \underline{I} \le i \le \overline{I}, (q^+, q^-) \in \mathbb{Z}^2_+, \\ 0 \le q^{\pm} \le Q^{\pm}_{\max}, (r^+, r^-) \in \overline{\mathbb{Z}}^2_+\big\}.$$

We now proceed to describe the set of admissible actions, denoted as $\mathcal{A}$. At the beginning of each time period, the market maker observes the LOB status and based on it, considers multiple possible actions. These actions include cancelling existing LOs and placing new bid and ask quotes, cancellation of existing orders without placing new orders, or adopting a "doing nothing" approach. More precisely, her actions at time t_k will be represented by the pair $(L^+_{t_k}, L^-_{t_k})$, where $L^+_{t_k}, L^-_{t_k} \in \mathbb{Z}_+ \cup \{\infty\} \cup \{o\}$ (hereafter, $\mathbb{Z}_+ = \{0, 1, 2 \dots\}$). If $L^+_{t_k}$ takes on a value in $\mathbb{Z}_+$, the market maker will cancel any outstanding ask order, and then proceed to quote a new ask order of volume $Q^+_{\max}$ at the price $S_{t_k} + L^+_{t_k}$. That is, $L^+_{t_k}$ is the relative price of the asset when compared to the fundamental price S_{t_k}. If

$L^+_{t_k} = \infty$, the market maker will cancel any outstanding ask orders without placing a new one, while if $L^+_{t_k} = o$, the market maker will take no action on the ask side, leaving intact any outstanding LOs at time t_k. Similar notation is used on the bid side.

We now show how the system transitions from time t_k to t_{k+1} for $k = 0, 1, \cdots, N-1$. Let $\Delta S_{t_k} := S_{t_{k+}} - S_{t_k}$ and $\Delta S_{t_{k+}} := S_{t_{k+1}} - S_{t_{k+}}$ denote the price changes on the intervals $[t_k, t_{k+})$ and $[t_{k+}, t_{k+1})$, respectively. It becomes clear that the fundamental price satisfies the relation

$$S_{t_{k+1}} = S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}}. \tag{2.3}$$

The cash holding and inventory processes are given as

$$W_{t_{k+1}} = W_{t_k} + P_{fill^+_{t_k}} Q_{fill^+_{t_k}} - P_{fill^-_{t_k}} Q_{fill^-_{t_k}} + P_{fill^+_{t_{k+}}} Q_{fill^+_{t_{k+}}} - P_{fill^-_{t_{k+}}} Q_{fill^-_{t_{k+}}}, \tag{2.4}$$

$$I_{t_{k+1}} = I_{t_k} - Q_{fill^+_{t_k}} + Q_{fill^-_{t_k}} - Q_{fill^+_{t_{k+}}} + Q_{fill^+_{t_{k+}}}, \tag{2.5}$$

where $P_{fill^+_{t_k}}$ and $Q_{fill^+_{t_k}}$ respectively denote the execution price and the number of filled shares of any outstanding ask LO on the interval $[t_k, t_{k+})$, while $P_{fill^+_{t_{k+}}}$ and $Q_{fill^+_{t_{k+}}}$ denote the execution price and the number of filled shares of ask LOs on the interval $[t_{k+}, t_{k+1})$. Analogously, $P_{fill^-_{t_k}}$, $Q_{fill^-_{t_k}}$, $P_{fill^-_{t_{k+}}}$, $Q_{fill^-_{t_{k+}}}$ represent the corresponding quantities on the bid side. From the explanation at the end of Subsection 2.1.1 and recalling that $r^+_{t_k}$ and $r^-_{t_k}$ are the prices of any outstanding ask and bid LO at time t_k, respectively, it is clear

that the execution prices are determined as follows:

$$P_{fill^{\pm}_{t_k}} = \begin{cases} S_{t_k} \pm r^{\pm}_{t_k}, & \text{if } r^{\pm}_{t_k} \in \mathbb{Z}_+, \\ 0, & \text{if } r^{\pm}_{t_k} = \infty, \end{cases} \tag{2.6}$$

$$P_{fill^{\pm}_{t_{k+}}} = \begin{cases} \mathbb{1}_{\left\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\right\}} \left[S_{t_k} + \Delta S_{t_k}\right] + \mathbb{1}_{\left\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\right\}} \left[S_{t_k} \pm L^{\pm}_{t_k}\right], & \text{if } L^{\pm}_{t_k} \in \mathbb{Z}_+, \\ S_{t_k} \pm r^{\pm}_{t_k}, & \text{if } L^{\pm}_{t_k} = o, \\ 0, & \text{if } L^{\pm}_{t_k} = \infty. \end{cases} \tag{2.7}$$

To describe the number of filled shares, it is important to separately consider the cases $L^{\pm}_{t_k} \in \mathbb{Z}_+ \cup \{\infty\}$ and $L^{\pm}_{t_k} = o$. To model the number of filled shares from an outstanding LO during a given time period, we adopt a more realistic version of the stochastic linear model of [12]. Broadly, the number of filled shares will depend on whether there are any MO arriving during that interval and the distance between the quoted price and the fundamental price. More specifically, for the ask side, the number of filled shares during $[u, v)$ from a LO of size q placed at level $S_u + r$ is given by

$$\mathbb{1}_{fill^+} c(p - r)_+ \wedge q, \tag{2.8}$$

where $x_+ = \max(x, 0)$, $a \wedge b = \min\{a, b\}$, and c, p are $\mathcal{F}_v$-measurable r.v.'s, $\mathbb{1}_{fill^+} \in \mathcal{F}_v$ is 1 if there are any buy MOs arriving during $[u, v)$ and 0, otherwise. The so-called reservation price p is understood as the maximum price, relative to S_u, achievable by all buy MOs placed during the time interval $[u, v)$. Meanwhile, the demand slope c measures the rate of increase in the number of filled shares of the ask LO as its ask price approaches the fundamental price S_u within the same interval $[u, v)$.

Following the prescription above and recalling that $Q^{\pm}_{out_{t_k}}$ denotes the volume of any outstanding LO at time t_k and $Q^{\pm}_{\max}$ is the volume of submitted LOs, for $L^{\pm}_{t_k} \in \overline{\mathbb{Z}}_+$, we have

$$Q_{fill^{\pm}_{t_k}} := \mathbb{1}_{fill^{\pm}_k} c^{\pm}_{t_k} \left[p^{\pm}_{t_k} - r^{\pm}_{t_k}\right]_+ \wedge Q^{\pm}_{out_{t_k}}, \tag{2.9}$$

$$Q_{fill^{\pm}_{t_{k+}}} := \mathbb{1}_{\left\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\right\}} Q^{\pm}_{\max} + \mathbb{1}_{\left\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\right\}} \mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} [p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}]_+ \wedge Q^{\pm}_{\max}, \tag{2.10}$$

where $c^{\pm}_{t_k}, p^{\pm}_{t_k} \in \mathcal{F}_{t_{k+}}$, $c^{\pm}_{t_{k+}}, p^{\pm}_{t_{k+}} \in \mathcal{F}_{t_{k+1}}$, and

$$\begin{aligned}
\mathbb{1}_{fill^{+}_{k}} &= \mathbb{1}_{\{\text{At least one buy MO arrives during } [t_k, t_{k+})\}}, \\
\mathbb{1}_{fill^{-}_{k}} &= \mathbb{1}_{\{\text{At least one sell MO arrives during } [t_k, t_{k+})\}}, \\
\mathbb{1}_{fill^{+}_{k+}} &= \mathbb{1}_{\{\text{At least one buy MO arrives during } [t_{k+}, t_{k+1})\}}, \\
\mathbb{1}_{fill^{-}_{k+}} &= \mathbb{1}_{\{\text{At least one sell MO arrives during } [t_{k+}, t_{k+1})\}}.
\end{aligned}$$

The precise conditions on the variables above are given in Assumption 2.1.1 below.

We now consider the case $L^{\pm}_{t_k} = o$ ("doing nothing"). Note that this means that any outstanding LOs at time t_k remain untouched at time t_{k+}. In that situation, $P_{fill^{\pm}_{t_k}} = P_{fill^{\pm}_{t_{k+}}} = S_{t_k} \pm r^{\pm}_{t_k}$ and the updated cash holding and inventory at time t_{k+1}, $W_{t_{k+1}}$ and $I_{t_{k+1}}$, only depend on the total demand $Q_{fill^{\pm}_{t_k}} + Q_{fill^{\pm}_{t_{k+}}}$ over the whole interval $[t_k, t_{k+1})$ out of $Q^{\pm}_{out_{t_k}}$. Then, it makes more sense to model such a demand jointly. That is, when $L^{\pm}_{t_k} = o$, we set

$$\widetilde{Q}_{fill^{\pm}_{t_k}} := Q_{fill^{\pm}_{t_k}} + Q_{fill^{\pm}_{t_{k+}}} := \widetilde{\mathbb{1}}_{fill^{\pm}_k} \tilde{c}^{\pm}_{t_k} \left[\tilde{p}^{\pm}_{t_k} - r^{\pm}_{t_k}\right]_+ \wedge Q^{\pm}_{out_{t_k}}. \tag{2.11}$$

where

$$\begin{aligned}\widetilde{\mathbb{1}}_{fill_k^+} &= \mathbb{1}_{\{\text{At least one buy MO arrives during } [t_k, t_{k+1})\}},\\ \widetilde{\mathbb{1}}_{fill_k^-} &= \mathbb{1}_{\{\text{At least one sell MO arrives during } [t_k, t_{k+1})\}},\end{aligned} \tag{2.12}$$

and $\tilde{c}^{\pm}_{t_k}, \tilde{p}^{\pm}_{t_k} \in \mathcal{F}_{t_{k+1}}$.

We can now specify the number of outstanding shares for ask and bid orders at time t_{k+1}, for $k = 0, 1, \ldots, N-1$,

$$Q^{\pm}_{out_{t_{k+1}}} := \begin{cases} Q^{\pm}_{\max} - Q_{fill^{\pm}_{t_{k+}}}, & \text{if } L^{\pm}_{t_k} \in \mathbb{Z}_+,\\ Q^{\pm}_{out_{t_k}} - \widetilde{Q}_{fill^{\pm}_{t_k}}, & \text{if } L^{\pm}_{t_k} = o,\\ 0, & \text{if } L^{\pm}_{t_k} = \infty, \end{cases} \tag{2.13}$$

and setting as initial value $Q^{\pm}_{out_{t_0}} := 0$. Finally, the ask and bid outstanding quote pair $r^{\pm}_{t_{k+1}}$ is determined as

$$r^{\pm}_{t_{k+1}} := \begin{cases} L^{\pm}_{t_k} \mp (\Delta S_{t_k} + \Delta S_{t_{k+}}), & \text{if } L^{\pm}_{t_k} \in \overline{\mathbb{Z}}_+,\\ r^{\pm}_{t_k} \mp (\Delta S_{t_k} + \Delta S_{t_{k+}}), & \text{if } L^{\pm}_{t_k} = o, \end{cases} \tag{2.14}$$

and initial value $r^{\pm}_{t_0} = \infty$.

Next, we outline the system dynamics from time t_N to t_{N+}, when the market maker does not post new quotes and only unwinds her inventory position using a MO at time t_N

(though, this will be executed until time $t_N + \Delta\tau$). Specifically, we have

$$\begin{aligned}
W_{t_{N+}} &= W_{t_N} + P_{fill^+_{t_N}} Q_{fill^+_{t_N}} - P_{fill^-_{t_N}} Q_{fill^-_{t_N}} \\
&= W_{t_N} + (S_{t_N} + r^+_{t_N})(\mathbb{1}_{fill^+_N} \cdot c^+_{t_N} \left[p^+_{t_N} - r^+_{t_N}\right]_+ \wedge Q^+_{out_{t_N}}) \\
&\qquad - (S_{t_N} - r^-_{t_N})(\mathbb{1}_{fill^-_N} \cdot c^-_{t_N} \left[p^-_{t_N} - r^-_{t_N}\right]_+ \wedge Q^-_{out_{t_N}}), \\
S_{t_{N+}} &= S_{t_N} + \Delta S_{t_N}, \\
I_{t_{N+}} &= I_{t_N} - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}} \\
&= I_{t_N} - \mathbb{1}_{fill^+_N} \cdot c^+_{t_N} \left[p^+_{t_N} - r^+_{t_N}\right]_+ \wedge Q^+_{out_{t_N}} + \mathbb{1}_{fill^-_N} \cdot c^-_{t_N} \left[p^-_{t_N} - r^-_{t_N}\right]_+ \wedge Q^-_{out_{t_N}}.
\end{aligned}$$

We now turn our attention to the assumptions on the variables involved in the demand formulas (these are similar to those given in [12]).

Assumption 2.1.1. *For $k = 0, 1, \ldots, N$, we have*

1. *$\mathbb{1}_{fill^\pm_k} \in \mathcal{F}_{t_{k+}}$ and $\mathbb{1}_{fill^\pm_{k+}}, \widetilde{\mathbb{1}}_{fill^\pm_k} \in \mathcal{F}_{t_{k+1}}$ are such that*

$$\pi^\pm_{t_k} = \mathbb{P}(\mathbb{1}_{fill^\pm_k} = 1 \mid \mathcal{F}_{t_k}),\ \pi^\pm_{t_{k+}} = \mathbb{P}(\mathbb{1}_{fill^\pm_{k+}} = 1 \mid \mathcal{F}_{t_k}),\ \tilde{\pi}^\pm_{t_k} = \mathbb{P}(\widetilde{\mathbb{1}}_{fill^\pm_k} = 1 \mid \mathcal{F}_{t_k}), \tag{2.15}$$

   for some deterministic probabilities $\pi^\pm_{t_k}$, $\pi^\pm_{t_{k+}}$, and $\tilde{\pi}^\pm_{t_k}$;

2. *$(\mathbb{1}_{fill^+_k}, \mathbb{1}_{fill^+_{k+}})$ and $(\mathbb{1}_{fill^-_k}, \mathbb{1}_{fill^-_{k+}})$ are independent, conditional on $\mathcal{F}_{t_k}$;*

3. *$(c^\pm_{t_k}, p^\pm_{t_k})$ are $\mathcal{F}_{t_{k+}}$-measurable, while $(c^\pm_{t_{k+}}, p^\pm_{t_{k+}})$ and $(\tilde{c}^\pm_{t_k}, \tilde{p}^\pm_{t_k})$ are $\mathcal{F}_{t_{k+1}}$-measurable;*

4. *The conditional distribution of $(c^\pm_{t_k}, p^\pm_{t_k})$ given $(\mathcal{F}_{t_k}, \mathbb{1}_{fill^\pm_k})$ is a function of only the indicator $\mathbb{1}_{fill^\pm_k}$ that does not depend on k;*

5. *The conditional distribution of $(c^\pm_{t_{k+}}, p^\pm_{t_{k+}})$ given $(\mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^\pm_{k+}})$ is a function of only the indicator $\mathbb{1}_{fill^\pm_{k+}}$ that does not depend on k;*

6. *The conditional distribution of* $(\tilde{c}^{\pm}_{t_k}, \tilde{p}^{\pm}_{t_k})$ *given* $(\mathcal{F}_{t_k}, \widetilde{\mathbb{1}}_{fill^{\pm}_k})$ *is a function of only the indicator* $\widetilde{\mathbb{1}}_{fill^{\pm}_k}$ *that does not depend on* k*;*

7. $(c^{+}_{t_k}, p^{+}_{t_k})$ *and* $(c^{-}_{t_k}, p^{-}_{t_k})$ *are conditionally independent given* $(\mathcal{F}_{t_k}, \mathbb{1}_{fill^{+}_k}, \mathbb{1}_{fill^{-}_k})$*. Similarly,* $(c^{+}_{t_{k+}}, p^{+}_{t_{k+}})$ *and* $(c^{-}_{t_{k+}}, p^{-}_{t_{k+}})$ *are conditionally independent given* $(\mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{+}_{k+}}, \mathbb{1}_{fill^{-}_{k+}})$*,*

8. $\mathbb{1}_{fill^{\pm}_{k+}}$ *and* $(\mathbb{1}_{fill^{\pm}_k}, c^{\pm}_{t_k}, p^{\pm}_{t_k})$ *are conditional independent given* $\mathcal{F}_{t_k}$*.*

Remark 2.1.1. *Assumptions 2.1.1.4-2.1.1.6 above mean that there exist cdfs* $F^{\pm}_{\Delta\tau} : \mathbb{R}^2 \times \{0,1\} \to [0,1]$*,* $\check{F}^{\pm}_{\Delta\tau,\Delta t} : \mathbb{R}^2 \times \{0,1\} \to [0,1]$*, and* $\tilde{F}^{\pm}_{\Delta t} : \mathbb{R}^2 \times \{0,1\} \to [0,1]$ *such that, for all* $k = 0, 1, \ldots, N$*,*

$$\begin{aligned}
&\mathbb{P}(c^{\pm}_{t_k} \le x, p^{\pm}_{t_k} \le y \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill^{\pm}_k}) = F^{\pm}_{\Delta\tau}(x, y, \mathbb{1}_{fill^{\pm}_k}),\\
&\mathbb{P}(c^{\pm}_{t_{k+}} \le x, p^{\pm}_{t_{k+}} \le y \mid \mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{\pm}_{k+}}) = \check{F}^{\pm}_{\Delta\tau,\Delta t}(x, y, \mathbb{1}_{fill^{\pm}_{k+}}),\\
&\mathbb{P}(\tilde{c}^{\pm}_{t_k} \le x, \tilde{p}^{\pm}_{t_k} \le y \mid \mathcal{F}_{t_k}, \widetilde{\mathbb{1}}_{fill^{\pm}_k}) = \tilde{F}^{\pm}_{\Delta t}(x, y, \widetilde{\mathbb{1}}_{fill^{\pm}_{k+}}).
\end{aligned} \tag{2.16}$$

Next, we establish the assumptions on the fundamental price. Below, we use the shorthand notation $\Delta S[u,v) := S_v - S_u$, for any times $u < v$

Assumption 2.1.2. *For* $k = 0, 1, \ldots, N$*, we have*

1. *The fundamental price* $\{S_t\}_t$ *is a martingale:*

$$\mathbb{E}[S_v \mid \mathcal{F}_u] = S_u, \quad \textit{for all} \quad u < v;$$

2. $\Delta S[t_k, t_{k+}]$ *and* $(\mathbb{1}_{fill^{\pm}_k}, c^{\pm}_{t_k}, p^{\pm}_{t_k}, \mathbb{1}_{fill^{\pm}_{k+}}, c^{\pm}_{t_{k+}}, p^{\pm}_{t_{k+}})$ *are conditional independent given* $\mathcal{F}_{t_k}$*;*

3. $\Delta S[t_k, t_{k+1}]$ *and* $(\widetilde{\mathbb{1}}_{fill^{\pm}_k}, \tilde{c}^{\pm}_{t_k}, \tilde{p}^{\pm}_{t_k})$ *are conditional independent given* $\mathcal{F}_{t_k}$*;*

4. $\Delta S[t_{k+}, t_{k+1}]$ *and* $(\mathbb{1}_{fill_{k+}^{\pm}}, c_{t_{k+}}^{\pm}, p_{t_{k+}}^{\pm})$ *are conditional independent given* $\mathcal{F}_{t_{k+}}$*;*

5. $\Delta S[t_k, t_{k+}]$ *is independent of* $\mathcal{F}_{t_k}$*.*

6. $\Delta S[t_k, t_{k+}] \overset{\mathcal{D}}{=} S_{\Delta\tau}$*.*

2.2 Characterization of the Optimal Market Making Strategy

Given the framework established in Section 2.1, we will proceed to reformulate the optimal market making problem as a MDP with an immediate reward structure, which is explicitly identified. In turn, this characterization will allow us to numerically find the optimal strategy for the market maker.

The objective of the market maker is to maximize her expected terminal wealth at the end of the trading period by performing an admissible action at the start of each time interval. The market maker's terminal Profit and Loss (P&L) is denoted as $\mathcal{W}$, and it is determined by the state of the market just before t_{N+}, represented by ($W_{t_{N+}}$, $S_{t_{N+}}$, $I_{t_{N+}}$, $Q^{+}_{out_{t_{N+}}}$, $Q^{-}_{out_{t_{N+}}}$, $r^{+}_{t_{N+}}$, $r^{-}_{t_{N+}}$). In our setting, the terminal P&L will be computed as

$$\mathcal{W} := W_{t_{N+}} + S_{t_{N+}} I_{t_{N+}} - \lambda I^2_{t_{N+}}, \tag{2.17}$$

where $W_{t_{N+}}$ and $I_{t_{N+}}$ are the market maker's cash holding and inventory at time t_{N+}, $S_{t_{N+}}$ is the fundamental price of the asset at t_{N+}, and λ is a positive inventory penalization constant. Our choice of the P&L function agrees with what is common in the literature (see [1, 12, 14]). The last two terms in Eq. (2.17) can be expressed as $(S_{t_{N+}} - \lambda I_{t_{N+}})I_{t_{N+}}$, where the term $S_{t_{N+}} - \lambda I_{t_{N+}}$ can be understood as the average price per share that the

market maker is expected to receive when liquidating her inventory $I_{t_{N+}}$ through a MO, assuming a linear instantaneous price impact. For instance, if $I_{t_{N+}} > 0$, the market maker will place a sell MO, resulting in the execution of trades at prices that may gradually "eat" into the bid side of the LOB. The same logic applies when $I_{t_{N+}} < 0$. Thus, the term $(S_{t_{N+}} - \lambda I_{t_{N+}})I_{t_{N+}}$ can be understood as the cash flow received by the market maker when liquidating her inventory, taking into account the impact on the LOB due to the execution of a MO.

We now describe the market maker's optimization problem. Let V_{t_0} denote the value function starting from the 0^{th} period, defined as the maximum expected terminal P&L at the end of the trading period under all Markovian admissible policies, given the initial state. Specifically,

$$V_{t_0} := \sup_{(L^{\pm}_{t_0},\ldots,L^{\pm}_{t_N})\in\mathcal{A}_{0,N}} \mathbb{E}\left[\mathcal{W} \mid \mathcal{F}_{t_0}\right], \tag{2.18}$$

where we say that $(L^{\pm}_{t_0},\ldots,L^{\pm}_{t_N}) \in \mathcal{A}_{0,N}$ if, for any $k = 0,1,\ldots,N$, $L^{\pm}_{t_k} \in \sigma(\Phi_{t_k})$ and $L^{\pm}_{t_k}$ takes values in $\mathbb{Z}_+ \cup \{\infty\} \cup \{o\}$ (recall $\Phi_{t_k} := (W_{t_k}, S_{t_k}, I_{t_k}, Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k})$). Similarly, V_{t_k} will represent the value function starting from the k^{th} period, for $k = 1,2,\ldots,N,N+$. In this sense,

$$V_{t_{N+}} := W_{t_{N+}} + S_{t_{N+}} I_{t_{N+}} - \lambda I^2_{t_{N+}},$$

while for $k = 0,1,\ldots,N$,

$$V_{t_k} := \sup_{(L^{\pm}_{t_k},\ldots,L^{\pm}_{t_N})\in\mathcal{A}_{k,N}} \mathbb{E}\left[\mathcal{W} \mid \mathcal{F}_{t_k}\right], \tag{2.19}$$

where we say that $(L^{\pm}_{t_k}, \ldots, L^{\pm}_{t_N}) \in \mathcal{A}_{k,N}$ if, for any $j = k, \ldots, N$, $L^{\pm}_{t_j} \in \sigma(\Phi_{t_j})$ and $L^{\pm}_{t_j}$ takes values in $\mathbb{Z}_+ \cup \{\infty\} \cup \{o\}$.

Having formulated the market making problem, we can formally write the Bellman equation that governs the value functions. By applying standard arguments, we obtain the following expressions:

$$V_{t_k} = \begin{cases} W_{t_{N+}} + S_{t_{N+}} I_{t_{N+}} - \lambda I^2_{t_{N+}}, & k = N+, \\ \mathbb{E}\left[V_{t_{N+}} \mid \mathcal{F}_{t_N}\right], & k = N, \\ \sup\limits_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\left[V_{t_{k+1}} \mid \mathcal{F}_{t_k}\right], & k = 0, \ldots, N-1, \end{cases} \tag{2.20}$$

where above we say $(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k$ if $L^{\pm}_{t_k} \in \sigma(\Phi_{t_k})$ and $L^{\pm}_{t_k} \in \mathbb{Z}_+ \cup \{\infty\} \cup \{o\}$.

2.2.1 Value of the Order

In this section, we introduce the concept of order value, which holds significant importance in our analysis of the optimal market making problem as a MDP. The order value serves as a fundamental metric that guides our exploration and understanding of optimal market making strategies.

Broadly, the "(intrinsic) value" of a LO is defined as its executed price relative to the fundamental value of the asset at the time of execution, times the number of LO shares that are executed or filled. More specifically, when $L^{\pm}_{t_k} \in \mathbb{Z}_+$, there are two active orders during the time interval $[t_k, t_{k+1})$: any outstanding LO left from the previous interval and the new LO which registers in the LOB at time t_{k+}. The value of the first LO is then

given by

$$VO_1^{\pm}(k) := (r_{t_k}^{\pm} \mp \Delta S_{t_k}) \cdot [\mathbb{1}_{fill_k^{\pm}} c_{t_k}^{\pm} (p_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm}], \tag{2.21}$$

while the value of the second order is

$$\begin{aligned} VO_2^{\pm}(k) := \mathbb{1}_{\{\pm\Delta S_{t_k} < L_{t_k}^{\pm}\}} \cdot (L_{t_k}^{\pm} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\mathbb{1}_{fill_{k+}^{\pm}} c_{t_{k+}}^{\pm} (p_{t_{k+}}^{\pm} - L_{t_k}^{\pm})_+ \wedge Q_{\max}^{\pm}] \\ + \mathbb{1}_{\{\pm\Delta S_{t_k} \geq L_{t_k}^{\pm}\}} \cdot (\mp \Delta S_{t_{k+}}) \cdot Q_{\max}^{\pm}. \end{aligned} \tag{2.22}$$

When $L_{t_k}^{\pm} = o$ ("doing nothing"), there is in fact only one LO during the whole interval $[t_k, t_{k+1})$ and its value is

$$VO_3^{\pm}(k) := (r_{t_k}^{\pm} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\widetilde{\mathbb{1}}_{fill_k^{\pm}} \tilde{c}_{t_k}^{\pm} (\tilde{p}_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm}]. \tag{2.23}$$

In the case that $L_{t_k}^{\pm} = \infty$ (cancelling any outstanding LO and not submitting any new LO), we have only one LO during $[t_k, t_{k+})$ and its value is the same as $VO_1^{\pm}(k)$.

We provide an explanation of equations (2.21)-(2.23) for the ask side, noting that the bid side follows a similar reasoning. Eq. (2.21) computes the value of an outstanding LO during the latency period. The execution price of this ask order is determined by its limit price, $S_{t_k} + r_{t_k}^{+}$, which relative to the fundamental price $S_{t_{k+}}$ at time t_{k+}, takes the form $S_{t_k} + r_{t_k}^{+} - S_{t_{k+}} = r_{t_k}^{+} - \Delta S_{t_k}$. Multiplying this quantity by the number of filled shares, as defined in Eq. (2.9), yields the expression in Eq. (2.21). Eq. (2.22) refers to value of LO for the non-latency period when $L_{t_k}^{\pm} \in \mathbb{Z}_+$. Order value can assume two distinct values, depending on the price change, ΔS_{t_k}, experienced during the latency period. If the price change during latency is less than the ask spread $L_{t_k}^{+}$, the ask LO enters the book with an execution price of $S_{t_k} + L_{t_k}^{+}$, which relative to the fundamental price at time t_{k+1}, $S_{t_{k+1}}$, is equal to $S_{t_k} + L_{t_k}^{+} - S_{t_{k+1}} = L_{t_k}^{+} - \Delta S_{t_k} - \Delta S_{t_{k+}}$, which multiplied by the

number of filled shares, results in the first term of (2.22). However, if the price change during the latency period is greater than or equal to the ask spread $L^+_{t_k}$, the ask order is executed immediately after being registered in the LOB at a price of $S_{t_{k+}}$, with the full quantity $Q^+_{\max}$, yielding the second term in Eq. (2.22). When $L^+_{t_k} = o$, the market maker takes no action, and the maximum number of filled shares during the whole period $[t_k, t_{k+1})$ is constrained by $Q^+_{out_{t_k}}$ and the execution price relative to $S_{t_{k+1}}$ is given by $S_{t_k} + r^+_{t_k} - S_{t_{k+1}} = r^+_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}$. Multiplying these two quantities results in the expression (2.23).

The expectations of the above intrinsic values will play an important role in the formulation of the problem as a MDP. Specifically, with the state variables

$$X_{t_k} := (I_{t_k}, Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k}), \quad k = 0, 1, \ldots, N, \tag{2.24}$$

for any $x = (i, q^+, q^-, r^+, r^-) \in \mathcal{X}$, where

$$\mathcal{X} := \big\{(i, q^+, q^-, r^+, r^-) : i \in \mathbb{Z}, \underline{I} \le i \le \overline{I}, (q^+, q^-) \in \mathbb{Z}^2_+, 0 \le q^\pm \le Q^\pm_{\max}, (r^+, r^-) \in \overline{\mathbb{Z}}^2_+\big\}, \tag{2.25}$$

we define

$$H^\pm_{t_k}(\Delta\tau, r^\pm, q^\pm) := \mathbb{E}\big[VO^\pm_1(k) \mid X_{t_k} = x\big], \tag{2.26}$$

$$H^\pm_{t_{k+}}(\Delta\tau, \Delta t, l^\pm) := \mathbb{E}\big[VO^\pm_2(k) \mid X_{t_k} = x\big], \tag{2.27}$$

$$\widetilde{H}^\pm_{t_k}(\Delta t, r^\pm, q^\pm) := \mathbb{E}\big[VO^\pm_3(k) \mid X_{t_k} = x\big]. \tag{2.28}$$

The following proposition shows that conditioning on X_{t_k} is equivalent to conditioning on the entire filtration $\mathcal{F}_{t_k}$.

Proposition 2.2.1. *Under Assumptions 2.1.1 and 2.1.2, for any* $L_{t_k}^{\pm} \in \sigma(X_{t_k})$ *and* $k = 0, 1, \ldots, N$*, we have*

$$\mathbb{E}[VO_1^{\pm}(k) \mid \mathcal{F}_{t_k}] = \mathbb{E}[VO_1^{\pm}(k) \mid X_{t_k}] = H_{t_k}^{\pm}(\Delta\tau, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}), \tag{2.29}$$

$$\mathbb{E}[VO_2^{\pm}(k) \mid \mathcal{F}_{t_k}] = \mathbb{E}[VO_2^{\pm}(k) \mid X_{t_k}] = H_{t_{k+}}^{\pm}(\Delta\tau, \Delta t, L_{t_k}^{\pm}), \tag{2.30}$$

$$\mathbb{E}[VO_3^{\pm}(k) \mid \mathcal{F}_{t_k}] = \mathbb{E}[VO_3^{\pm}(k) \mid X_{t_k}] = \widetilde{H}_{t_k}^{\pm}(\Delta t, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}). \tag{2.31}$$

Furthermore, the following explicit formulas hold:

$$\begin{aligned}
H_{t_k}^{\pm}(\Delta\tau, r, q) &= \pi_{t_k}^{\pm} h_{\Delta\tau}^{\pm}(1, r, q), \\
H_{t_{k+}}^{\pm}(\Delta\tau, \Delta t, l) &= \pi_{t_{k+}}^{\pm} \check{h}_{\Delta\tau,\Delta t}^{\pm}(1, l, Q_{\max}^{\pm}) \mathbb{E}[(l \mp S_{\Delta\tau})_+], \\
\widetilde{H}_{t_k}^{\pm}(\Delta t, r, q) &= \tilde{\pi}_{t_k}^{\pm} \tilde{h}_{\Delta t}^{\pm}(1, r, q),
\end{aligned}$$

where, using the cdfs in Eq. (2.16), for $j \in \{0, 1\}, r \in \mathbb{R}_+, q \in \mathbb{Z}_+$*,*

$$\begin{aligned}
h_{\Delta\tau}^{\pm}(j, r, q) &= \int [x(y - r)_+ \wedge q] \, F_{\Delta\tau}^{\pm}(dx, dy, j), \\
\check{h}_{\Delta\tau,\Delta t}^{\pm}(j, r, q) &= \int [x(y - r)_+ \wedge q] \, \check{F}_{\Delta\tau,\Delta t}^{\pm}(dx, dy, j), \\
\tilde{h}_{\Delta t}^{\pm}(j, r, q) &= \int [x(y - r)_+ \wedge q] \, \tilde{F}_{\Delta t}^{\pm}(dx, dy, j).
\end{aligned} \tag{2.32}$$

2.2.2 Value Functions

The following result shows that the value function defined in Eq. (2.19) corresponds to that of a MDP with a one-period reward during $[t_k, t_{k+1})$ given by

$$H^{\pm}_{act_{t_k}}(r^{\pm}, l^{\pm}, q^{\pm}) := \begin{cases} H^{\pm}_{t_k}(\Delta\tau, r^{\pm}, q^{\pm}) + H^{\pm}_{t_{k+}}(\Delta\tau, \Delta t, l^{\pm}), & \text{if } l^{\pm} \in \mathbb{Z}_+, \\ \widetilde{H}^{\pm}_{t_k}(\Delta t, r^{\pm}, q^{\pm}), & \text{if } l^{\pm} = o, \\ H^{\pm}_{t_k}(\Delta\tau, r^{\pm}, q^{\pm}), & \text{if } l^{\pm} = \infty. \end{cases} \tag{2.33}$$

This total value serves as a measure of the market maker's performance and guides the decision-making process by quantifying the immediate rewards associated with each possible action at time t_k.

Theorem 2.2.2. *Let $X_{t_k} := (I_{t_k}, Q^{+}_{out_{t_k}}, Q^{-}_{out_{t_k}}, r^{+}_{t_k}, r^{-}_{t_k})$ with dynamics determined by Eqs. (2.5), (2.13), and (2.14). Then, under Assumptions 2.1.1 and 2.1.2, we have*

$$V_{t_{N+}} = W_{t_{N+}} + S_{t_{N+}} I_{t_{N+}} - \lambda I^2_{t_{N+}},$$

and, for $k = 0, 1, 2, \ldots, N$,

$$V_{t_k} = W_{t_k} + S_{t_k} I_{t_k} + g_k(X_{t_k}), \tag{2.34}$$

where, for $x = (i, q^{+}, q^{-}, r^{+}, r^{-}) \in \mathcal{X}$, as defined in Eq. (2.25),

$$g_k(x) := \begin{cases} H^+_{t_N}\left(\Delta\tau, r^+, q^+\right) + H^-_{t_N}\left(\Delta\tau, r^-, q^-\right) \\ \qquad - \lambda\mathbb{E}\big[(i - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}})^2 \mid X_{t_N} = x\big], \quad k = N, \\ \\ \max\limits_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} G_k(x, l^+, l^-), \qquad\qquad k = 0, 1, \ldots, N-1, \end{cases} \tag{2.35}$$

and

$$\begin{aligned} G_k(x, l^+, l^-) = \mathbb{E}\big[g_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x, L^+_{t_k} = l^+, L^-_{t_k} = l^-\big] \\ + H^+_{act_{t_k}}(r^+, l^+, q^+) + H^-_{act_{t_k}}(r^-, l^-, q^-). \end{aligned} \tag{2.36}$$

In particular, the optimal strategies are given by

$$(L^{+,*}_{t_k}, L^{-,*}_{t_k}) = \underset{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}}{\arg\max} G_k(X_{t_k}, l^+, l^-). \tag{2.37}$$

Theorem 2.2.2 provides a reduction in the computational complexity of the value function by reducing the number of state variables from seven to five, specifically (i, q^+, q^-, r^+, r^-), in the backward recursion outlined in (2.35). By examining Eq. (2.34), we observe that the value function V_{t_k} can be decomposed into three distinct components:

1. W_{t_k} represents the current wealth of the market maker,
2. $S_{t_k} I_{t_k}$ denotes the value of the inventory marked to the market at the mid-price,
3. $g_k(X_{t_k})$ captures the additional value resulting from following the optimal strategy, which is contingent upon the market maker's outstanding orders and inventory.

The backward recursion and maximization problem presented in Eq. (2.35) and (2.36) establish a trade-off between the value of orders, encompassing both outstanding orders

and newly placed orders, in the current period, and the anticipated extra value g_{k+1} in the subsequent period.

Remark 2.2.1. *In Theorem 2.2.2, we don't impose any constrains on the set of admissible placements A_k, which, in particular, may result in inventories I_{t_k} exceeding the boundaries $(\underline{I}, \overline{I})$. However, we can restrict A_k to guarantee the desired boundaries. Recall that*

$$I_{t_{k+1}} = I_{t_k} - Q_{fill^+_{t_k}} + Q_{fill^-_{t_k}} - Q_{fill^+_{t_{k+}}} + Q_{fill^+_{t_{k+}}},$$

and $Q_{fill^\pm_{t_k}}, Q_{fill^\pm_{t_{k+}}} \leq Q^\pm_{\max}$. Then, if $I_{t_k} > \overline{I} - Q^+_{\max}$, we fix $L^-_{t_k} = \infty$ (no more buying) and $I_{t_k} < \underline{I} + Q^-_{\max}$, we fix $L^+_{t_k} = \infty$ (no more selling). Then, in (2.35), we take $(l^+, l^-) \in \mathcal{A}(x)$, where

$$\begin{aligned}\mathcal{A}(x) = \{(l^+, l^-) : l^\pm \in \mathbb{Z}_+ \cup \{\infty\} \cup \{o\} \text{ if } i \in [\underline{I} + Q^-_{\max}, \overline{I} - Q^+_{\max}],\\ l^+ = \infty \text{ if } i < \underline{I} + Q^-_{\max}, \text{ and } l^- = \infty \text{ if } i > \overline{I} - Q^+_{\max}\}.\end{aligned}$$

2.3 Variations in the Market Making Model: Special Cases Analysis

For ease of simplicity, we introduce some notation related to $(c^\pm_{t_k}, c^\pm_{t_{k+}}, p^\pm_{t_k}, p^\pm_{t_{k+}}, \tilde{c}^\pm_{t_k}, \tilde{p}^\pm_{t_k})$. For $a, b \in \{0, 1, 2\}$,

$$\begin{aligned}\mu^\pm_{c^a p^b} &:= \mathbb{E}\big[(c^\pm_{t_k})^a (p^\pm_{t_k})^b \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill^\pm_k} = 1\big],\\ \mu^\pm_{c^a p^b_+} &:= \mathbb{E}\big[(c^\pm_{t_{k+}})^a (p^\pm_{t_{k+}})^b \mid \mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^\pm_{k+}} = 1\big],\\ \tilde{\mu}^\pm_{c^a p^b} &:= \mathbb{E}\big[(\tilde{c}^\pm_{t_k})^a (\tilde{p}^\pm_{t_k})^b \mid \mathcal{F}_{t_k}, \tilde{\mathbb{1}}_{fill^\pm_k} = 1\big].\end{aligned} \tag{2.38}$$

2.3.1 Simplified Model with Latency

In this section, a simplified model is presented, which makes the assumption that $p^{\pm} - r^{\pm}$ and $p^{\pm} - l^{\pm}$ are always positive, and that the number of filled shares will not exceed the bounds $Q^{\pm}_{\max}$ or $Q^{\pm}_{out}$.

Under these assumptions, the value of LO given in Eq. (2.21) can be written as

$$VO^{\pm}_1(k) := (r^{\pm}_{t_k} \mp \Delta S_{t_k}) \cdot \mathbb{1}_{fill^{\pm}_k} c^{\pm}_{t_k} (p^{\pm}_{t_k} - r^{\pm}_{t_k}), \tag{2.39}$$

while the value of the second order defined in Eq. (2.22) becomes

$$\begin{aligned} VO^{\pm}_2(k) := {} & \mathbb{1}_{\{\pm \Delta S_{t_k} < L^{\pm}_{t_k}\}} \cdot (L^{\pm}_{t_k} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})] \\ & + \mathbb{1}_{\{\pm \Delta S_{t_k} \geq L^{\pm}_{t_k}\}} \cdot (\mp \Delta S_{t_{k+}}) \cdot Q^{\pm}_{\max}. \end{aligned} \tag{2.40}$$

When $L^{\pm}_{t_k} = o$, Eq. (2.23) simplifies to

$$VO^{\pm}_3(k) := (r^{\pm}_{t_k} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot \widetilde{\mathbb{1}}_{fill^{\pm}_k} \tilde{c}^{\pm}_{t_k} (\tilde{p}^{\pm}_{t_k} - r^{\pm}_{t_k}). \tag{2.41}$$

In the case that $L^{\pm}_{t_k} = \infty$, the LO value is the same as $VO^{\pm}_1(k)$ given in Eq. (2.39). $H^{\pm}_{t_k}(\Delta\tau, r^{\pm}, q^{\pm})$, $H^{\pm}_{t_{k+}}(\Delta\tau, \Delta t, l^{\pm})$ and $\widetilde{H}^{\pm}_{t_k}(\Delta t, r^{\pm}, q^{\pm})$ are defined similarly as Eqs. (2.26)-(2.28).

Proposition 2.3.1. *Under Assumptions 2.1.1 and 2.1.2, and the simplified conditions mentioned above, for $k = 0, 1, \ldots, N$,*

$$H^{\pm}_{t_k}(\Delta\tau, r) = \pi^{\pm}_{t_k} r^{\pm}(\mu^{\pm}_{cp} - \mu^{\pm}_{c} r^{\pm}), \tag{2.42}$$

$$H^{\pm}_{t_{k+}}(\Delta\tau, \Delta t, l) = \pi^{\pm}_{t_{k+}}(\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} l^{\pm})\mathbb{E}[(l^{\pm} \mp S_{\Delta\tau})_+], \tag{2.43}$$

$$\widetilde{H}^{\pm}_{t_k}(\Delta t, r) = \tilde{\pi}^{\pm}_{t_k} r^{\pm}(\tilde{\mu}^{\pm}_{cp} - \tilde{\mu}^{\pm}_{c} r^{\pm}). \tag{2.44}$$

Without $Q^{\pm}_{out}$ in state space, we introduce the notation X' to refer to the tuple (I, r^+, r^-). Specifically, for $k = 0, 1, ..., N$, we define

$$X'_{t_k} := (I_{t_k}, r^+_{t_k}, r^-_{t_k}).$$

Remark 2.3.1. *Similar to Proposition 2.2.1, we can show that for the simplified model defined above, conditioning on X'_{t_k} is same as conditioning on $\mathcal{F}_{t_k}$.*

We can compute an approximate optimal placing strategy using Theorem 2.2.2. First, we compute $g_N(x')$, for x' in a constrained finite domain Ψ'. Recall Eq. (2.35) and the

simplified model assumptions mentioned above, we have

$$
\begin{aligned}
g_N(x') &= H^+_{t_N}(\Delta\tau, r^+) + H^-_{t_N}(\Delta\tau, r^-) \\
&\quad - \lambda\mathbb{E}\big[(i - \mathbb{1}_{fill^+_N} \cdot c^+_{t_N}(p^+_{t_N} - r^+) + \mathbb{1}_{fill^-_N} \cdot c^-_{t_N}(p^-_{t_N} - r^-))^2 \mid X'_{t_N} = x'\big] \\
&= H^+_{t_N}(\Delta\tau, r^+) + H^-_{t_N}(\Delta\tau, r^-) \\
&\quad - \lambda\mathbb{E}\big[i^2 + \mathbb{1}_{fill^+_N} \cdot (c^+_{t_N})^2(p^+_{t_N} - r^+)^2 + \mathbb{1}_{fill^-_N} \cdot (c^-_{t_N})^2(p^-_{t_N} - r^-)^2 \\
&\qquad\quad - 2i \cdot \mathbb{1}_{fill^+_N} \cdot (c^+_{t_N} p^+_{t_N} - c^+_{t_N} r^+) + 2i \cdot \mathbb{1}_{fill^-_N} \cdot (c^-_{t_N} p^-_{t_N} - c^-_{t_N} r^-) \\
&\qquad\quad - 2 \cdot \mathbb{1}_{fill^+_N}(c^+_{t_N} p^+_{t_N} - c^+_{t_N} r^+) \cdot \mathbb{1}_{fill^-_N}(c^-_{t_N} p^-_{t_N} - c^-_{t_N} r^-) \mid X'_{t_N} = x'\big] \\
&= r^+\pi^+_{t_N}(\mu^+_{cp} - \mu^+_c r^+) + r^-\pi^-_{t_N}(\mu^-_{cp} - \mu^-_c r^-) \\
&\quad - \lambda\big[i^2 + \pi^+_{t_N}\mu^+_{c^2}(\mu^+_{p^2} - 2\mu^+_p r^+ + (r^+)^2) + \pi^-_{t_N}\mu^-_{c^2}(\mu^-_{p^2} - 2\mu^-_p r^- + (r^-)^2) \\
&\qquad - 2i\pi^+_{t_N}(\mu^+_{cp} - \mu^+_c r^+) + 2i\pi^-_{t_N}(\mu^-_{cp} - \mu^-_c r^-) \\
&\qquad - 2 \cdot \pi^+_{t_N}(\mu^+_{cp} - \mu^+_c r^+) \cdot \pi^-_{t_N}(\mu^-_{cp} - \mu^-_c r^-)\big].
\end{aligned}
$$

For the recursive step in Eq. (2.35), for $x' \in \Psi'$, we have

$$
\begin{aligned}
g_{N-1}(x') &= \max_{(l,l)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} G_{N-1}(x', l^+, l^-) \\
&= \max_{(l,l)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} \big\{\mathbb{E}\big[g_N(X'_{t_N}) \mid X'_{t_{N-1}} = x', L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-\big] \\
&\qquad\qquad + H^+_{act_{t_{N-1}}}(r^+, l^+) + H^-_{act_{t_{N-1}}}(r^-, l^-)\big\},
\end{aligned}
$$

where

$$
H^\pm_{act_{t_{N-1}}}(r^\pm, l^\pm) = \begin{cases} H^\pm_{t_{N-1}}(\Delta\tau, r^\pm) + H^\pm_{t_{N-1}}(\Delta\tau, \Delta t, l^\pm), & \text{if } l^\pm \in \mathbb{Z}_+, \\ \widetilde{H}^\pm_{t_{N-1}}(\Delta t, r^\pm), & \text{if } l^\pm = o, \\ H^\pm_{t_{N-1}}(\Delta\tau, r^\pm), & \text{if } l^\pm = \infty. \end{cases} \tag{2.45}
$$

2.3.2 Market Making without Latency

This section is devoted to a specific scenario within our proposed model, where the market making activity takes place without any latency. This scenario assumes that the market maker has immediate access to all available information and can execute trades without any delay. We start by presenting the system dynamics, a structure similar to that described in Section 2.1.2.

System Dynamics

Now, we present the system dynamics $(W, S, I, Q^+_{out}, Q^-_{out}, r^+, r^-)$ from time t_k to t_{k+1} for $k = 0, 1, \cdots, N$, under the assumption of zero latency. To allow for a slight abuse of notation, let $\Delta S_{t_{k+}} := S_{t_{k+1}} - S_{t_k}$ denote the price change on the interval $[t_k, t_{k+1})$. It is clear that the fundamental price follows the relationship

$$S_{t_{k+1}} = S_{t_k} + \Delta S_{t_{k+}}. \tag{2.46}$$

The cash holding and inventory processes are defined as follows:

$$W_{t_{k+1}} = W_{t_k} + P_{fill^+_{t_{k+}}} Q_{fill^+_{t_{k+}}} - P_{fill^-_{t_{k+}}} Q_{fill^-_{t_{k+}}}, \tag{2.47}$$

$$I_{t_{k+1}} = I_{t_k} - Q_{fill^+_{t_{k+}}} + Q_{fill^+_{t_{k+}}}, \tag{2.48}$$

where $P_{fill^+_{t_{k+}}}$ and $Q_{fill^+_{t_{k+}}}$ respectively denote the execution price and the number of filled shares of ask LOs during $[t_k, t_{k+1})$. Analogously, $P_{fill^-_{k+}}$, $Q_{fill^-_{k+}}$ are the corresponding

quantities on the bid side. These quantities follow

$$P_{fill^{\pm}_{t_{k+}}} = \begin{cases} S_{t_k} \pm L^{\pm}_{t_k}, & \text{if } L^{\pm}_{t_k} \in \mathbb{Z}_+, \\ S_{t_k} \pm r^{\pm}_{t_k}, & \text{if } L^{\pm}_{t_k} = o, \\ 0, & \text{if } L^{\pm}_{t_k} = \infty, \end{cases} \tag{2.49}$$

$$Q_{fill^{\pm}_{t_{k+}}} = \begin{cases} \mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})_+ \wedge Q^{\pm}_{\max}, & \text{if } L^{\pm}_{t_k} \in \overline{\mathbb{Z}}_+, \\ \mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - r^{\pm}_{t_k})_+ \wedge Q^{\pm}_{out_{t_k}}, & \text{if } L^{\pm}_{t_k} = o, \end{cases} \tag{2.50}$$

where

$$\mathbb{1}_{fill^{+}_{k+}} = \mathbb{1}_{\{\text{At least one buy MO arrives during } [t_k, t_{k+1})\}},$$
$$\mathbb{1}_{fill^{-}_{k+}} = \mathbb{1}_{\{\text{At least one sell MO arrives during } [t_k, t_{k+1})\}}.$$

In a manner similar to Section 2.2, we make the assumption that $\mathbb{1}_{fill^{\pm}_{k+}} \in \mathcal{F}_{t_{k+1}}$, and we introduce the notation $\pi^{\pm}_{t_{k+}} := \mathbb{P}(\mathbb{1}_{fill^{\pm}_{k+}} = 1 \mid \mathcal{F}_{t_k})$. Additionally, we define the random variables $c^{\pm}_{t_{k+}}$ and $p^{\pm}_{t_{k+}}$ in a manner similar to Section 2.2, but now specifically within the interval $[t_k, t_{k+1})$.

Also, the number of outstanding shares for ask and bid orders at time t_{k+1} follow the relation

$$Q^{\pm}_{out_{t_{k+1}}} = \begin{cases} Q^{\pm}_{\max} - Q_{fill^{\pm}_{t_{k+}}}, & \text{if } L^{\pm}_{t_k} \in \mathbb{Z}_+, \\ Q^{\pm}_{out_{t_k}} - Q_{fill^{\pm}_{t_{k+}}}, & \text{if } L^{\pm}_{t_k} = o, \\ 0, & \text{if } L^{\pm}_{t_k} = \infty. \end{cases} \tag{2.51}$$

Finally, the ask and bid outstanding quote pair is determined as

$$r^{\pm}_{t_{k+1}} = \begin{cases} L^{\pm}_{t_k} \mp \Delta S_{t_{k+}}, & \text{if } L^{\pm}_{t_k} \in \overline{\mathbb{Z}}_+, \\ r^{\pm}_{t_k} \mp \Delta S_{t_{k+}}, & \text{if } L^{\pm}_{t_k} = o. \end{cases} \tag{2.52}$$

We make analogous assumptions about $c^{\pm}_{\cdot}$ and $p^{\pm}_{\cdot}$ to those presented in Assumption 2.1.1. Note that Assumption 2.3.1, although sharing some notation with a part of Assumption 2.1.1, holds distinct implications in this context.

Assumption 2.3.1. *For $k = 0, 1, \ldots, N$, we have*

1. *$(c^{\pm}_{t_{k+}}, p^{\pm}_{t_{k+}})$ are $\mathcal{F}_{t_{k+1}}$-measurable;*
2. *The conditional distribution of $(c^{\pm}_{t_{k+}}, p^{\pm}_{t_{k+}})$ given $(\mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{\pm}_{k+}})$ is a function of only the indicator $\mathbb{1}_{fill^{\pm}_{k+}}$ that does not depend on k;*
3. *$(c^{+}_{t_{k+}}, p^{+}_{t_{k+}})$ and $(c^{-}_{t_{k+}}, p^{-}_{t_{k+}})$ are conditionally independent given $(\mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{+}_{k+}}, \mathbb{1}_{fill^{-}_{k+}})$.*

Furthermore, we introduce assumptions about the fundamental price, mirroring the structure of Assumption 2.1.2.

Assumption 2.3.2. *For $k = 0, 1, \ldots, N$, we have*

1. *The fundamental price $S_{t_k} \in \mathcal{F}_{t_k}$ is a martingale:*

$$\mathbb{E}[S_{t_{k+1}} \mid \mathcal{F}_{t_k}] = S_{t_k};$$

2. *$S_{t_{k+}} - S_{t_k}$ and $(\mathbb{1}_{fill^{\pm}_{k+}}, c^{\pm}_{t_{k+}}, p^{\pm}_{t_{k+}})$ are conditional independent given $\mathcal{F}_{t_k}$.*

Optimal Market Making Strategy

Given the outlined system dynamics, we will now proceed to reframe the optimal market making problem without latency as a MDP. Same as Section 2.2, the market maker takes an admissible action at the beginning of each time interval aiming to maximize her expected P&L at the termination of the trading interval. It is worth noting that, due to the absence of latency, the final action of unwinding using MOs occurs at the terminal time t_{N+1}, which differs from the case with latency where the final unwinding inventory action occurs at time t_{N+}.

Denote the market maker's terminal P&L as $\mathcal{W}$. Assuming at terminal time t_{N+1}, the state $\Phi_{t_{N+1}}$ is $(W_{t_{N+1}}, S_{t_{N+1}}, I_{t_{N+1}}, Q^{+}_{out_{t_{N+1}}}, Q^{-}_{out_{t_{N+1}}}, r^{+}_{t_{N+1}}, r^{-}_{t_{N+1}})$. Then, we have

$$\mathcal{W} := W_{t_{N+1}} + S_{t_{N+1}} I_{t_{N+1}} - \lambda I^2_{t_{N+1}}, \tag{2.53}$$

where $\lambda > 0$ is a inventory penalization constant.

Following the notation introduced in Equation (2.24), we can now present the market maker's optimization problem. Recall from Eq. (2.18) that

$$V_{t_0} := \sup_{(L^{\pm}_{t_0},\dots,L^{\pm}_{t_N})\in\mathcal{A}_{0,N}} \mathbb{E}\left[\mathcal{W} \mid \mathcal{F}_{t_0}\right],$$

and from Eq. (2.19), for $k = 1, 2, \dots, N$, we have

$$V_{t_k} := \sup_{(L^{\pm}_{t_k},\dots,L^{\pm}_{t_N})\in\mathcal{A}_{k,N}} \mathbb{E}\left[\mathcal{W} \mid \mathcal{F}_{t_k}\right].$$

Now, in the absence of latency, for $k = N + 1$, we have

$$V_{t_{N+1}} := W_{t_{N+1}} + S_{t_{N+1}} I_{t_{N+1}} - \lambda I^2_{t_{N+1}}.$$

With standard arguments, bellman equation for above value functions is given as follows:

$$V_{t_k} = \begin{cases} W_{t_{N+1}} + S_{t_{N+1}} I_{t_{N+1}} - \lambda I^2_{t_{N+1}}, & k = N+1, \\ \sup\limits_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\left[V_{t_{k+1}} \mid \mathcal{F}_{t_k} \right], & k = 0, 1, \ldots, N, \end{cases} \tag{2.54}$$

where above we say $(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k$ if $L^\pm_{t_k} \in \sigma(\Phi_{t_k})$ and $L^\pm_{t_k} \in \mathbb{Z}_+ \cup \{\infty\} \cup \{o\}$.

Value of the Order

Now, we expand the concept of the order value, as defined in Section 2.2.1, to the current scenario without latency. In the absence of latency, the expression for $H^\pm_{t_k}(\Delta\tau, r^\pm, q^\pm)$ given in Eq. (2.26) is not applicable, and Eq. (2.27) can be simplified. Note that when there is no latency, Eq. (2.22) simplifies to

$$VO^\pm_2(k) = (L^\pm_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\mathbb{1}_{fill^\pm_{k+}} c^\pm_{t_{k+}} (p^\pm_{t_{k+}} - L^\pm_{t_k})_+ \wedge Q^\pm_{\max}], \tag{2.55}$$

and Eq. (2.23) becomes

$$VO^\pm_3(k) = (r^\pm_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\widetilde{\mathbb{1}}_{fill^\pm_k} \tilde{c}^\pm_{t_k} (\tilde{p}^\pm_{t_k} - r^\pm_{t_k})_+ \wedge Q^\pm_{out_{t_k}}]. \tag{2.56}$$

Then, the definitions of $H^\pm_{t_{k+}}(\Delta t, l^\pm)$ and $\widetilde{H}^\pm_{t_k}(\Delta t, r^\pm, q^\pm)$ mirror those introduced in Eq. (2.27)-(2.28). $H^\pm_{act_{t_k}}(r^\pm, l^\pm, q^\pm)$ are also defined in the same manner as in Eq. (2.33).

Similar to the Theorem 2.2.2, we have the following result.

Theorem 2.3.2. *Let* $X_{t_k} = (I_{t_k}, Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k})$ *with dynamics determined by Eqs. (2.48), (2.51), and (2.52). Then under Assumptions 2.3.1 and 2.3.2, we have*

$$V_{t_{N+1}} = W_{t_{N+1}} + S_{t_{N+1}} I_{t_{N+1}} - \lambda I^2_{t_{N+1}}$$

and, for $k = 0, 1, 2, \ldots, N$,

$$V_{t_k} = W_{t_k} + S_{t_k} I_{t_k} + g_k(X_{t_k}),$$

where, for $x = (i, q^+, q^-, r^+, r^-) \in \mathcal{X}$, *as defined in Eq. (2.25),*

$$g_k(x) := \begin{cases} \max\limits_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} \big\{ H^+_{act_{t_k}}(r^+, l^+, q^+) + H^-_{act_{t_k}}(r^-, l^-, q^-) \\ \qquad - \lambda \mathbb{E}\big[(i - Q_{fill^+_{t_{N+}}} + Q_{fill^-_{t_{N+}}})^2 \mid X_{t_N} = x, L^+_{t_k} = l^+, L^-_{t_k} = l^-\big]\big\}, \quad k = N, \\ \max\limits_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} G_k(x, l^+, l^-), \qquad\qquad k = 0, \ldots, N-1, \end{cases} \tag{2.57}$$

and

$$\begin{aligned} G_k(x, l^+, l^-) = \mathbb{E}\big[g_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x, L^+_{t_k} = l^+, L^-_{t_k} = l^-\big] \\ + H^+_{act_{t_k}}(r^+, l^+, q^+) + H^-_{act_{t_k}}(r^-, l^-, q^-). \end{aligned}$$

In particular, the optimal strategies are given by

$$(L^{+,*}_{t_k}, L^{-,*}_{t_k}) = \operatorname*{arg\,max}_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} G_k(X_{t_k}, l^+, l^-).$$

2.3.3 Simplified Market Making without Latency

In this section, we consider a setting where the market maker is not allowed to take the actions of "doing nothing" or cancellation. Specifically, the market maker places ask and bid LOs at the beginning of every second, which is consistent with the setup in [12]. Under this arrangement, there are no outstanding orders or spreads to consider. Hence, the state variable is just I_{t_k} itself, for $k = 0, 1, \ldots, N$.

Using Theorem 2.3.2, we can compute an approximate optimal strategy. We start with the calculation of $g_N(i)$, for $i \in [\underline{I}, \overline{I}]$,

$$\begin{aligned}
g_N(i) = \max_{(l^+,l^-)\in\mathcal{A}} \big\{ & H^+_{act}(l^+) + H^-_{act}(l^-) \\
& - \lambda\mathbb{E}\big[i^2 + \mathbb{1}_{fill^+_{N+}} \cdot (c^+_{t_{N+}})^2(p^+_{t_{N+}} - l^+)^2 + \mathbb{1}_{fill^-_{N+}} \cdot (c^-_{t_{N+}})^2(p^-_{t_{N+}} - l^-)^2 \\
& - 2i \cdot \mathbb{1}_{fill^+_{N+}} \cdot (c^+_{t_{N+}} p^+_{t_{N+}} - c^+_{t_{N+}} l^+) + 2i \cdot \mathbb{1}_{fill^-_{N+}} \cdot (c^-_{t_{N+}} p^-_{t_{N+}} - c^-_{t_{N+}} l^-) \\
& - 2 \cdot \mathbb{1}_{fill^+_{N+}}(c^+_{t_{N+}} p^+_{t_{N+}} - c^+_{t_{N+}} l^+)\mathbb{1}_{fill^-_{N+}}(c^-_{t_{N+}} p^-_{t_{N+}} - c^-_{t_{N+}} l^-) \mid I_{t_N} = i\big]\big\} \\
= \max_{(l^+,l^-)\in\mathcal{A}} \big\{ & l^+\pi^+_{t_{N+}}(\mu^+_{cp_+} - \mu^+_{c_+} l^+) + l^-\pi^-_{t_{N+}}(\mu^-_{cp_+} - \mu^-_{c_+} l^-) \\
& - \lambda\big[i^2 + \pi^+_{t_{N+}}\mu^+_{c^2_+}(\mu^+_{p^2_+} - 2\mu^+_{p_+} l^+ + (l^+)^2) \\
& + \pi^-_{t_{N+}}\mu^-_{c^2_+}(\mu^-_{p^2_+} - 2\mu^-_{p_+} l^- + (l^-)^2) - 2i\pi^+_{t_{N+}}(\mu^+_{cp_+} - \mu^+_{c_+} l^+) \\
& + 2i\pi^-_{t_{N+}}(\mu^-_{cp_+} - \mu^-_{c_+} l^-) - 2 \cdot \pi^+_{t_{N+}}(\mu^+_{cp_+} - \mu^+_{c_+} l^+) \cdot \pi^-_{t_{N+}}(\mu^-_{cp_+} - \mu^-_{c_+} l^-)\big]\big\}.
\end{aligned}$$

For the recursive step in Eq. (2.57), we have

$$g_{N-1}(i) = \max_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} \big\{ H^+_{act}(l^+) + H^-_{act}(l^-) + \mathbb{E}\big[g_N(I_{t_N}) \mid I_{t_{N-1}} = i, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-\big]\big\},$$

where

$$H^\pm_{act}(l^\pm) = H^\pm_{t_{N-1}}(\Delta t, l^\pm).$$

2.4 Implementation and Empirical Performance

In this section, we empirically evaluate the effectiveness of the optimal order placement strategy formulated based on Theorem 2.2.2. This evaluation is conducted using authentic transaction data sourced from the stock market and corresponding LOB records. We utilized LOB data of three prominent companies, namely Apple Inc. (AAPL), Amazon.com Inc. (AMZN), and Microsoft Corporation (MSFT), over the course of the year 2019. The dataset encompasses 252 trading days, during which we recorded the activity of the LOB from 10am to 4pm each day. Our data is composed of two principal components: books and orders. The books component provides us with a comprehensive snapshot of the top 20 best ask and bid prices in the LOB along with their corresponding volume for each timestamp during the trading hours from 10am to 4pm each day, where a timestamp denotes every time that there is a change in any of the top levels at either side of the book. The orders component, on the other hand, records all actions submitted by market participants in LOB, including LOs, MOs, and cancellations, throughout the trading day.

We assume that the mid-price $\{S_t\}_{t\geq 0}$ follows a compound Poisson process, following

$$S_t = S_0 + \delta \sum_{i=0}^{\mathcal{N}_t} D_i, \tag{2.58}$$

where $\{\mathcal{N}_t\}_{t\geq 0}$ is a homogeneous Poisson process with price change rate ν and δ is the tick size. Additionally, $(D_i)_{i=1,2,\dots}$ denote a sequence of independent and identically distributed random variables. These random variables take on values of either +1 or -1, each with a probability of 0.5.

Table 2.1 presents the fundamental parameter settings for our implementation and performance analysis. Note that the values in Upper/Lower Inventory Boundaries, Outstanding/New Order Size are presented in units of hundred shares. Also, integer values in Outstanding/New Order Spread are presented in units of 1 tick.

Table 2.1: Default values of the parameters in simulation in Section 2.4.

Parameter	**Value(s)**
Upper Inventory Boundary $\overline{I}$	11
Lower Inventory Boundary $\underline{I}$	-11
Outstanding Order Spread $r^{\pm}$	0, 1, 2, 3, ∞
Outstanding Order Size $Q^{\pm}_{out}$	0, 1, 2, 3, 4, 5
New Order Spread/Action $L^{\pm}$	0, 1, 2, 3, o("doing nothing"), ∞("cancellation")
New Order Size $Q^{\pm}_{\max}$	5
Tick Size δ	\$0.01

2.4.1 Parameter Estimation

Rate of Price Changes ν. To estimate the rate at which the price changes, we considered two steps. First, a preprocessing step, where we trimmed the data to exclude the first and last 30 minutes of each trading day in order to avoid the high volatility of the opening and closing periods. Second, we performed a computational step where we tracked the mid-price, defined as the average of the best ask price and best bid price, between the trading hours of 10:00am to 3:30pm. To estimate the rate of price changes, we counted the number of times the mid-price changed over the course of the trading day and divided this amount by the total number of seconds between 10:00:00am and 3:30:00pm. Table 2.2 provides the mean of these estimated daily mid-price change rates averaged over all the year.

Table 2.2: Average rate (per second) of price changes over 252 trading days in 2019.

Stock	**Upward Rate**	**Downward Rate**
AAPL	2.213405	2.215102
AMZN	1.264078	1.267465
MSFT	1.034977	1.034514

Arrival Probability of MOs. We discuss now the estimation of the arrival probability of MOs, focusing on the case of buy MOs during the latency period. The arrival probabilities for other periods can be computed similarly. We set the latency to 0.02 seconds, which is consistent with the setup employed in other studies. (see, e.g., [24]).

For computational simplicity, we make the assumption that the arrival probability of MOs remain constant across all time instances t_k. Consequently, we will henceforth denote $\pi^{\pm}_{t_k}$, $\pi^{\pm}_{t_{k+}}$ and $\tilde{\pi}^{\pm}_{t_k}$ as $\pi^{\pm}_{\Delta\tau}$, $\pi^{\pm}_{\Delta\tau,\Delta t}$ and $\tilde{\pi}^{\pm}_{\Delta t}$, respectively. To determine the arrival probability of at least one buy MO within the latency period for a given trading day, we tally the number of seconds from 10:00am to 3:30pm during which at least one buy MO appears within the first 0.02 seconds of each 1-second subinterval. This count is then divided by 19800, which is the total number of seconds between 10:00am and 3:30pm, and repeat this procedure for every trading day. The mean of these estimated probabilities over 252 trading days is reported in Table 2.3, along with the estimates of the arrival probabilities for other time periods.

Table 2.3: Average arrival probability of market orders over 252 trading days in 2019. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second.

Stock	$\widehat{\pi}^{+}_{\Delta\tau}$	$\widehat{\pi}^{-}_{\Delta\tau}$	$\widehat{\pi}^{+}_{\Delta\tau,\Delta t}$	$\widehat{\pi}^{-}_{\Delta\tau,\Delta t}$	$\widehat{\tilde{\pi}}^{+}_{\Delta t}$	$\widehat{\tilde{\pi}}^{-}_{\Delta t}$
AAPL	0.009769	0.009936	0.258041	0.262947	0.263407	0.268212
AMZN	0.003435	0.003514	0.097630	0.100175	0.099906	0.102493
MSFT	0.007425	0.007499	0.201744	0.204104	0.206397	0.208745

Demand Function. To estimate the demand function for the ask and bid sides, on both the latency and non-latency periods, within a 1-second sub-interval, we fit a weighted linear regression to the empirical demand function obtained from the data. The slope and intercept of the fitted line are then used to estimate c and p as described below.

For example, consider the ask side during the latency period. We begin by determining the demand during a given interval $[t_k, t_k + \Delta\tau)$ for different price levels $P_l \in \{0.005, 0.015, 0.025, \ldots\}$ relative to S_{t_k}. That is, for a given P_l, the demand will be the number of shares sold during $[t_k, t_k + \Delta\tau)$ if the market maker had a LO placed at level P_l, at time t_k. To this end, when a buy MO (MO_m) with a volume of V_{MO_m} enters the market at time $t_m \in [t_k, t_k + \Delta\tau)$, we first record the volume of existing ask LOs with prices lower than P_l, at that precise moment. This recorded volume is denoted as V_{LO_m}. Next, we compute $(V_{MO_m} - V_{LO_m}) \vee 0$, which is the number of shares from the market maker's LO that would be executed by this buy MO. This process is repeated for all buy MOs occurring within the interval $[t_k, t_k + \Delta\tau)$. The cumulative quantity, represented as $\sum_m \big((V_{MO_m} - V_{LO_m}) \vee 0\big)$, serves as a proxy of the actual demand at the price level P_l during this time frame. We use the same procedure to compute the demand function for the bid side, as well as during non-latency periods.

The above procedure enables us to estimate the actual demand within a $\Delta\tau$ subinterval for a given day. Subsequently, for each price level P_l, we average the estimated demands across all seconds during which at least one MO arrives on that particular day, and obtain the daily average demand $\overline{D}_l$. Then, we fit a linear model via weighted least squares estimation. Specifically, we employ the average demand $\overline{D}_l$ at each price level as the response variable, with the price level P_l serving as the predictor variable. We assign higher weights to price levels closer to the mid-price and lower weights to price levels

deeper in the LOB. The slope and intercept of the fitted line are used to estimate c and cp, from where we estimate the parameters c and p for that day.

This procedure is repeated for all 252 trading days in the year 2019, yielding 252 daily estimates of c and p. The annual sample averages and second moments of these daily estimates are presented in Table 2.4.

Table 2.4: Average values of $\hat{\mu}^{\pm}_{\{c,p\}}$ over 252 trading days for 2019 AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second.

$\bar{\mu}^{+}_{c} = 29.73$	$\bar{\mu}^{-}_{c} = 32.29$
$\bar{\mu}^{+}_{p} = 3.979$	$\bar{\mu}^{-}_{p} = 3.972$
$\bar{\mu}^{+}_{cp} = 113.76$	$\bar{\mu}^{-}_{cp} = 124.13$
$\bar{\mu}^{+}_{c^2} = 955.70$	$\bar{\mu}^{-}_{c^2} = 1111.46$
$\bar{\mu}^{+}_{p^2} = 16.40$	$\bar{\mu}^{-}_{p^2} = 16.33$

(a) Latency period

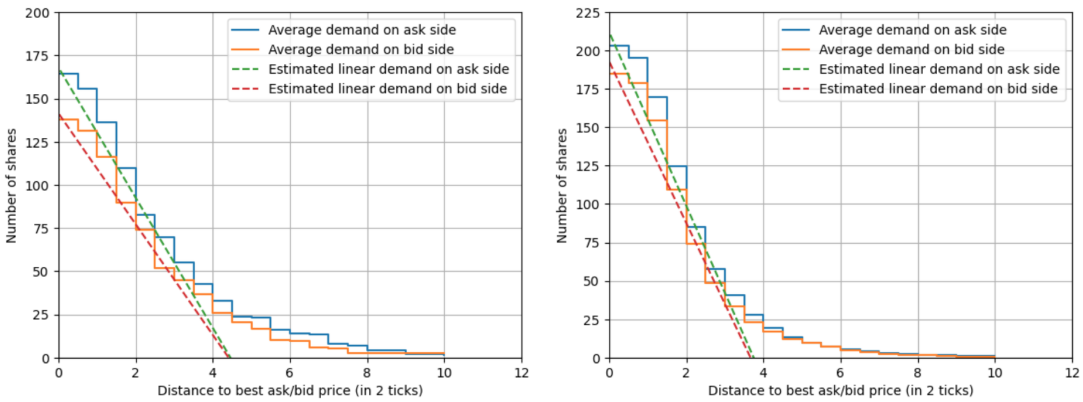
$\bar{\mu}^{+}_{c_+} = 48.06$	$\bar{\mu}^{-}_{c_+} = 53.54$
$\bar{\mu}^{+}_{p_+} = 3.269$	$\bar{\mu}^{-}_{p_+} = 3.317$
$\bar{\mu}^{+}_{cp_+} = 156.43$	$\bar{\mu}^{-}_{cp_+} = 176.49$
$\bar{\mu}^{+}_{c^2_+} = 2361.22$	$\bar{\mu}^{-}_{c^2_+} = 2941.92$
$\bar{\mu}^{+}_{p^2_+} = 10.78$	$\bar{\mu}^{-}_{p^2_+} = 11.11$

(b) Non-latency period

$\bar{\tilde{\mu}}^{+}_{c} = 48.40$	$\bar{\tilde{\mu}}^{-}_{c} = 53.89$
$\bar{\tilde{\mu}}^{+}_{p} = 3.258$	$\bar{\tilde{\mu}}^{-}_{p} = 3.305$
$\bar{\tilde{\mu}}^{+}_{cp} = 157.03$	$\bar{\tilde{\mu}}^{-}_{cp} = 177.08$
$\bar{\tilde{\mu}}^{+}_{c^2} = 2394.32$	$\bar{\tilde{\mu}}^{-}_{c^2} = 2979.13$
$\bar{\tilde{\mu}}^{+}_{p^2} = 10.71$	$\bar{\tilde{\mu}}^{-}_{p^2} = 11.03$

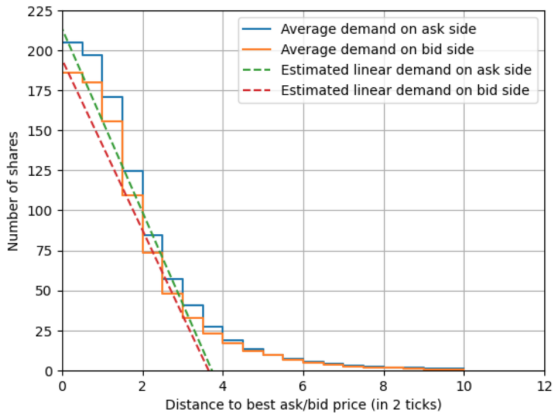
(c) Whole period

Figure 2.1a provides a comparison between the average actual demands $\overline{D}_l$ and the estimated linear demand for the latency period on the trading day of January 2nd, 2019. Analogously, Figure 2.1b shows a comparable analysis, focusing on non-latency periods, while Figure 2.1c encompasses the entire 1-second interval. These figures show that our linear demand assumption in Eq. (2.8) is reasonable.



(a) Latency period.

(b) Non-latency period.



(c) Whole period.

Figure 2.1: Average demand vs. estimated linear demand during a 1-second trading interval on January 2, 2019, for AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second.

Value Functions. To compute the optimal strategy, we need to evaluate the functions $g_k(\cdot)$ and $G_k(\cdot)$ defined in Eq. (2.35). To this end, we need to evaluate the relevant quantities. We employed a Monte Carlo based approach for the computation of $H^{\pm}_{act_{t_k}}(r^{\pm}, L^{\pm}, q^{\pm})$, as defined in Eq. (2.33). For illustration purpose, we will focus on the calculation of $H^{+}_{t_k}(\Delta\tau, r^{+}, q^{+})$, noting that the computation of the other involved quantities follows a similar method. Recall from Eq. (2.26) that

$$H^{+}_{t_k}(\Delta\tau, r^{+}, q^{+}) = \mathbb{E}\big[(r^{+} - \Delta S_{t_k}) \cdot [\mathbb{1}_{fill^{+}_{k}} c^{+}_{t_k}(p^{+}_{t_k} - r^{+})_{+} \wedge q^{+}] \mid X_{t_k} = x\big].$$

Under the assumption of constant arrival probabilities of MOs and time-invariant distribution for the demand parameters c and p across all time instances t_k, we can simplify $H^+_{t_k}(\Delta\tau, r^+, q^+)$ as follows:

$$H^+(\Delta\tau, r^+, q^+) = \mathbb{E}\big[(r^+ - S_{\Delta\tau}) \cdot [\mathbb{1}_{fill^+_{\Delta\tau}} c^+_{\Delta\tau}(p^+_{\Delta\tau} - r^+)_+ \wedge q^+]\big]. \tag{2.59}$$

To obtain these estimates, we conducted 1,000,000 simulation trials for each admissible combination of (r^+, q^+) in a finite domain. In our numerical experiments, we take $r^+ \in \{0, 1, 2, 3, \infty\}$ and $q^+ \in \{0, 1, \ldots, 5\}$. For each trial, we generated all the random variables therein according the following procedure:

- $S_{\Delta\tau}$ was simulated according to the price process specified in Eq (2.58). Specifically, the Poisson rate ν was set to 4.428507, which results from the sum of the upward and downward rates detailed in Table 2.2.
- $\mathbb{1}_{fill^+_{\Delta\tau}}$ follows a Bernoulli distribution with parameter $\pi^+_{\Delta\tau}$ (see Table 2.3 for the estimate of $\pi^+_{\Delta\tau}$).
- $c^+_{\Delta\tau}$ and $p^+_{\Delta\tau}$ were drawn from independent Gamma distributions. We used the Method of Moments to find the shape and scale parameters so that so recover the sample moments given in Table 2.4. The values of the estimated gamma parameters are outlined in Table 2.5.

2.4.2 Inventory and Optimal Spread Process

Utilizing Theorem 2.2.2, we can compute an approximate optimal placing strategy as illustrated in this subsection. We start with the calculation of $g_N(x)$, for x in a constrained

Table 2.5: Parameter values for c and p in simulation for 2019 AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second.

Parameter	**Value**	**Unit**
Slope of demand function $c^+_{\Delta\tau}$	$\Gamma(shape = 12.32, scale = 2.411)$	share/tick
Slope of demand function $c^-_{\Delta\tau}$	$\Gamma(shape = 15.23, scale = 2.120)$	share/tick
Slope of demand function $c^+_{\Delta\tau,\Delta t}$	$\Gamma(shape = 45.38, scale = 1.059)$	share/tick
Slope of demand function $c^-_{\Delta\tau,\Delta t}$	$\Gamma(shape = 38.57, scale = 1.388)$	share/tick
Slope of demand function $\tilde{c}^+_{\Delta t}$	$\Gamma(shape = 45.55, scale = 1.062)$	share/tick
Slope of demand function $\tilde{c}^-_{\Delta t}$	$\Gamma(shape = 39.01, scale = 1.381)$	share/tick
Reservation price $p^+_{\Delta\tau}$	$\Gamma(shape = 28.17, scale = 0.141)$	tick
Reservation price $p^-_{\Delta\tau}$	$\Gamma(shape = 28.20, scale = 0.140)$	tick
Reservation price $p^+_{\Delta\tau,\Delta t}$	$\Gamma(shape = 106.95, scale = 0.030)$	tick
Reservation price $p^-_{\Delta\tau,\Delta t}$	$\Gamma(shape = 98.19, scale = 0.033)$	tick
Reservation price $\tilde{p}^+_{\Delta t}$	$\Gamma(shape = 109.99, scale = 0.029)$	tick
Reservation price $\tilde{p}^-_{\Delta t}$	$\Gamma(shape = 100.96, scale = 0.032)$	tick

finite domain Ψ. In our numerical experiments, we take $\Psi := \{(i, q^+, q^-, r^+, r^-) : i \in \{-11, -10, \ldots, 11\}, q^\pm \in \{0, 1, \ldots, 5\}, r^\pm \in \{0, 1, 2, 3, \infty\}\}$. Recall from Eq. (2.35) that

$$g_N(x) = H^+_{t_N}(\Delta\tau, r^+, q^+) + H^-_{t_N}(\Delta\tau, r^-, q^-) - \lambda\mathbb{E}\big[(i - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}})^2 \mid X_{t_N} = x\big].$$

Here, $H^+_{t_N}(\Delta\tau, r^+, q^+)$ and $H^-_{t_N}(\Delta\tau, r^-, q^-)$ are obtained from our Monte Carlo estimates as described after Eq. (2.59).

Under the same time-invariant assumptions applied in the derivation of Eq. (2.59),

$$\begin{aligned}
&\mathbb{E}[(i - Q_{fill^+_{\Delta\tau}} + Q_{fill^-_{\Delta\tau}})^2 \mid X = x] \\
&= \mathbb{E}[(i - \mathbb{1}_{fill^+_{\Delta\tau}} c^+_{\Delta\tau}(p^+_{\Delta\tau} - r^+)_+ \wedge q^+ + \mathbb{1}_{fill^-_{\Delta\tau}} c^-_{\Delta\tau}(p^-_{\Delta\tau} - r^-)_+ \wedge q^-].
\end{aligned}$$

We then carried out a Monte Carlo estimation to compute the above quantity with 1,000,000 simulation trials for all $x \in \Psi$, generating the random variables $\mathbb{1}_{fill^\pm_{\Delta\tau}}$, $c^\pm_{\Delta\tau}$ and $p^\pm_{\Delta\tau}$ as outlined after Eq. (2.59).

For the recursive step in Eq. (2.35), for $x \in \Psi$, we have

$$\begin{aligned} g_{N-1}(x) &= \max_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} G_{N-1}(x, l^+, l^-) \\ &= \max_{(l^+,l^-)\in\mathbb{Z}_+\cup\{\infty\}\cup\{o\}} \big\{\mathbb{E}\big[g_N(X_{t_N}) \mid X_{t_{N-1}} = x, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-\big] \\ &\qquad + H^+_{act_{t_{N-1}}}(r^+, l^+, q^+) + H^-_{act_{t_{N-1}}}(r^-, l^-, q^-)\big\}. \end{aligned} \tag{2.60}$$

To compute $\mathbb{E}\big[g_N(X_{t_N}) \mid X_{t_{N-1}} = x, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-\big]$, we rewrite it as

$$\sum g_N(x') \cdot \mathbb{P}(X_{t_N} = x' \mid X_{t_{N-1}} = x, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-), \tag{2.61}$$

where the summation in (2.61) is over all $x' \in \Psi$. Since the distribution of X_{t_N} depends on $X_{t_{N-1}}, L^\pm_{t_{N-1}}, Q_{fill^\pm_{t_{N-1}}}, Q_{fill^\pm_{t_{(N-1)+}}}$ and $\Delta S_{t_{N-1}}$, for each $L^\pm_{t_{N-1}} \in \{0, 1, 2, 3\}$, we have that

$$\begin{aligned} &\mathbb{P}(X_{t_N} = x' \mid X_{t_{N-1}} = x, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-) \\ &= \sum \mathbb{P}\big(Q_{fill^\pm_{t_{N-1}}} = q^\pm_{\Delta\tau}, Q_{fill^\pm_{t_{(N-1)+}}} = q^\pm_{\Delta\tau,\Delta t}, S_{\Delta t} = s \mid X_{t_{N-1}} = x, L^+_{t_{N-1}} = l^+, L^-_{t_{N-1}} = l^-\big), \end{aligned} \tag{2.62}$$

where the summation is over all $q^\pm_{\Delta\tau} \in \{0, 1, \ldots, 5\}$, $q^\pm_{\Delta\tau,\Delta t} \in \{0, 1, \ldots, 5\}$ and $s \in \{-0.02, -0.015, \ldots, 0.02\}$ such that $X_{t_N} = x'$.

To compute the probability in the right-hand side of Eq. (2.62), we resort, once more, to Monte Carlo, conducting 1,000,000 simulation trials for each admissible combination of (x, l^+, l^-), where $x \in \Psi$ and $l^\pm \in \{0, 1, 2, 3, o, \infty\}$. In each trial, we simulated the number of filled shares for both the ask and bid sides over three time intervals: latency, nonlatency, and the entire 1-second period. The simulation methodology for determining the number of filled shares aligns with the approach utilized in the computation of the value functions, as detailed in the paragraph after Eq. (2.59). The parameters used for generating the

values of $(c^{\pm}_{t_k}, c^{\pm}_{t_{k+}}, p^{\pm}_{t_k}, p^{\pm}_{t_{k+}}, \tilde{c}^{\pm}_{t_k}, \tilde{p}^{\pm}_{t_k})$ can be found in Table 2.5. A precautionary measure was taken to ensure that the number of filled shares during the latency periods did not exceed the outstanding shares, if such outstanding shares existed.

2.4.3 Results

In this subsection, we assess the performance of the optimal strategy by applying it to the AAPL stock during the year 2019. To quantify the effectiveness of the strategy, we execute it against observed market data, which allows us to compute the terminal cash flow W_T and inventory I_T for each trading day. Considering the dynamic nature of the market and the wide variation in the arrival probabilities of MOs throughout the day, we analyze the terminal P&L value $W_T + \bar{S}_T I_T$ at different time points. Here, $\bar{S}_T$ represents the average price per share that the market maker would obtain when liquidating inventory I_T through a MO at time T, taking into account the state of the LOB at time T. Therefore, $\bar{S}_T I_T$ is the cash flow generated by the liquidation of the inventory I_T using a MO at time T. It is important to note that, due to the high computational burden (see section 2.4.4), our trading horizon is set to only 1 minute.

Table 2.7 presents the outcomes of our analysis for a specific trading scenario, where $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds and the inventory penalty parameter λ, is 0.01. The "overall mean" reflects the average terminal P&L value obtained across all 252 trading days for that particular minute of trading. Observe that the overall means of the terminal P&L values are all negative across different trading times. This outcome can be attributed to the fact that the rate of arrival of MOs is time-dependent, which can significantly differ from the constant value that was estimated and used in the computation of the optimal strategy. When the actual prevailing arrival rate deviates significantly from our estimations, it is

expected our placement strategy won't be suitable, leading to unfavorable outcomes and resulting in negative terminal P&L values. As an example, Table 2.6 shows the actual arrival probabilities of MOs at different time instances t_k.

Table 2.6: Average arrival probability of market orders during half-hour intervals at different times of day averaged over 252 trading days in 2019 for AAPL. $\Delta\tau = 0.02$ seconds and $\Delta t = 1$ second.

Time	$\widehat{\pi}^{+}_{\Delta\tau}$	$\widehat{\pi}^{-}_{\Delta\tau}$	$\widehat{\pi}^{+}_{\Delta\tau,\Delta t}$	$\widehat{\pi}^{-}_{\Delta\tau,\Delta t}$	$\widehat{\tilde{\pi}}^{+}_{\Delta t}$	$\widehat{\tilde{\pi}}^{-}_{\Delta t}$
10:00am-10:30am	0.016158	0.016387	0.388880	0.386192	0.395843	0.393124
12:00pm-12:30pm	0.009143	0.009199	0.236026	0.245169	0.241600	0.250398
14:30pm-15:00pm	0.008059	0.008287	0.227097	0.233576	0.231792	0.238209

To improve the suitability of our trading strategy, we implement a pre-screening process that takes into account the recently observed arrival probabilities of MOs. Specifically, prior to initiating our trading process and placing LOs, we estimate the arrival probabilities of MOs using the 5-minute window immediately preceding the potential trading period. Then, we proceed with our optimal strategies, but only if the arrival probabilities of both, buy and sell MOs, fall within a certain margin of error (MOE) from the average arrival probabilities displayed in Table 2.3.

The filtered trading days, which are the subset of trading days that passed the pre-screening process, demonstrate much more favorable results as they consistently yield positive terminal P&L values.

The "Filtered Days" column in the terminal P&L table provide the count of filtered trading days corresponding to different margin of error thresholds across various trading times. To further explore the dynamics of the inventory management, we introduce the mean of the absolute value of the final trading inventory. This measure provides insights into the magnitude of the inventory levels at the conclusion of each trading period.

The findings presented in Table 2.7 show the importance of the pre-screening process in improving the reliability and profitability of our trading strategy. By incorporating real-time estimations of MO arrival rates, we are able to adapt to current market conditions and achieve more favorable outcomes.

Subsequently, we present the outcomes of the terminal P&L values analysis for various combinations of latency and inventory penalty parameter.

Table 2.7: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-186105.47	10%	25	156911.60	2.76
		20%	72	113662.01	2.41
12:00-01pm	-61237.04	10%	19	458.09	2.28
		20%	54	440.44	2.44
13:00-01pm	-6357.26	10%	12	541.66	1.50
		20%	41	424.39	1.68
14:00-01pm	-7089.52	10%	11	80.83	2.18
		20%	40	163185.56	2.81
15:00-01pm	-93871.98	10%	11	127.27	3.09
		20%	51	77216.41	2.30

Latency

In this section, we investigate the impact of latency on the trading outcomes, by considering additional latency value, namely 0.2s, in addition to the previously discussed 0.02s case. Detailed information regarding the estimated parameters associated with the latency value 0.2s can be found in the appendix. We present the results of our analysis in Tables 2.8.

When considering a latency of 0.2 seconds, we observe that the mean terminal P&L values, even for the filtered trading days, are consistently negative, indicating the difficulty of generating profits under such latency conditions. This outcome might be attributed to the fact that a latency of 0.2 seconds is excessively large for a one-second trading interval. Consequently, the market maker faces significant challenges in cancelling unfavorable LOs in a timely manner. The delayed cancellation process exacerbates the impact of adverse placements, leading to further negative outcomes.

From the numerical results of optimal actions taken, we can see when latency gets larger, we see more aggressive placements. One of the potential reasons is that the increase in latency appears to influence the risk perception of market makers. With delays in order execution, market maker might prefer more aggressive placements to offset potential losses or capitalize on favorable market movements. Another potential reason is that market maker who faces higher latency might feel pressured to secure her positions swiftly due to the risk of market changes during the delay in execution. This pressure could lead to more assertive order placements.

These findings signal that latency might be an important factor in market making and display the difficulties associated with trading under prolonged latency periods. The results demonstrate the adverse effects of increased latency on profitability and emphasize the importance of minimizing latency to enhance trading performance.

Inventory Penalty Parameter λ

Here, we analyze the P&L obtained when setting the inventory penalty parameter λ, to 0.1 and 1. By referring to Table 2.9 and 2.10, we can compare these outcomes with the results presented in Table 2.7 where $\lambda = 0.01$.

Table 2.8: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.20$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-335916.46	10%	22	365030.62	3.16
		20%	64	33493.35	3.20
12:00-01pm	-149763.88	10%	23	-257235.0	2.66
		20%	54	-110954.46	3.00
13:00-01pm	-211093.51	10%	8	-875.62	4.25
		20%	33	-800243.15	3.65
14:00-01pm	-39949.52	10%	11	-683701.36	3.18
		20%	36	-209331.94	2.91
15:00-01pm	-161078.49	10%	12	-187377.69	3.15
		20%	49	285476.12	2.48

When a higher inventory penalty is applied, the terminal inventory is reduced while also the terminal P&L values decrease because the market maker aims to minimize her inventory holdings throughout the trading horizon. While maintaining lower terminal inventory levels mitigates the risk of unwinding the inventory at unfavorable prices through MOs, it also means that the market maker will likely miss out on potential trading opportunities to generate greater profits when favorable market conditions arise.

The trade-off between inventory penalty and potential profit demonstrates the importance of striking a balance between risk mitigation and capitalizing on advantageous trading situations. By adjusting the inventory penalty parameter, market maker can tailor her trading strategies to align with her risk tolerance and profit objectives.

Table 2.9: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.1$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-260323.13	10%	25	78537.0	2.76
		20%	72	27264.02	2.34
12:00-01pm	-53448.31	10%	19	610.47	2.04
		20%	54	420.80	2.23
13:00-01pm	-3115.83	10%	12	458.33	1.33
		20%	41	439.02	1.46
14:00-01pm	-59233.03	10%	11	606.36	1.72
		20%	40	163300.79	2.63
15:00-01pm	-29997.14	10%	11	625.0	2.7
		20%	51	461.42	2.02

Table 2.10: Terminal P&L for AAPL stock in 2019 across different trading times. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 1$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-243442.95	10%	25	78275.00	1.20
		20%	72	27128.61	0.80
12:00-01pm	-61488.19	10%	19	239.04	0.80
		20%	54	55.62	0.83
13:00-01pm	-37329.46	10%	12	135.76	0.07
		20%	41	-134188.88	0.55
14:00-01pm	-71551.15	10%	11	-281.66	1.16
		20%	40	163237.61	1.02
15:00-01pm	-33537.46	10%	11	77.27	0.63
		20%	51	-351.32	0.83

"Doing Nothing" Action

We now investigate the significance of the "doing nothing" action in our trading strategy. Table 2.11 presents the results obtained when the market maker has to place new LOs at each time interval.

The results demonstrate that the model that allows the market maker to "do nothing", outperforms the model that forces the market maker to place new LOs at the beginning to each period. This outcome can be attributed to the potential drawbacks associated with sending new orders, particularly in the presence of latency. When the market price experiences fluctuations within the latency window, the newly placed orders from the market maker may inadvertently enter the LOB at undesirable price levels. Also, when the market maker already has outstanding orders at favorable prices, opting for the "doing nothing" action can yield better outcomes. By refraining from cancelling the current position and sending additional orders, the market maker avoids entering the market at worse prices due to the latency-induced price movements.

Table 2.11: Terminal P&L for AAPL stock in 2019 across different trading times when placing new LOs only at all time intervals. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.1$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-289923.55	10%	25	156602.0	3.16
		20%	72	77802.43	3.09
12:00-01pm	-82806.68	10%	19	-137386.66	3.09
		20%	54	-51481.33	2.83
13:00-01pm	-77844.98	10%	12	363.33	2.00
		20%	41	365.36	2.09
14:00-01pm	-3875.69	10%	11	-348.33	2.58
		20%	40	127365.56	3.04
15:00-01pm	-86349.96	10%	11	160.0	3.90
		20%	51	6139.71	3.16

Compared with Gao's Method

[24] focuses on market making strategies specifically tailored for large-tick assets in the presence of latency, where only a single unit of filled shares is allowed. In contrast, our

research extends the scope by enabling partial fills for LOs. In this section, we illustrate the advantages these partial fulfillments offer in improving the final P&L outcomes.

The results from [24] are presented in Table 2.12, which illustrates a distinct characteristic where LOs are executed either in full quantity or not at all. This method results in fewer instances of order fulfillment. Consequently, the recorded terminal inventory exhibits significant size differences compared to those outlined in Table 2.7. The substantially larger terminal inventory could plausibly contribute to the markedly negative terminal P&L. Such negative results may arise due to the necessity for the market maker to unwind the substantial inventory employing MOs, which could potentially execute at unfavorable prices.

Table 2.12: Terminal P&L for AAPL Stock in 2019 across different trading times when LOs can only be filled at full quantity. The model parameters are: $\underline{I} = -11$, $\overline{I} = 11$, $I_{t_0} = 0$, $\Delta\tau = 0.02$ seconds, $\Delta t = 1$ second and $\lambda = 0.01$.

Trading Time	**Overall Mean**	**MOE**	**Filtered Days**	**Filtered Mean**	**Filtered Abs. Inv. Mean**
11:00-01am	-530579.28	10%	25	-394.20	7.00
		20%	72	-54.23	6.80
12:00-01pm	-164860.0	10%	19	-290902.14	6.71
		20%	54	-108943.30	6.35
13:00-01pm	-95249.34	10%	12	-616091.15	7.38
		20%	41	-177579.11	6.91
14:00-01pm	-315684.94	10%	11	-912.08	7.08
		20%	40	-94.77	7.15
15:00-01pm	-56337.55	10%	11	3003.18	8.18
		20%	51	-115092.54	7.18

2.4.4 Computational Challenge

During our simulation process, we encountered some computational difficulties stemming from the exhaustive search required in (2.35), particularly when trading over longer time horizons. To obtain the optimal spreads, we needed to consider all admissible combinations of $(i, q^+, q^-, r^+, r^-, l^+, l^-)$ for each time step. Specifically, inventory levels i ranged from -11 to 11, while outstanding shares $q^{\pm}$ ranged from {0, 1, 2, 3, 4, 5}. Outstanding spreads $r^{\pm}$ could take values from {0, 1, 2, 3, 99}, where 99 represented no outstanding orders, and new actions $l^{\pm}$ could take values from {0, 1, 2, 3, 10, 99}, where 10 and 99 denoted "doing nothing" and "canceling the LO (if any)", respectively. This resulted in a total of 745,200 possible combinations that we had to search for optimal spreads for each time step. Even for a modest 1-hour trading window (3600 seconds), this calculation could take about 40 hours to complete due to limited computational resources.

Consequently, we turned to a sub-optimal solution, opting for a local search with certain assumptions instead of an exhaustive search. When calculating the function $g_k(i)$ backward in time, we proceed as follows.

$$\begin{aligned} g_{k-1}(x) &= \max_{(l^+,l^-)\in\mathcal{N}_s} G_{k-1}\left(x, l^+, l^-\right) \\ &= \max_{(l^+,l^-)\in\mathcal{N}_s} \big\{\mathbb{E}\big[g_k(X_{t_N}) \mid X_{t_{k-1}} = x, (L^+_{t_{k-1}}, L^-_{t_{k-1}}) = (l^+, l^-)\big] \\ &\qquad + H^+_{act}\left(r^+, l^+, q^+\right) + H^-_{act}\left(r^-, l^-, q^-\right)\big\}. \end{aligned} \tag{2.63}$$

Here, we introduce the concept of the neighborhood $\mathcal{N}_s$, which is defined as the "neighbor" of previous sub-optimal spreads. To address the computational difficulties arising from the exhaustive search required to identify the optimal spreads at each time step, we assume that the optimal spreads from time t_k to t_{k-1} for a given inventory level i, outstanding ask and bid orders (q^+, q^-), and outstanding ask and bid spreads (r^+, r^-),

should not differ significantly. Based on this assumption, we perform a local search for the optimal spreads at time t_{k-1} within the neighborhood of the optimal spreads at time t_k. This approach reduces the computational complexity significantly, allowing us to obtain sub-optimal solutions in a feasible amount of time and with limited computational resources.

Our proposed approach involves an initial exhaustive search over a specified number of iterations to determine a set of optimal spreads. Subsequently, we proceed with the neighborhood search strategy to refine sub-optimal spreads. The optimal balance between the number of full searches and neighborhood searches, as well as the performance of this method, could be a subject for future investigation.

Chapter 3

Reinforcement Learning in Market Making with Latency

In this chapter, our focus lies on evaluating the efficacy of RL algorithms in addressing the market making problem with a unique distinction from existing literature. Our model not only incorporates the crucial element of latency but also allows for partial fills of LOs. Unlike conventional approaches, our framework acknowledges the possibility of unfilled LOs that can later be filled. This becomes particularly important when opting for the "do nothing" action. This nuanced consideration of partial fills and the temporal dynamics of unfilled orders adds a layer of realism to our model, capturing the intricate nature of market dynamics and enhancing the representational capacity of the RL algorithms in the context of latency-infused market making.

3.1 Simulation Study

In this section, we train a market making agent using RL methods in an artificial market crafted based on the model delineated in Section 2.1.1. Our objective is to assess the performance of the RL-driven strategy and draw comparisons with the analytical optimal control outlined in Section 2.2.

3.1.1 RL Settings

Simulator. We establish a simulated trading environment based on the market-making with latency model described in Section 2.1.1 to facilitate the training of the RL agent in making strategic decisions.

Each episode, representing a temporal trading segment, unfolds from t_0 to the terminal time T. The configuration of T in Table 3.1 designates each episode as equivalent to 1 minute, aligning with the setup detailed in Section 2.4.3. Within this episode, the RL agent executes actions on the LOB at discrete time $0 = t_0 < t_1 < \ldots < t_N < t_{N+1} = T$. These action times adhere to a regular interval of dt. Due to the constant latency $\Delta\tau \in [0, dt]$, the LOB registers the submitted actions only at time t_{k+}, as defined in Eq. (2.2). The starting fundamental price, setting the initial reference point for the market, is randomly drawn from a uniform distribution. The price process S_{t_k} is delineated in Eq. (2.58). The slopes of the demand functions, reservation prices, and MO arrival probabilities can be referenced in the provided tables.

Table 3.1: Default values of the parameters in simulation in Section 3.1.

Parameter	**Value(s)**	**Unit**
Terminal time T	60	second
Time interval dt	1	second
Latency $\Delta\tau$	0.02	second
Initial Price S_0	$uniform(1.4 \times 10^4, 2.2 \times 10^4)$	tick
Price Process S_{t_k}	See Eq. (2.58), $\nu = 4.43$	tick
Slopes of demand functions $c^{\pm}_{\Delta\tau}, c^{\pm}_{\Delta\tau,\Delta t}, \tilde{c}^{\pm}_{\Delta t}$	See Table 2.5	share/tick
Reservation prices $p^{\pm}_{\Delta\tau}, p^{\pm}_{\Delta\tau,\Delta t}, \tilde{p}^{\pm}_{\Delta t}$	See Table 2.5	tick
MO arrival probabilities $\pi^{\pm}_{\Delta\tau}$, $\pi^{\pm}_{\Delta\tau,\Delta t}$, $\tilde{\pi}^{\pm}_{\Delta t}$	See Table 2.3	

State Space. The state space of the environment consists of the following variables:

- Inventory level I_{t_k} of the RL agent;

- Remaining time till the end the trading horizon $T - t_k$;
- Ask/bid outstanding order sizes $Q^{\pm}_{out_{t_k}}$ and its quote pairs $r^{\pm}_{t_k}$.

Refer to Table 2.1 for the potential values that each state variable could take.

Action Space. The RL agent adheres to the strategy outlined in Section 2.1, executing actions on the LOB at specific times $0 = t_0 < t_1 < \ldots < t_N$, where $t_N < T$. In the expansive action space, the RL agent's choices $L^{\pm}_{t_k}$ range from 0 to 3, o, and ∞ on each side of the market. A choice of 0 to 3 prompts the RL agent to cancel any existing outstanding LOs and replace them with new ones at the corresponding quote. An action of o signifies a moment of inactivity, while ∞ prompts the cancellation of existing outstanding LOs without placing any new ones. This nuanced strategy unfolds through 36 distinct actions in the RL agent's action space.

Immediate reward function. We adopt the immediate reward function employed in [52],

$$R_{t_k} = (W_{t_{k+1}} - W_{t_k}) + (S_{t_{k+1}} I_{t_{k+1}} - S_{t_k} I_{t_k}) - [(e^{-\frac{T-t_{k+1}}{2000}})\lambda I^2_{t_{k+1}} - (e^{-\frac{T-t_k}{2000}})\lambda I^2_{t_k}], \quad (3.1)$$

where $W_{t_{k+1}}$, $S_{t_{k+1}}$, $I_{t_{k+1}}$ are defined in Eq. (2.4), (2.3), (2.5), respectively.

It quantifies alterations in the agent's cash holdings and inventory value, coupled with a penalty term on the agent's inventory level at time t_k. This penalty term, initially negligible, gains prominence as the trading period progresses, reaching its zenith as time converges towards T. As in [52], we also opt for a discount rate, γ in Eq. (1.1), of 1,

resulting in an aggregated reward scheme that is formulated as follows:

$$R = \sum_{k=0}^{N} R_{t_k} = W_T + S_T I_T - \lambda I_T^2.$$

Hence, the goal to be maximized by the RL agent aligns with the objective in the control problem of Section 2.2.

Action-value function approximation. We apply tile coding scheme to the state-action space, which includes state variables $(I_{t_k}, T - t_k, Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k})$ and action variables $(L^+_{t_k}, L^-_{t_k})$. Continuous state variables, I and t, spanning distinct ranges, are finely divided into intervals. For I_{t_k}, ranging from -11 to 11, and t_k, ranging from 0 to 60, a resolution of $\frac{1}{256}$ is recommended to capture subtle variations. This resolution is derived from the reciprocal of the product of the number of intervals per variable in each tiling (set to 8) and the number of layers of tilings (32).

In contrast, categorical variables, such as $Q^+_{out_{t_k}}, Q^-_{out_{t_k}}, r^+_{t_k}, r^-_{t_k}, L^+_{t_k}, L^-_{t_k}$, are treated as discrete tiles, with each unique value representing a distinct category. This configuration aims to strike a balance between granularity and computational efficiency, applying a finer resolution to continuous variables and a category-based approach for discrete ones.

RL algorithms. We train the RL agent employing SARSA and Q-learning, leveraging tile coding for the approximation of the action-value function. The SARSA and Q-learning algorithms for the simulation study are delineated in Algorithm 1 and 2. The default parameters used in the learning algorithms are given in Table 3.2.

Algorithm 1 SARSA Algorithm with Tile Coding

Input: Initialize action-value function weights: $w_i = 0$ for all tile i
Output: Learned policy π

for each episode **do**
  Initialize the environment: generate fundamental price S_{t_0}
  Initialize state S_{t_0}: inventory $I_{t_0} = 0$, remaining time $T - t_0 = T$, outstanding order sizes $Q^{\pm}_{out_{t_k}} = 0$ and outstanding quote pairs $r^{\pm}_{t_k} = \infty$
  Choose action A (e.g., ϵ-greedy)
  for each step t_k of the episode **do**
    Take action A
    Update the environment:
      Generate slopes of demand functions $c^{\pm}_{\Delta\tau}, c^{\pm}_{\Delta\tau,\Delta t}, \tilde{c}^{\pm}_{\Delta t}$
      Generate reservation prices $p^{\pm}_{\Delta\tau}, p^{\pm}_{\Delta\tau,\Delta t}, \tilde{p}^{\pm}_{\Delta t}$
      Generate Price change $S_{t_{k+1}} - S_{t_k}$
    Observe reward R and new state S_{new}
    Choose A_{new} from S_{new} using $q(S_{new}, a)$ (e.g., ϵ-greedy)
    for each tile j that covers the state-action pair (S, A) **do**
      $w_j \leftarrow w_j + \alpha[R + \gamma\hat{q}(S_{new}, A_{new}) - \hat{q}(S, A)]$
    end for
    $S \leftarrow S_{new}$
    $A \leftarrow A_{new}$
  end for
end for

Algorithm 2 Q-Learning Algorithm with Tile Coding

Input: Initialize action-value function weights: $w_i = 0$ for all tile i
Output: Learned policy π

for each episode **do**
    Initialize the environment: generate fundamental price S_{t_0}
    Initialize state S_{t_0}: inventory $I_{t_0} = 0$, remaining time $T - t_0 = T$, outstanding order sizes $Q^{\pm}_{out_{t_k}} = 0$ and outstanding quote pairs $r^{\pm}_{t_k} = \infty$
    Choose action A (e.g., ϵ-greedy)
    for each step t_k of the episode **do**
        Take action A
        Update the environment:
            Generate slopes of demand functions $c^{\pm}_{\Delta\tau}, c^{\pm}_{\Delta\tau,\Delta t}, \tilde{c}^{\pm}_{\Delta t}$
            Generate reservation prices $p^{\pm}_{\Delta\tau}, p^{\pm}_{\Delta\tau,\Delta t}, \tilde{p}^{\pm}_{\Delta t}$
            Generate Price change $S_{t_{k+1}} - S_{t_k}$
        Observe reward R and new state S_{new}
        Choose A_{new} from S_{new} using $q(S_{new}, a)$ (e.g., ϵ-greedy)
        for each tile j that covers the state-action pair (S, A) **do**
            $w_j \leftarrow w_j + \alpha[R + \gamma \max_a \hat{q}(S_{new}, a) - \hat{q}(S, A)]$
        end for
        $S \leftarrow S_{new}$
        $A \leftarrow A_{new}$
    end for
end for

Table 3.2: Default parameters for RL algorithms in Section 3.1.

Parameter	**Value**
Number of tilings	32
Tile resolution	8
Learning rate	10^{-5}/(Number of tilings)
Discount factor	1
Exploration rate	0.8

3.1.2 Simulation Results

Simulations were performed with different configurations of training and testing episodes, as detailed in Table 3.3. Specifically, we considered combinations such as 2000 training episodes and 500 testing episodes, 1000 training episodes and 250 testing episodes, and 120 training episodes with 30 testing episodes. In each testing episode, the objective value is computed as follows

$$W_T + S_T I_T - \lambda I_T^2,$$

serving as the performance criterion. A comparative analysis is then performed, contrasting the performance of RL with that of a random strategy. The random strategy involves the random selection of actions from the RL action space at each step.

Constrained by computational resources[2], the duration of each training and testing episode is approximately 5 minutes. This temporal constraint restricts our ability to obtain outcomes from a more comprehensive set of training and testing episodes. The comparative results are delineated in Table 3.3, which reveal the superior performance

[2]The experiments were conducted on a 2019 MacBook Pro with a 2.6 GHz 6-Core Intel Core i7 processor, 16 GB of 2667 MHz DDR4 RAM, and a 512 GB SSD. The operating system used was macOS Ventura. The program was implemented in Python 3.7 and executed using VS Code. The experiments were run without parallelization.

of both RL algorithms over the random strategy. Evidently, an increase in the number of training episodes of RL algorithms correlates with improved performance. This improvement is reflected in a larger terminal value and reduced standard deviation, indicating greater robustness. The observed trend also suggests that the RL algorithms have not reached convergence yet. However, constrained by temporal and computational limitations, this represents the extent of our current findings. The prospect of conducting further training and testing episodes stands as a potential avenue for future exploration in this section.

Table 3.3: Mean and standard deviation of objective values $W_T + S_T I_T - \lambda I_T^2$ (given in units of 10^4), and average terminal inventory (given in unit of 1 share). $\lambda = 0.0005$.

	Reinforcement Learning			**Random**	
Train / Test	Algorithm	P&L	Terminal Inv.	P&L	Terminal Inv.
120 / 30	SARSA	-9.5 (18.5)	-1.93	-27.9 (15.9)	-10.25
	Q-Learning	-3.6 (7.0)	2.42		
1000 / 250	SARSA	-1.1 (3.7)	1.66	-26.2 (16.4)	-9.96
	Q-Learning	-1.4 (3.8)	1.51		
2000 / 500	SARSA	-0.6 (2.9)	-1.33	-26.3 (16.6)	-11.26
	Q-Learning	-0.5 (2.5)	-1.37		

Chapter 4

Market Making with Running Inventory Penalty

This chapter builds upon the work presented in [12], which offers a comprehensive analysis of the optimal policy for a HFM engaged in bid and ask LO placement. The authors explore the strategic execution of these LOs at predetermined time intervals, aiming to maximize profitability through efficient round trip transactions, while considering the terminal inventory penalty. In this chapter, we extend the existing approach proposed in [12] by incorporating a penalty associated with the running inventory throughout the entire trading horizon. This novel approach allows for more effective management of the market maker's exposure to inventory risk. By integrating this running inventory penalty framework, we enhance the market maker's risk management capabilities and establish a more robust framework for evaluating trading strategies.

4.1 Model Setup

For the sake of completeness and clarity, we include the following notation and details, which are adapted from [12] and are similar to those in Chapter 2.

As before, we assume that the market making strategy spans from time 0 to a predetermined time point denoted as $T > 0$. Within this timeframe, the market maker strategically places bid and ask LOs simultaneously on both sides of the LOB for a specific asset. These LOs are executed at prearranged time intervals $0 = t_0 < t_1 < \cdots < t_N < T$, which are defined in Eq. (2.1).

We also let $t_{N+1} = T$, and the set $\mathcal{T}$ is defined as $\{t_0, t_1, \ldots, t_{N+1}\}$. Adopting the notation established in [12], all variables are defined within the context of a probability space denoted as $(\Omega, \mathbb{P}, \mathcal{F})$, which is equipped with a filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$.

As outlined in [12], the dynamics of buy and sell MOs are captured using a framework based on two Bernoulli processes. With a slight abuse of notation, we define $\mathbb{1}^{\pm}_{t_{k+1}}$ to be the same as $\widetilde{\mathbb{1}}_{fill^{\pm}_k}$ as defined in Eq. (2.12). Specifically, $\mathbb{1}^{+}_{t_{k+1}}$ and $\mathbb{1}^{-}_{t_{k+1}}$ are Bernoulli random variables that indicate the presence or absence of at least one buy or sell MO, respectively, within the time interval $[t_k, t_{k+1})$.

We make the assumption that $\mathbb{1}^{+}_{t_{k+1}}, \mathbb{1}^{-}_{t_{k+1}} \in \mathcal{F}_{t_{k+1}}$, implying that these variables are measurable with respect to the filtration $\mathcal{F}_{t_{k+1}}$. Moreover, we consider the joint probability distribution of $\mathbb{1}^{+}_{t_{k+1}}$ and $\mathbb{1}^{-}_{t_{k+1}}$ given $\mathcal{F}_{t_{k+1}}$, denoted as $\pi_{t_{k+1}}(j^+, j^-)$,

$$\mathbb{P}(\mathbb{1}^{+}_{t_{k+1}} = j^+, \mathbb{1}^{-}_{t_{k+1}} = j^- \mid \mathcal{F}_{t_{k+1}}) = \pi_{t_{k+1}}(j^+, j^-),$$

for $j^{\pm} \in \{0, 1\}$, where $\pi_{t_{k+1}} : \{0, 1\} \times \{0, 1\} \to [0, 1]$. This probability distribution $\pi_{t_{k+1}}$ characterizes the likelihood of specific combinations of buy and sell MOs occurring within the time interval.

To further explore the characteristics of its distribution, we introduce the notation $\pi^{\pm}_{t_{k+1}}$ to represent the marginal conditional probabilities. Specifically,

$$\pi^{\pm}_{t_{k+1}} := \mathbb{P}(\mathbb{1}^{\pm}_{t_{k+1}} = 1 \mid \mathcal{F}_{t_{k+1}}).$$

Throughout our analysis, we assume that $\pi^{+}_{t_{k+1}}$ and $\pi^{-}_{t_{k+1}}$ are both strictly positive.

Remark 4.1.1. *By definition of marginal probabilities, we have*

$$\pi^{+}_{t_{k+1}} = \pi_{t_{k+1}}(1,1) + \pi_{t_{k+1}}(1,0),$$
$$\pi^{-}_{t_{k+1}} = \pi_{t_{k+1}}(1,1) + \pi_{t_{k+1}}(0,1).$$

The following relation between $\pi_{t_{k+1}}(1,1)$ and $\pi^{\pm}_{t_{k+1}}$ holds for each t_{k+1}:

$$(\pi^{+}_{t_{k+1}} + \pi^{-}_{t_{k+1}} - 1) \vee 0 \leq \pi_{t_{k+1}}(1,1) \leq \pi^{+}_{t_{k+1}} \wedge \pi^{-}_{t_{k+1}}. \tag{4.1}$$

The ask LO is strategically placed at time t_k using the available information $\mathcal{F}_{t_{k+1}}$ at the price level $a_{t_k} \in \mathcal{F}_{t_{k+1}}$. Similarly, the bid LO is placed at time t_k at a price level $b_{t_k} \in \mathcal{F}_{t_{k+1}}$. We reparameterize a_{t_k} and b_{t_k} as follows:

$$a_{t_k} = S_{t_k} + L^{+}_{t_k}, \quad b_{t_k} = S_{t_k} - L^{-}_{t_k}.$$

Here, $L^{\pm}_{t_k} \in \mathcal{F}_{t_{k+1}}$ represent the market maker's spreads, while $S_{t_k} \in \mathcal{F}_{t_{k+1}}$ denotes the fundamental price of the asset at time t_k. Further details regarding the assumptions governing the fundamental price process $\{S_{t_k}\}_{k=0,\ldots,N+1}$ will be provided later in the analysis.

It is important to note that the LOs placed at time t_k may experience full or partial execution within the time interval $[t_k, t_{k+1})$ only if at least one MO arrives during that specific period. Following [12], the number of filled shares on the bid side during the interval $[t_k, t_{k+1})$ is assumed to be determined by

$$Q^-_{t_{k+1}} \triangleq \mathbb{1}^-_{t_{k+1}} c^-_{t_{k+1}} \left[b_{t_k} - (S_{t_k} - p^-_{t_{k+1}}) \right] = \mathbb{1}^-_{t_{k+1}} c^-_{t_{k+1}} (p^-_{t_{k+1}} - L^-_{t_k}).$$

The quantity of executed shares on the buy side during the interval $[t_k, t_{k+1})$ is influenced by the positive random variables $c^-_{t_{k+1}}$ and $p^-_{t_{k+1}}$, both of which belong to the information set $\mathcal{F}_{t_{k+1}}$. We will specify their distribution in Assumption 4.1.1.

When no sell MO arrives within the interval $[t_k, t_{k+1})$, the indicator variable $\mathbb{1}^-_{t_{k+1}}$ takes a value of 0, resulting in no executions on the buy side. Here, $p^-_{t_{k+1}}$ is defined such that $S_{t_k} - p^-_{t_{k+1}}$ represents the lowest price at which all sell MOs arriving during $[t_k, t_{k+1})$ can be executed. In other words, the bid LOs placed by the HFT will not be executed during $[t_k, t_{k+1})$ if the price falls below $S_{t_k} - p^-_{t_{k+1}}$. We refer to $p^-_{t_{k+1}}$ as the reservation price for sellers.

The demand slope $c^-_{t_{k+1}}$ quantifies the rate at which the number of executed shares on the bid order increases as the bid price b_{t_k} approaches the fundamental price S_{t_k}. It captures the sensitivity of the HFM's bid order to changes in price, reflecting the trading behavior in response to the proximity to the fundamental value.

Similarly, the number of shares filled by the HFM's ask LO during $[t_k, t_{k+1})$ is expressed as

$$Q^+_{t_{k+1}} \triangleq \mathbb{1}^+_{t_{k+1}} c^+_{t_{k+1}} \left[(S_{t_k} + p^+_{t_{k+1}}) - a_{t_k} \right] = \mathbb{1}^+_{t_{k+1}} c^+_{t_{k+1}} (p^+_{t_{k+1}} - L^+_{t_k}).$$

Next, we state the main assumptions on $c_\cdot^\pm$ and $p_\cdot^\pm$.

Assumption 4.1.1 (General Properties of $(c_\cdot^\pm, p_\cdot^\pm)$)**.** *For $k = 0, \ldots, N$, we have*

1. *$(c_\cdot^\pm, p_\cdot^\pm)$ are $\mathcal{F}_{t_{k+1}}$ − measurable,*
2. *the conditional distribution of $(c_{t_{k+1}}^+, p_{t_{k+1}}^+, c_{t_{k+1}}^-, p_{t_{k+1}}^-)$ given $(\mathcal{F}_{t_k}, \mathbb{1}_{t_{k+1}}^+, \mathbb{1}_{t_{k+1}}^-)$ does not depend on k and is nonrandom,*
3. *$(c_{t_{k+1}}^+, p_{t_{k+1}}^+)$ and $(c_{t_{k+1}}^-, p_{t_{k+1}}^-)$ are independent given $(\mathcal{F}_{t_k}, \mathbb{1}_{t_{k+1}}^+, \mathbb{1}_{t_{k+1}}^-)$.*

Next, we introduce some further notation related to $(c_\cdot^\pm, p_\cdot^\pm)$. For $a, b \in \{0, 1, 2\}$,

$$\mu_{c^a p^b}^\pm := \mathbb{E}\big[(c_{t_{k+1}}^\pm)^a (p_{t_{k+1}}^\pm)^b \mid \mathcal{F}_{t_k}, \mathbb{1}_{t_{k+1}}^\pm = 1\big]. \tag{4.2}$$

In order to incorporate the price impact resulting from liquidating a significant net position using MOs at the end of the trading horizon, we introduce a penalty term in the objective function that accounts for both the running inventory and the terminal inventory. This penalty term captures the adverse effects on the market caused by the HFM's efforts to unwind a substantial position.

The primary objective of the HFM is to maximize the following expression:

$$\max_{(L_\cdot^+, L_\cdot^-) \in \mathcal{A}} \mathbb{E}\big[W_T + S_T I_T - \lambda I_T^2 - \phi \sum_{j=k+1}^{N+1} I_{t_j}^2\big], \tag{4.3}$$

where $\mathcal{A}$ represents the set of all $\mathcal{F}$-adapted processes. Specifically, $\mathcal{A}$ is a set that includes all processes whose values at each time are determined by the information available up to that time, given by the sigma-algebra $\mathcal{F}$; and inventory penalty parameters $\lambda, \phi \geq 0$. Within this context, W_T signifies the cash holdings of the market maker, while I_T denotes

the inventory maintained by the market maker at the conclusion of the period $[0, T]$. Additionally, I_{t_j} represents the running inventory at time t_j, where t_j is a specific point within the trading horizon. This running inventory captures the evolving net position held by the market maker throughout the course of trading.

In Equation (4.3), the term λI_T^2 serves as the penalty for the terminal inventory. This penalty component reflects the price impact or cost associated with liquidating a significant net position using MOs at the end of the trading horizon.

Moreover, the term $\phi \sum_{j=k+1}^{N+1} I_{t_j}^2$ constitutes the penalty for the running inventory. This penalty term captures the adverse effects arising from maintaining a running inventory over the trading horizon. The parameter ϕ scales the running inventory penalty, influencing the weight assigned to the running inventory component in the overall objective. It is crucial to recognize that holding inventory can introduce inherent risks and uncertainties. The market maker is exposed to potential price fluctuations and market movements that may adversely impact the profitability of her inventory holdings. Therefore, it becomes imperative for the market maker to diligently assess the trade-off between holding inventory and the associated risks it entails. By striking the right balance between minimizing inventory and maximizing trading opportunities, the market maker can effectively navigate the trade-offs and make informed decisions that align with her overarching goal of maximizing expected terminal wealth.

By incorporating both the penalty for terminal inventory and the penalty for running inventory within the objective function, the model appropriately accounts for the potential price impact of liquidating a large net position at the end of the trading horizon, as well as the ongoing impact of maintaining a running inventory throughout the trading period.

Drawing from the established notation employed by [12], the cash holding and inventory processes, $\{W_{t_k}\}$ and $\{I_{t_k}\}$, respectively, satisfy the following equations:

$$\begin{aligned} W_{t_{k+1}} &= W_{t_k} + a_{t_k} Q^+_{t_{k+1}} - b_{t_k} Q^-_{t_{k+1}} \\ &= W_{t_k} + (S_{t_k} + L^+_{t_k}) \mathbb{1}^+_{t_{k+1}} c^+_{t_{k+1}} (p^+_{t_{k+1}} - L^+_{t_k}) - (S_{t_k} - L^-_{t_k}) \mathbb{1}^-_{t_{k+1}} c^-_{t_{k+1}} (p^-_{t_{k+1}} - L^-_{t_k}) \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} I_{t_{k+1}} &= I_{t_k} - Q^+_{t_{k+1}} + Q^-_{t_{k+1}} \\ &= I_{t_k} - \mathbb{1}^+_{t_{k+1}} c^+_{t_{k+1}} (p^+_{t_{k+1}} - L^+_{t_k}) + \mathbb{1}^-_{t_{k+1}} c^-_{t_{k+1}} (p^-_{t_{k+1}} - L^-_{t_k}), \end{aligned} \tag{4.5}$$

where $W_{t_0} = 0$ and $I_{t_0} = 0$.

4.2 Analytical Optimal Control

At time t_k, the value function of the control problem described in Eq. (4.3) is denoted by V_{t_k}. It is defined as the maximum expected value obtained by optimizing the control variables $(L^+_\cdot, L^-_\cdot)$ within the set $\mathcal{A}$. The value function can be expressed as follows:

$$V_{t_k} = \sup_{(L^+_\cdot, L^-_\cdot) \in \mathcal{A}} \mathbb{E}\Big[W_T + S_T I_T - \lambda I_T^2 - \phi \sum_{j=k+1}^{N+1} I^2_{t_j} \Big],$$

By applying the dynamic programming principle, we can establish a recursive relationship satisfied by the value function. This relationship, known as the **Bellman equation**, is given by:

$$V_{t_k} = \sup_{(L^+_\cdot, L^-_\cdot) \in \mathcal{A}} \mathbb{E}\big[V_{t_{k+1}} - \phi I^2_{t_{k+1}} \mid \mathcal{F}_{t_k} \big]. \tag{4.6}$$

4.2.1 Optimal Strategy Under a Martingale Fundamental Price Process

In this subsection, we assume that the fundamental price $S_{t_k} \in \mathcal{F}_{t_k}$ of the asset is a martingale:

$$\mathbb{E}[S_{t_{k+1}} \mid \mathcal{F}_{t_k}] = S_{t_k}, \quad k = 0, \ldots, N. \tag{4.7}$$

Furthermore, we assume conditional independence between $S_{t_{k+1}} - S_{t_k}$ and $(\mathbb{1}^+_{t_{k+1}}, \mathbb{1}^-_{t_{k+1}}, c^+_{t_{k+1}}, p^+_{t_{k+1}}, c^-_{t_{k+1}}, p^-_{t_{k+1}})$, given $\mathcal{F}_{t_k}$.

We begin the analysis by introducing an ansatz for the value function V_{t_k}.

$$V_{t_k} = v\left(t_k, S_{t_k}, W_{t_k}, I_{t_k}\right) = W_{t_k} + \alpha_{t_k} I^2_{t_k} + S_{t_k} I_{t_k} + h_{t_k} I_{t_k} + g_{t_k}, \tag{4.8}$$

where $\alpha : \mathcal{T} \to \mathbb{R}$, $h : \mathcal{T} \to \mathbb{R}$, and $g : \mathcal{T} \to \mathbb{R}$ are deterministic functions defined on the time grid $\mathcal{T} = \{t_0, t_1, \ldots, t_{N+1}\}$. By plugging the ansatz into the Bellman equation (4.6), we have the following iterative representation for the value function

$$\begin{aligned} & W_{t_k} + \alpha_{t_k} I^2_{t_k} + S_{t_k} I_{t_k} + h_{t_k} I_{t_k} + g_{t_k} \\ = & \sup_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}} \mathbb{E}\big[W_{t_{k+1}} + \alpha_{t_{k+1}} I^2_{t_{k+1}} + S_{t_{k+1}} I_{t_{k+1}} + h_{t_{k+1}} I_{t_{k+1}} + g_{t_{k+1}} - \phi I^2_{t_{k+1}} \mid \mathcal{F}_{t_k}\big]. \end{aligned} \tag{4.9}$$

Since $V_T = W_T + S_T I_T - \lambda I^2_T$, we obtain the terminal conditions $\alpha_T = -\lambda, g_T = 0$, and $h_T = 0$.

The optimal control can be determined by solving Eq. (4.9) in a backward fashion. The proposition below provides the expression for the maximizing point $(L^{+,*}_{t_k}, L^{-,*}_{t_k})$. A similar result was obtained in [12] in the absence of the running inventory penalization.

Proposition 4.2.1 (Optimal Controls)**.** *The optimal controls that solve the optimization problem (4.9) using the ansatz (4.8) and state dynamics (4.4)-(4.5) are given by*

$$\begin{aligned} L^{+,*}_{t_k} &= {}^{(1)}A^{+}_{t_k} I_{t_k} + {}^{(2)}A^{+}_{t_k} + {}^{(3)}A^{+}_{t_k}, \\ L^{-,*}_{t_k} &= -{}^{(1)}A^{-}_{t_k} I_{t_k} - {}^{(2)}A^{-}_{t_k} + {}^{(3)}A^{-}_{t_k}, \end{aligned} \tag{4.10}$$

for $k = 0, \ldots, N$. The coefficients above are specified as

$$\begin{aligned} {}^{(1)}A^{\pm}_{t_k} &= \frac{\beta^{\pm}_{t_k}(\alpha_{t_{k+1}} - \phi)}{\gamma_{t_k}}, \quad {}^{(2)}A^{\pm}_{t_k} = \frac{\beta^{\pm}_{t_k} h_{t_{k+1}}}{2\gamma_{t_k}}, \\ {}^{(3)}A^{\pm}_{t_k} &= \frac{\pi^{\mp}_{t_{k+1}}}{2\gamma_{t_k}} \left((\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2} - \mu^{\mp}_{c}\right) \Big[\pi^{\pm}_{t_{k+1}}(\mu^{\pm}_{cp} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\pm}_{c^2 p}) \\ &\qquad + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{\pm}_{c}\mu^{\mp}_{cp}\Big] \\ &\quad + \pi_{t_{k+1}}(1,1)\frac{(\alpha_{t_{k+1}} - \phi)}{2\gamma_{t_k}}\mu^{+}_{c}\mu^{-}_{c}\Big[\pi^{\mp}_{t_{k+1}}(\mu^{\mp}_{cp} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2 p}) \\ &\qquad + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{\mp}_{c}\mu^{\pm}_{cp}\Big], \end{aligned} \tag{4.11}$$

and

$$\begin{aligned} \gamma_{t_k} &:= \left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^{+}_{c}\mu^{-}_{c}\right]^2 \\ &\qquad - \pi^{+}_{t_{k+1}}\pi^{-}_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu^{+}_{c^2} - \mu^{+}_{c}\right)\left((\alpha_{t_{k+1}} - \phi)\mu^{-}_{c^2} - \mu^{-}_{c}\right), \end{aligned} \tag{4.12}$$

$$\beta^{\pm}_{t_k} := \pi^{+}_{t_{k+1}}\pi^{-}_{t_{k+1}}\mu^{\pm}_{c}\left((\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2} - \mu^{\mp}_{c}\right) - \pi^{\mp}_{t_{k+1}}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^{\pm}_{c}\left(\mu^{\mp}_{c}\right)^2. \tag{4.13}$$

In the expressions above, $\alpha : \mathcal{T} \to \mathbb{R}$ and $h : \mathcal{T} \to \mathbb{R}$ are specified using the following backward equations: $\alpha_T = -\lambda, h_T = 0$ at $T = t_{N+1}$ and, for $k = 0, \ldots, N$:

$$\begin{aligned} \alpha_{t_k} = \alpha_{t_{k+1}} - \phi &+ \sum_{\delta=\pm}\pi^{\delta}_{t_{k+1}}\left[\left((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\right)\left({}^{(1)}A^{\delta}_{t_k}\right)^2 + 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c}\left({}^{(1)}A^{\delta}_{t_k}\right)\right] \\ &+ 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}\left({}^{(1)}A^{+}_{t_k}\,{}^{(1)}A^{-}_{t_k}\right) \end{aligned} \tag{4.14}$$

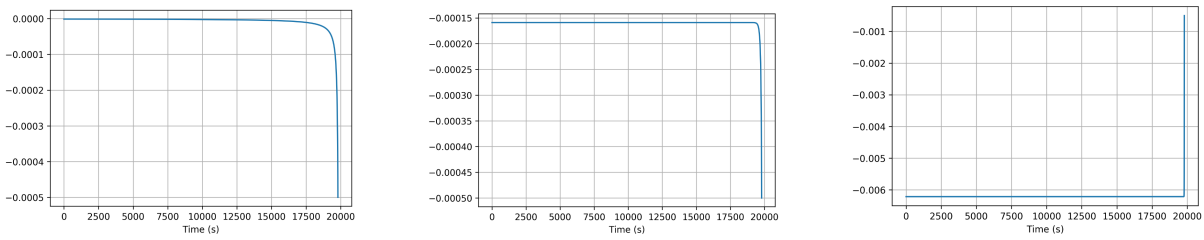
and

$$
\begin{aligned}
h_{t_k} =& h_{t_{k+1}} + \sum_{\delta=\pm} \pi^{\delta}_{t_{k+1}} \Big\{ 2\left((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\right)\left[{}^{(1)}A^{\delta}_{t_k}\left(\delta^{(3)}A^{\delta}_{t_k} + {}^{(2)}A^{\delta}_{t_k}\right)\right] \\
&\qquad + 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c}\left(\delta^{(3)}A^{\delta}_{t_k} + {}^{(2)}A^{\delta}_{t_k}\right) - 2(\alpha_{t_{k+1}} - \phi)\left(\delta\mu^{\delta}_{cp}\right) \\
&\qquad + \left(\delta^{(1)}A^{\delta}_{t_k}\right)\left(\mu^{\delta}_{cp} + \delta h_{t_{k+1}}\mu^{\delta}_{c} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2p}\right) \Big\} \qquad (4.15) \\
&- 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}\Big[{}^{(1)}A^{+}_{t_k}\left({}^{(3)}A^{-}_{t_k} - {}^{(2)}A^{-}_{t_k}\right) - {}^{(1)}A^{-}_{t_k}\left({}^{(2)}A^{+}_{t_k} + {}^{(3)}A^{+}_{t_k}\right) \\
&\qquad + \frac{\mu^{+}_{cp}}{\mu^{+}_{c}}\left({}^{(1)}A^{-}_{t_k}\right) - \frac{\mu^{-}_{cp}}{\mu^{-}_{c}}\left({}^{(1)}A^{+}_{t_k}\right)\Big].
\end{aligned}
$$

Upon comparing our findings with Proposition 1 in [12], we observe a similarity in the structure of the optimal controls. However, a notable distinction arises due to the incorporation of the running inventory penalty. In our case, the presence of the penalty parameter ϕ modifies the expression $\alpha_{t_{k+1}}$ to $\alpha_{t_{k+1}} - \phi$. This adjustment allows us to effectively manage the running inventory by accounting for the penalty at each trading interval when updating α_{t_k}. Consequently, our approach facilitates enhanced control over the inventory dynamics. Lemma 4.2.2 shows the negativity of α_{t_k}. This characteristic of α_{t_k} serves as a crucial element in the subsequent proofs of the following theorems and propositions.

Lemma 4.2.2. *The quantity α_{t_k} defined in Eq. (4.14) is negative for every t_k.*

In contrast to Lemma 1 presented in [12], we observe that the claim of strict decreasing behavior of α_{t_k} with respect to t_k does not hold. The monotonic behavior of the modified α_{t_k}, as defined in Eq. (4.14), is contingent upon the value of the parameter ϕ, as illustraded in Figure 4.1. Specifically, when ϕ takes on small values, α_{t_k} exhibits a strict decreasing trend with t_k. Conversely, for larger values of ϕ, α_{t_k} demonstrates a strict increasing pattern with t_k.



(a) $\phi = 0$ (b) $\phi = 1 \times 10^{-6}$ (c) $\phi = 1 \times 10^{-3}$

Figure 4.1: Paths of α_{t_k} for different values of running inventory penalty ϕ under symmetric market conditions. The agent acts every one second from 0 to 19800 seconds. Plots are generated with the following parameters: $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$ and $\pi_{t_k}(1,1) \equiv 0.1$.

Remark 4.2.1. *Through the solution of the Bellman Equation (4.9) following the approach outlined in Proposition 4.2.1, we are able to derive an explicit expression for the coefficient g_{t_k} in the value function given by Eq. (4.8). This computation enables us to obtain the value of g_{t_k} for each time step $k = 0, 1, \ldots, N$ by means of a backward iteration process.*

$$
\begin{aligned}
g_{t_k} = g_{t_{k+1}} + \sum_{\delta=\pm} \pi_{t_{k+1}}^{\delta} \Big[& \left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{\delta} - \mu_c^{\delta}\right) \left({}^{(3)}A_{t_k}^{\delta} + \left(\delta^{(2)}A_{t_k}^{\delta}\right)\right)^2 \\
& + (\alpha_{t_{k+1}} - \phi)\mu_{c^2p^2}^{\delta} - \left(\delta h_{t_{k+1}}\right)\mu_{cp}^{\delta} \\
& + \left(\mu_{cp}^{\delta} + \left(\delta h_{t_{k+1}}\right)\mu_c^{\delta} - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2p}^{\delta}\right) \left({}^{(3)}A_{t_k}^{\delta} + \left(\delta^{(2)}A_{t_k}^{\delta}\right)\right) \Big] \\
- 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^{+}\mu_c^{-} \Big[& \left({}^{(2)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{+}\right)\left({}^{(3)}A_{t_k}^{-} - {}^{(2)}A_{t_k}^{-}\right) + \frac{\mu_{cp}^{+}\mu_{cp}^{-}}{\mu_c^{-}\mu_c^{+}} \\
& - \frac{\mu_{cp}^{+}}{\mu_c^{+}}\left({}^{(3)}A_{t_k}^{-} - {}^{(2)}A_{t_k}^{-}\right) - \frac{\mu_{cp}^{-}}{\mu_c^{-}}\left({}^{(2)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{+}\right) \Big],
\end{aligned}
\tag{4.16}
$$

with $g_T = 0$.

The admissibility of the control pair $(L_{t_k}^{+,*}, L_{t_k}^{-,*})$, as defined in Proposition 4.2.1, can be formally established. This admissibility condition is ensured by the property that

$I_{t_k} \in \mathcal{F}_{t_k}$, and by the deterministic nature of the functions ${}^{(1)}A^{\pm}_{t_k}, {}^{(2)}A^{\pm}_{t_k}, {}^{(3)}A^{\pm}_{t_k}$. Up to this point, we have obtained a preliminary solution to the control problem (4.3) that satisfies the necessary condition specified by the Bellman equation (4.9). However, in order to establish its optimality, we present a verification theorem that rigorously confirms the optimality of the control pair $(L^{+,*}_{t_k}, L^{-,*}_{t_k})$ as the solution to the original control problem.

Theorem 4.2.3 (Verification Theorem)**.** *The optimal value function V_{t_k} for the control problem (4.3) can be expressed as:*

$$V_{t_k} = v(t_k, S_{t_k}, W_{t_k}, I_{t_k}),$$

where, for $t_k \in \mathcal{T}$,

$$v\left(t_k, s, w, i\right) = w + \alpha_{t_k} i^2 + si + h_{t_k} i + g_{t_k},$$

The coefficients α_{t_k} and h_{t_k} are determined according to Proposition 4.2.1, while g_{t_k} follows the expression given by Eq. (4.16). The control pair $(L^{+,*}_{t_k}, L^{-,*}_{t_k})$ given in Proposition 4.2.1 represents an optimal control strategy for the control problem (4.3).

The proof of Theorem 4.2.3 closely mirrors the derivation presented in Appendix A.3 of [12]. Interested readers are encouraged to consult that reference for further details.

4.2.2 Optimal Strategy with a General Adapted Fundamental Price Process

In this subsection, we relax the previous assumption of the fundamental price process $\{S_{t_k}\}_{t_k \in \mathcal{T}}$ being a martingale and instead consider a more general adapted process. Additionally, we introduce the assumption that, conditional on $\mathcal{F}_{t_k}$, the increments $\{S_{t_{j+1}} - S_{t_j}\}_{j \geq k}$ and the variables $(\mathbb{1}^+_{t_{k+1}}, \mathbb{1}^-_{t_{k+1}}, c^+_{t_{k+1}}, p^+_{t_{k+1}}, c^-_{t_{k+1}}, p^-_{t_{k+1}})$ are independent.

To facilitate our analysis, we introduce the notations:

$$\Delta_{t_k} := \mathbb{E}[S_{t_{k+1}} - S_{t_k} \mid \mathcal{F}_{t_k}].$$

and

$$\Delta^{t_k}_{t_j} := \mathbb{E}[\Delta_{t_j} \mid \mathcal{F}_{t_k}] = \mathbb{E}[S_{t_{j+1}} - S_{t_j} \mid \mathcal{F}_{t_k}], \quad j \geq k,$$

The quantity Δ_{t_k} represents the HFM's forecast regarding the change in the asset price during the interval $[t_k, t_{k+1})$, based on the available information at time t_k, while $\Delta^{t_k}_{t_j}$ represents the conditional expectation of the future price change in $[t_j, t_{j+1})$.

The incorporation of these forecasted price changes significantly enhances the flexibility of our model, enabling us to capture more realistic and diverse price dynamics. By considering both the HFM's immediate forecast Δ_{t_k} and the conditional expectations $\Delta^{t_k}_{t_j}$ over different time intervals, our framework becomes more comprehensive and adaptable. This broader perspective allows us to examine a wide range of scenarios and provides a robust foundation for studying the optimal control problem in a dynamic market environment.

The following theorem presents the optimal control strategy for the control problem (4.3) when considering a general fundamental price process. It highlights the impact of price

jump forecasts on the optimal placement spreads, contrasting it with the solution derived under the assumption of a martingale price process, as illustrated in Eq (4.10).

By relaxing the martingale price assumption, the theorem unveils the adjustments required in the optimal placement spreads to account for price jump forecasts. These forecasts introduce additional information that significantly influences the control strategy. Comparing the derived solution under the martingale assumption with the optimal control obtained in the general case sheds light on the changes in optimal behavior resulting from incorporating price jump forecasts into the model. The theorem provides valuable insights into the dynamic nature of the optimal control problem, capturing the impact of non-martingale price dynamics on the decision-making process.

Theorem 4.2.4 (Optimal Control with General Adapted Fundamental Price Process)**.** *The solution to the Bellman equation (4.9) with a general adapted fundamental price process are given, for* $k = 0, 1, \ldots, N$*, by*

$$\begin{aligned}
\widetilde{L}^{+,*}_{t_k} &= L^{+,*}_{t_k} + \frac{\beta^{+}_{t_k}}{2\gamma_{t_k}}\Delta_{t_k} + \left(\frac{\beta^{+}_{t_k}}{2\gamma_{t_k}}\right)\sum_{j=k+1}^{N}\prod_{\ell=k+1}^{j}\xi_\ell\Delta^{t_k}_{t_j}, \\
\widetilde{L}^{-,*}_{t_k} &= L^{-,*}_{t_k} - \frac{\beta^{-}_{t_k}}{2\gamma_{t_k}}\Delta_{t_k} - \left(\frac{\beta^{-}_{t_k}}{2\gamma_{t_k}}\right)\sum_{j=k+1}^{N}\prod_{\ell=k+1}^{j}\xi_\ell\Delta^{t_k}_{t_j},
\end{aligned} \tag{4.17}$$

where $L^{\pm,*}_{t_k}$ *is the optimal control with a martingale price process as derived in Proposition 4.2.1.* $\beta^{\pm}_{t_k}$ *and* γ_{t_k} *are the deterministic sequences consistent with Eq.(4.13)-(4.12), and the quantity* ξ_k *is defined as:*

$$\begin{aligned}
\xi_k = 1 &+ \frac{\alpha_{t_{k+1}} - \phi}{\gamma_{t_k}}\sum_{\delta=\pm}\pi^{\delta}_{t_{k+1}}\beta^{\delta}_{t_k}\left[\frac{\beta^{\delta}_{t_k}}{\gamma_{t_k}}\left((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\right) + 2\mu^{\delta}_{c}\right] \\
&+ 2\frac{(\alpha_{t_{k+1}} - \phi)^2}{\gamma^2_{t_k}}\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}\beta^{+}_{t_k}\beta^{-}_{t_k}.
\end{aligned} \tag{4.18}$$

Remark 4.2.2. *In the Appendix B.1, we demonstrate that when the fundamental price follows a general adapted process, the proposed ansatz for the value function V_{t_k} is given by:*

$$V_{t_k} = v\left(t_k, S_{t_k}, W_{t_k}, I_{t_k}\right) := W_{t_k} + \alpha_{t_k} I_{t_k}^2 + S_{t_k} I_{t_k} + \widetilde{h}_{t_k} I_{t_k} + \widetilde{g}_{t_k}. \tag{4.19}$$

Here, $\alpha : \mathcal{T} \to \mathbb{R}$ is a deterministic function, same as the case of the martingale price assumption in (4.14). Additionally, $\{\widetilde{h}_{t_k}\}_{t_k \in \mathcal{T}}$ and $\{\widetilde{g}_{t_k}\}_{t_k \in \mathcal{T}}$ are dependent on $\{\Delta_{t_k}, \Delta_{t_j}^{t_k}\}_{j \geq k}$ and adapted to the filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$. The specific expressions for $\widetilde{h}$ and $\widetilde{g}$ can be found in Eqs. (B.17)-(B.18), utilizing the notation provided in (B.2). The proof of the corresponding verification theorem follows a similar approach to the proof of Theorem 4.2.3.

The optimal placement strategies at time t_k, considering a non-martingale dynamics for the fundamental price process, can be expressed as follows:

$$\widetilde{a}_{t_k}^* = S_{t_k} + \widetilde{L}_{t_k}^{+,*}, \qquad \widetilde{b}_{t_k}^* = S_{t_k} - \widetilde{L}_{t_k}^{-,*}. \tag{4.20}$$

Here, $\widetilde{a}_{t_k}^*$ represents the price for the ask LO, while $\widetilde{b}_{t_k}^*$ represents the price for the bid LO. Equation (4.17) emphasizes that we can decompose the task of determining the optimal trading strategy into two subproblems.

First, we compute the recursive expressions (4.11)-(4.15). This computation is performed "offline" at the onset of each trading day. In other words, all the parameters required to calculate $L_{t_k}^{\pm,*}$ are predetermined at the beginning of the day.

Second, we address the forecasting problem by determining $\{\Delta_{t_j}^{t_k}\}_{j=k,\ldots,N}$ and calculating $\widetilde{L}_{t_k}^{+,*}$ using the expression for $L_{t_k}^{+,*}$ as given in Eq. (4.17). This step is carried out "online" at each time point t_k. Consequently, under a general adapted fundamental price process,

the optimal strategy $\widetilde{L}^{+,*}_{t_k}$ incorporates the perspectives of the HFM regarding changes in the fundamental based on the information available to her at time t_k.

Next proposition provides necessary conditions for a positive bid-ask spread of the HFM's optimal placement.

Proposition 4.2.5 (Conditions for a optimal Positive Spread)**.** *Under both martingale and non-martingale price processes, the optimal placement strategy yields positive spreads at all times (i.e., $a_{t_k} > b_{t_k}$, for all $k = 0, \ldots, N$), provided that the following three conditions hold:*

1. *The first and second conditional moments of $c^{\pm}$ defined in Eq. (4.2) satisfy*

$$\mu_c := \mu_c^+ = \mu_c^-, \quad \mu_{c^2} := \mu_{c^2}^+ = \mu_{c^2}^-. \tag{4.21}$$

2. *Buy and sell MOs arrive with the same probability:*

$$\pi^+_{t_{k+1}} = \pi^-_{t_{k+1}}. \tag{4.22}$$

3. *The conditional expectations of $(cp)^{\pm}$ and $(c^2p)^{\pm}$ defined in Eq. (4.2) satisfy*

$$\mu^{\pm}_{cp} = \mu^{\pm}_c \mu^{\pm}_p, \quad \mu^{\pm}_{c^2p} = \mu^{\pm}_{c^2} \mu^{\pm}_p. \tag{4.23}$$

The conditions presented as (4.21) and (4.22) establish a symmetric market framework. Specifically, under the condition (4.21), the mean and variance of the bid demand slope are identical to those on the ask side. Furthermore, condition (4.22) posits that the arrival probabilities of buy and sell MOs are equal within each time interval $[t_{t_k}, t_{k+1})$.

Additionally, condition (4.23) assumes that the demand slope and the reservation price are uncorrelated. These assumptions are supported by empirical evidence.

4.3 Properties of the Optimal Placement Strategies

In this section, we will discuss the behavior of the optimal placement strategies and their sensitivities to model parameters, such as the arrival rate $\pi_{t_k}(1,1)$, the inventory level I, and the penalty λ on the terminal inventory.

4.3.1 Sensitivity to Simultaneous Arrival Probability

Case $\pi_{t_k}(1,1) \equiv 0$. We first focus on the case of $\pi_{t_k}(1,1) \equiv 0$, which means that only one type of MOs, either buy or sell, would arrive during each subinterval $[t_k, t_{k+1})$. Denote the optimal ask and bid price under this condition as $\widetilde{a}^{*,0}_{t_k}$ and $\widetilde{b}^{*,0}_{t_k}$, respectively. Following from Eq.(4.20), $\widetilde{a}^{*,0}_{t_k}$ and $\widetilde{b}^{*,0}_{t_k}$ are given by:

$$\widetilde{a}^{*,0}_{t_k} = S_{t_k} + \overbrace{\frac{(\alpha^0_{t_{k+1}} - \phi)\mu^+_c}{\mu^+_c - (\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2}} I_{t_k} + \frac{\mu^+_{cp} - 2(\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2p}}{2[\mu^+_c - (\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2}]} + \frac{(\Delta_{t_k} + \widetilde{h}^0_{t_k})\mu^+_c}{2[\mu^+_c - (\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2}]}}^{\widetilde{L}^{+,*,0}_{t_k}}, \tag{4.24}$$

$$\widetilde{b}^{*,0}_{t_k} = S_{t_k} + \overbrace{\frac{(\alpha^0_{t_{k+1}} - \phi)\mu^-_c}{\mu^-_c - (\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2}} I_{t_k} - \frac{\mu^-_{cp} - 2(\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2p}}{2[\mu^-_c - (\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2}]} + \frac{(\Delta_{t_k} + \widetilde{h}^0_{t_k})\mu^-_c}{2[\mu^-_c - (\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2}]}}^{-\widetilde{L}^{-,*,0}_{t_k}}, \tag{4.25}$$

where

$$\alpha^0_{t_k} = (\alpha^0_{t_{k+1}} - \phi) + \sum_{\delta=\pm} \pi^{\delta}_{t_{k+1}} \frac{\left((\alpha^0_{t_{k+1}} - \phi)\mu^{\delta}_c\right)^2}{\mu^{\delta}_c - (\alpha^0_{t_{k+1}} - \phi)\mu^{\delta}_{c^2}},$$

$$\widetilde{h}^0_{t_k} = h^0_{t_k} + \sum_{j=k+1}^{N} \prod_{\ell=k+1}^{j} \xi^0_\ell \Delta^{t_k}_{t_j},$$

and $h^0_{t_k}, \xi^0_k$ are given by Eq. (4.15) and Eq. (4.18) respectively, setting $\pi_{t_{k+1}}(1,1) = 0$ therein.

Remark 4.3.1 (A Weaker Condition For a Positive Spread ($\widetilde{a}^{*,0}_{t_k} > \widetilde{b}^{*,0}_{t_k}$))**.** *In the case of $\pi_{t_k}(1,1) \equiv 0$, only the first symmetry condition (4.21) in Proposition 4.2.5 is needed to guarantee a positive spread:*

$$\widetilde{a}^{*,0}_{t_k} - \widetilde{b}^{*,0}_{t_k} = \widetilde{L}^{+,*,0}_{t_k} + \widetilde{L}^{-,*,0}_{t_k} = \frac{\mu^+_{cp} - 2(\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2p}}{2\left(\mu^+_c - (\alpha^0_{t_{k+1}} - \phi)\mu^+_{c^2}\right)} + \frac{\mu^-_{cp} - 2(\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2p}}{2\left(\mu^-_c - (\alpha^0_{t_{k+1}} - \phi)\mu^-_{c^2}\right)} > 0.$$

It follows from $\alpha^0_{t_k} < 0$ as shown in Lemma 4.2.2.

The inventory level affects prices of the HFM's LOs through the second terms in Eqs.(4.24)-(4.25) (i.e., $\frac{(\alpha^0_{t_{k+1}} - \phi)\mu^{\pm}_c}{\mu^{\pm}_c - (\alpha^0_{t_{k+1}} - \phi)\mu^{\pm}_{c^2}} I_{t_k}$). This term determines the shadow cost of inventory at time t_k. The coefficient of I_{t_k} can be written as

$$\frac{(\alpha^0_{t_{k+1}} - \phi)\mu^{\pm}_c}{\mu^{\pm}_c - (\alpha^0_{t_{k+1}} - \phi)\mu^{\pm}_{c^2}} = \frac{(\alpha^0_{t_{k+1}} - \phi)}{1 - (\alpha^0_{t_{k+1}} - \phi)\mu^{\pm}_c - \alpha^0_{t_{k+1}} \operatorname{Var}\left(c^{\pm}_{t_{k+1}} \mid \mathcal{F}_{t_k}\right)/\mu^{\pm}_c}.$$

Case $\pi_{t_k}(1,1) \not\equiv 0$. The occurrence of simultaneous arrivals of buy and sell MOs during a single time step, denoted as $\pi_{t_k}(1,1)$, is generally low in high-frequency trading settings, such as intervals of 1 second or less. In our empirical analysis presented in a later section, we have observed that $\pi_{t_k}(1,1)$ is approximately 0.05 for a trading period of 1 second. However, this situation changes when the trading frequency is reduced, such as in the case of intervals of 5 seconds or longer. In such scenarios, it becomes crucial to account for the

possibility of joint arrivals. To further explore the behavior of optimal placement under the conditions outlined in (4.21)-(4.23), along with additional conditions, we present the following corollary, which provides valuable insights.

Corollary 4.3.1. *Under Assumptions (4.21)-(4.22), the optimal spreads are*

- *invariant to the local drifts* $\{\Delta_{t_k}\}_{k=0,\dots,N}$*;*
- *independent on the inventory level.*

Suppose that, in addition to (4.21)-(4.22), the condition (4.23) as well as the following conditions hold:

$$\mu_c^2 = \mu_{c^2}, \quad \pi_{t_k}^{\pm} \equiv \pi^{\pm}, \quad \pi_{t_k}(1,1) \equiv \pi(1,1), \tag{4.26}$$

for some constants $\pi^{\pm} \in (0,1)$ *and* $\pi(1,1) \in [(\pi^+ + \pi^- - 1) \vee 0, \pi^+ \wedge \pi^-]$*. Then,*

- *the monotonicity of the optimal spreads over time is determined by the value of* ϕ*. And if* $\pi(0,1) = \pi(1,0) = 0$*, the optimal spreads remain constant throughout the trading horizon, exhibiting a flat behavior.*
- *the optimal spreads display a decreasing trend, when considering* $\pi(1,1)$ *at a specific time point.*

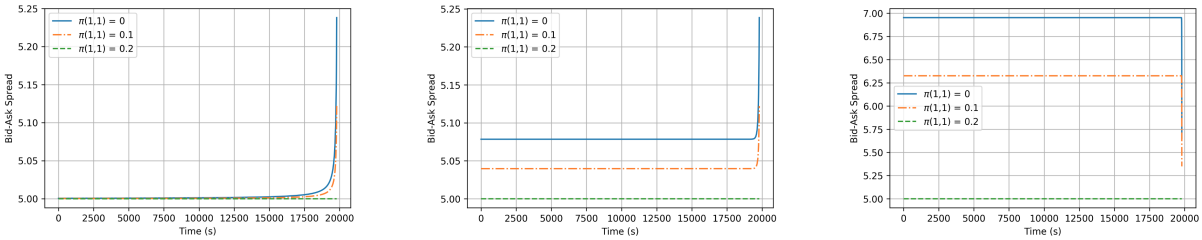
Although the HFM's inventory level and price drift play a role in determining the optimal bid and ask prices, it is interesting to note that the difference between the optimal ask and bid prices, also known as the optimal spread, remains unaffected by these factors. In other words, the spread does not depend on the specific inventory level or price drift considered.

The optimal spread exhibits distinct behavior depending on the value of the running inventory penalty, ϕ, as depicted in Figure (4.2). When ϕ is small, resulting in a relatively mild penalty for maintaining a running inventory, the optimal spread displays a non-decreasing pattern over time, escalating notably towards the conclusion of the trading horizon in response to penalties on terminal inventory. A broader spread empowers a HFM to regulate her inventory by executing more pronounced trades on a specific side of the LOB. The augmented spread accommodates the increased likelihood of round-trip transactions between consecutive actions, particularly as $\pi(1,1)$ rises.

Conversely, when ϕ attains a large value, signifying a substantial penalty on running inventory, the optimal spread demonstrates a non-increasing trend with time. Towards the termination of the trading horizon, it experiences a marked decline. The HFM, faced with a higher penalty for maintaining inventory positions, opts for a more risk-averse stance, reflected in the larger the bid-ask spread from the beginning. As discussed earlier, a wider spread empowers the HFM to exercise greater control over her inventory level. While the bid-ask spread experiences a significant decrease towards the end of the trading horizon, it remains notably larger compared to scenarios with a smaller ϕ.

4.3.2 Sensitivity to Inventory Level

In this section, we perform a sensitivity analysis to examine the impact of inventory levels I_{t_k} on the optimal prices. Specifically, we consider the scenario where both buy and sell MOs arrive simultaneously (i.e., $\pi(1,1) > 0$). The following corollary investigates the relationship between inventory levels and the monotonic behavior of the optimal prices. It sheds light on how the prices respond to changes in inventory.



(a) $\phi = 0$ (b) $\phi = 1 \times 10^{-6}$ (c) $\phi = 1 \times 10^{-3}$

Figure 4.2: Optimal bid-ask spread for different values of running inventory penalty ϕ under symmetric market conditions. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, we set $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1)$ ranges from 0 to $\pi^{\pm}$.

Corollary 4.3.2. *The optimal ask price $\widetilde{a}_{t_k}^*$ and the bid price $\widetilde{b}_{t_k}^*$ as defined in Eq. (4.20), are strictly decreasing with inventory I_{t_k}.*

Corollary 4.3.2 underscores the persistence of the decreasing characteristic in optimal prices concerning inventory, irrespective of the specific value assigned to ϕ. However, the parameter ϕ does exert influence on the behavior of the ask/bid spread. In instances where ϕ is negligible or equal to zero, and for a positive inventory level, both bid and ask prices undergo a decremental adjustment, facilitating the orderly sale of the HFM shares. Conversely, when the HFM holds a negative inventory, there is an inclination to elevate bid and ask prices, a strategic move to procure additional shares.

Contrarily, in scenarios where ϕ assumes a relatively larger value, the observed behavior is inverted. For a positive inventory, bid and ask prices experience an upward adjustment, aimed at encouraging share sales. Similarly, in the case of a negative inventory, the HFM opts to lower bid and ask prices, aligning with a strategy focused on accumulating more shares.

In Figure 4.3, the dynamics of optimal bid and ask spreads are illustrated over the final 500 seconds within a 5.5-hour trading interval, considering a spectrum of inventory levels from -1000 to 1000. As in Figure 4.3(a)-(b), when ϕ is small or zero, in accordance with Eqs. (4.24)-(4.25), both optimal ask and bid spreads consistently hover around $\mu_p^+/2$ and $\mu_p^-/2$, respectively, for the majority of the trading period. This observation holds true irrespective of the specific inventory level.

Nevertheless, as the terminal time approaches, the influence of inventory levels on the optimal bid and ask spread becomes more pronounced. Figure 4.3 reveals two discernible strategies contingent on varying inventory levels:

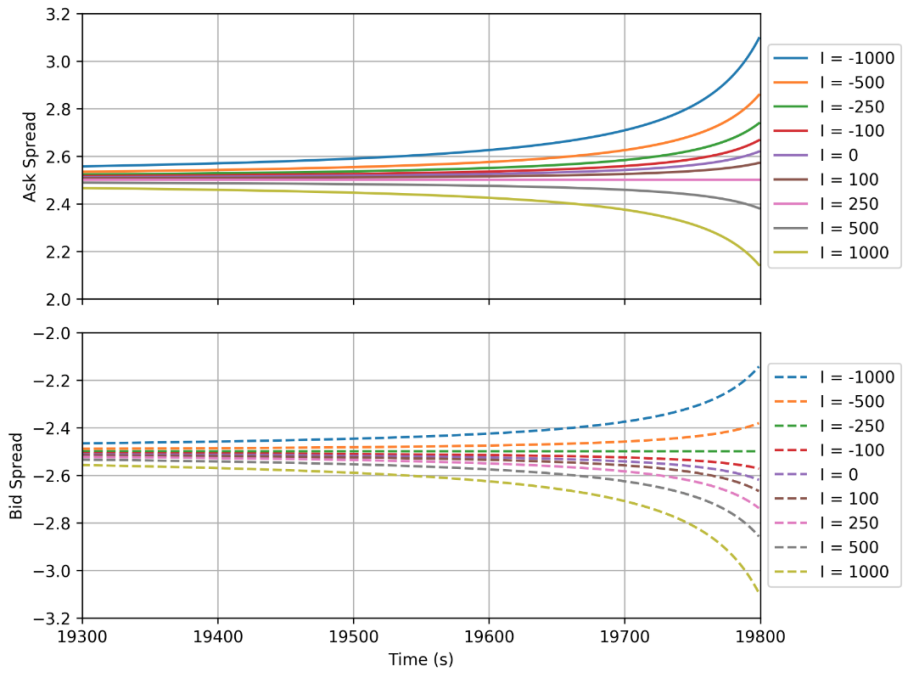
- For low inventory levels (e.g., between -250 to 250 shares), both ask and bid spreads exhibit an increasing trend over time. Consequently, both selling and buying activities are subdued as the trading period concludes. This approach enables the HFM to sustain a low inventory level until the closing moments.
- Conversely, in the scenario of highly positive inventory levels (e.g., surpassing 250 shares), the ask spread adopts a decreasing trajectory with time. This pattern reflects the HFM's heightened inclination to sell more shares as time progresses. Similarly, if the inventory level is markedly negative (e.g., falling below -250 shares), the bid spread decreases over time.

When ϕ assumes a large value, as illustrated in Figure 4.3(c), the following observations, contrary to the earlier scenario, are noted:

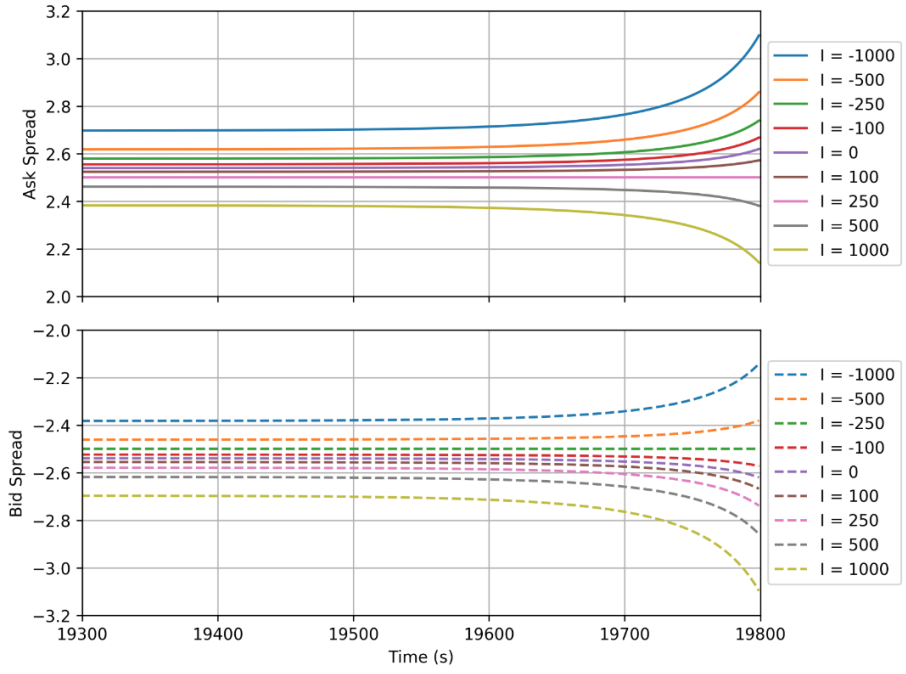
- For low inventory levels (e.g., ranging from -250 to 250 shares), both ask and bid spreads exhibit a decreasing trend over time. Notably, this decrease is from a much higher spread compared to the low ϕ value case. Despite the continuous

decrease, the spreads remain higher than those observed for the low ϕ value case. Consequently, both selling and buying activities are restrained as the trading period concludes, allowing the HFM to maintain a low inventory level until the closing moments.

- Conversely, in the case of highly positive inventory levels (e.g., exceeding 250 shares), the ask spread follows an increasing trajectory with time. However, this increase is from a much lower spread compared to the low ϕ value case. Even with the overall increase, the spreads remain lower than those observed for the low ϕ value case. This pattern signifies the HFM's heightened inclination to sell more shares as time progresses. Similarly, if the inventory level is significantly negative (e.g., falling below -250 shares), the bid spread increases over time.

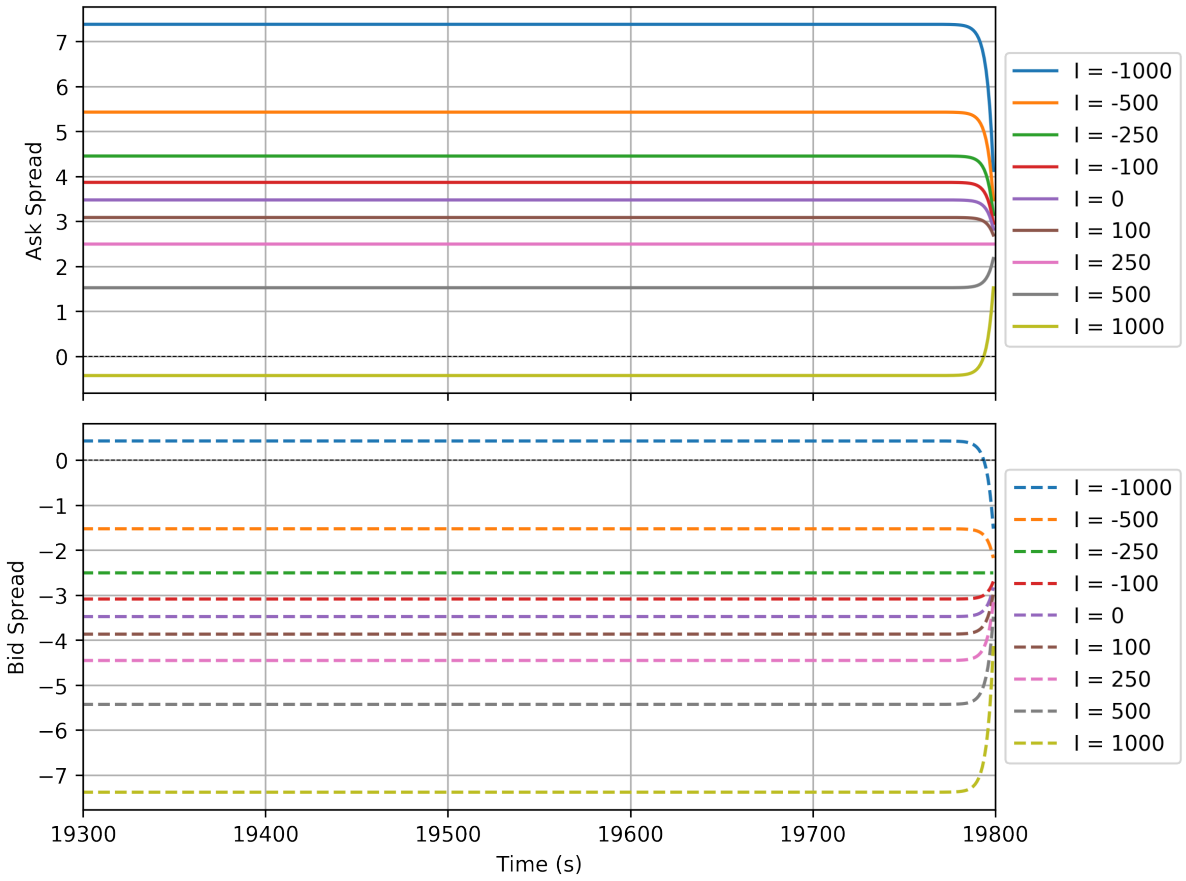


(a) $\phi = 0$



(b) $\phi = 1 \times 10^{-6}$

Figure 4.3: Optimal spreads in the last 500-seconds with various inventory levels for different values of ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, we set $\lambda = 0.0005$, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} = 0.2$, $\pi(1,1) = 0$ and $\Delta_{t_k} = 0$ for $k = 0, \ldots, N$.



(c) $\phi = 1 \times 10^{-3}$

Figure 4.3: Optimal spreads in the last 500-seconds with various inventory levels for different values of ϕ. (cont.)

4.3.3 Sensitivity to Terminal Inventory Penalty

In this section, we explore the relationship between the monotonicity of the optimal bid and ask spread and the inventory level over time. Specifically, we examine how the monotonic behavior changes as the inventory level varies. We provide an analytical expression for the inventory threshold and investigate the impact of different terminal inventory penalty levels and $\pi(1,1)$ values on the monotonicity of the optimal bid and ask spread with time. The following corollary focuses on the scenario of $\pi_{t_{k+1}}(1,1)$ and provides further insights into the relationship. There are similar results in [12].

Corollary 4.3.3. *Under the following assumptions:*

- *the market is symmetric, i.e., conditions (4.21)-(4.23) of the Proposition 4.2.5 hold and that* $\mu_p := \mu_p^+ = \mu_p^-$,
- $\pi(1,1) = 0$, *i.e., only one type of MOs can arrive during each subinterval,*
- *the fundamental price is a martingale process,*

there exists a threshold for the inventory level,

$$\bar{I}^{\pm} = \pm \frac{\mu_{c^2}\mu_p}{2\mu_c},$$

such that the following statements hold for any penalty level $\lambda > 0$*:*

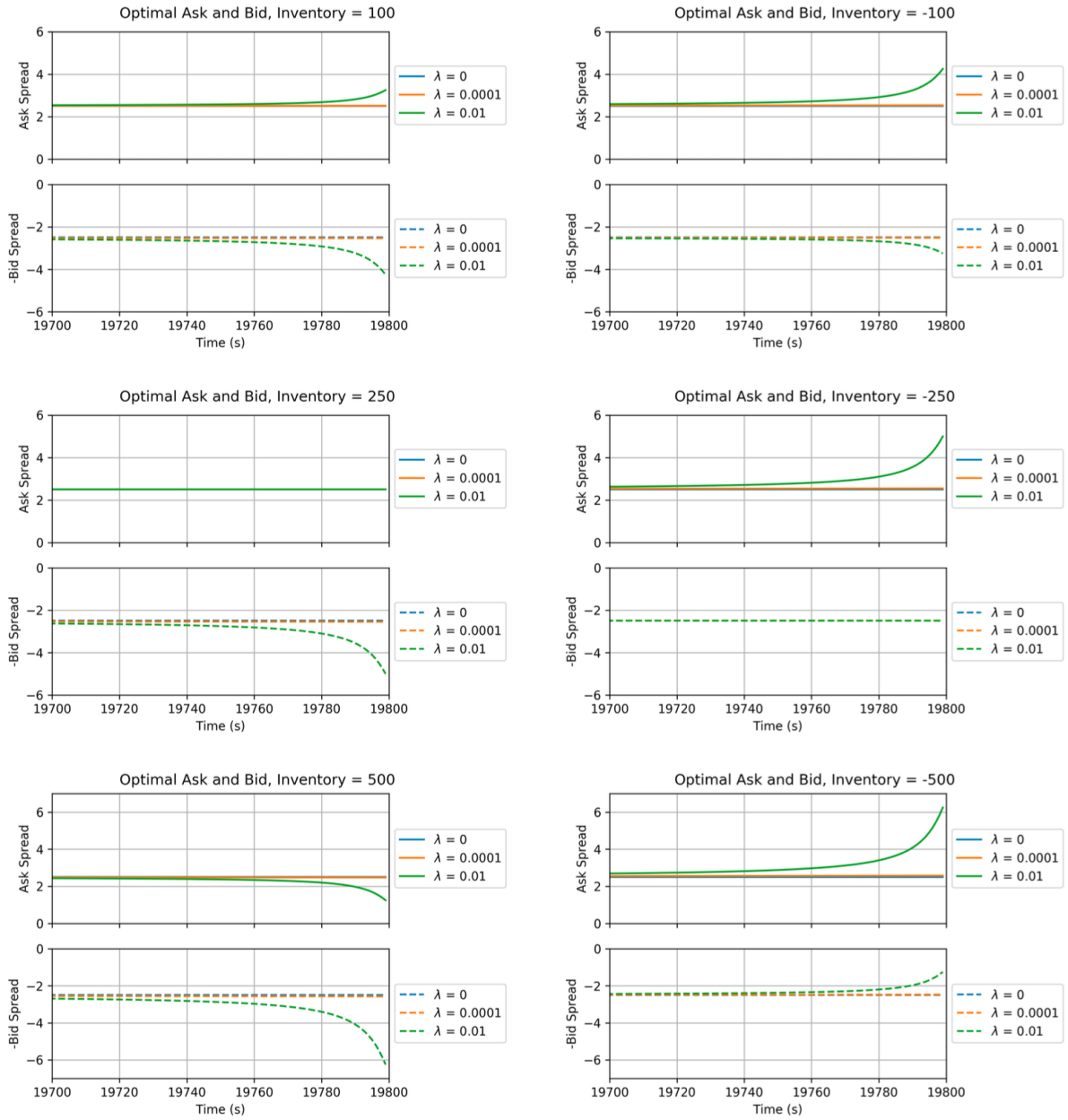
- *When the inventory level* $I_{t_k} = \bar{I}^+(\bar{I}^-)$*, the optimal ask (bid) spread stays at* $S_{t_k} + \mu_p/2(S_{t_k} - \mu_p/2)$ *as time goes to* T*;*
- *When the inventory level* $I_{t_k} \in (\bar{I}^-, \bar{I}^+)$*, the optimal ask and bid spreads increase as time goes to* T*;*

- *When the inventory level $I_{t_k} > \bar{I}^+ (I_{t_k} < \bar{I}^-)$, the optimal ask (bid) spread decreases, while the optimal bid (ask) spread increases as time goes to T.*

When there is no terminal inventory penalty, the optimal bid and ask spreads maintain a consistent level throughout the trading day, irrespective of the inventory level, as shown in Figure 4.4.

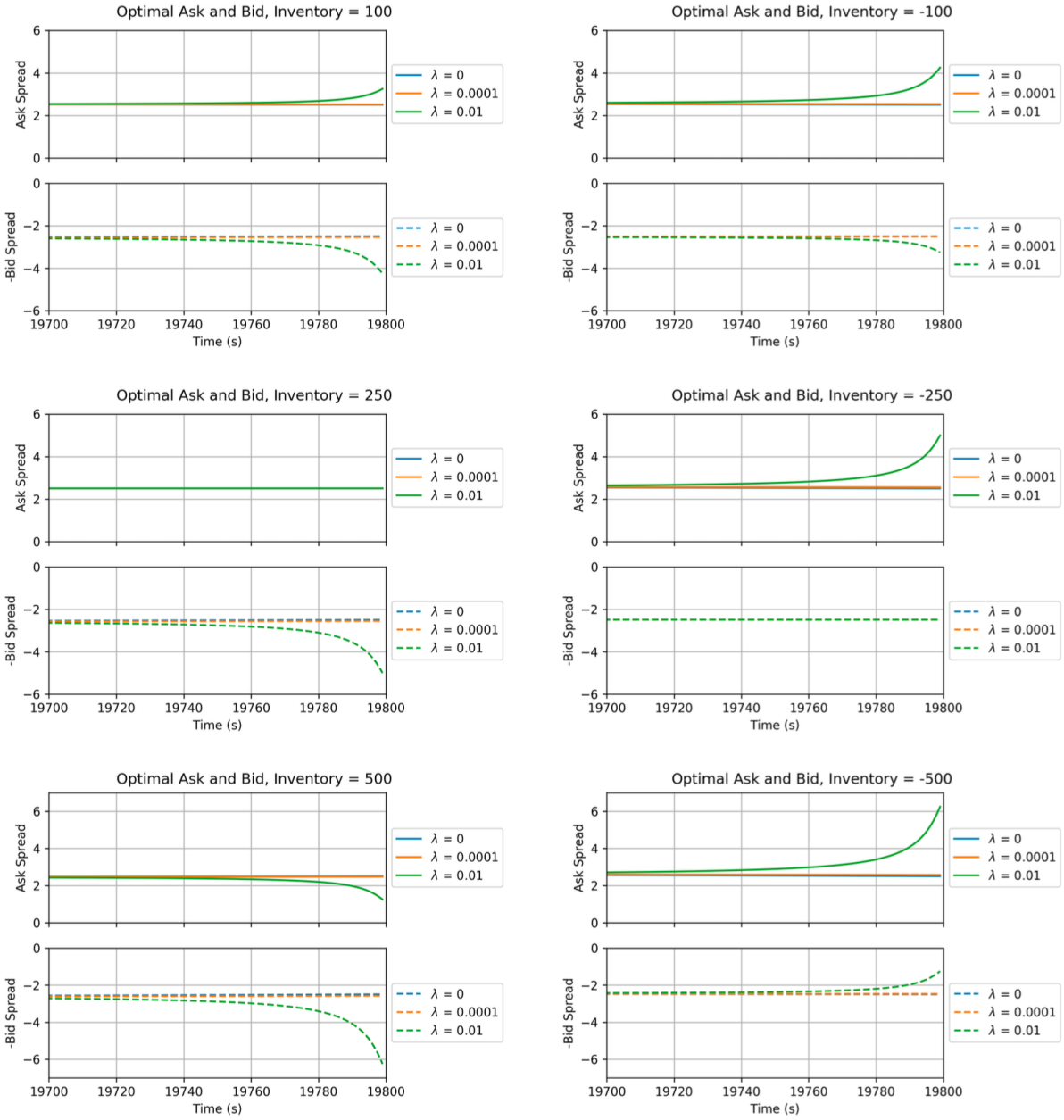
As the terminal inventory penalty increases, a HFM with a positive inventory position consistently positions bid LOs deeper to deter purchases. On the ask side, three strategies are viable: (1) placing ask LOs deeper into the book, (2) positioning ask LOs closer to S_{t_k}, and (3) maintaining the ask spread unchanged from its current value. According to Corollary 4.3.3, the selection among these strategies hinges on the relationship between the current inventory level I_{t_k} and the inventory threshold value $\bar{I}^+$.

The left panel of Figure 4.4 depicts optimal placements for positive inventory levels with thresholds $\bar{I}^{\pm} = \pm 250$. For an inventory position below 250 shares, the HFM strategically positions both bid and ask orders deeper into the book as time approaches T to maintain the current inventory level. The magnitude of the penalty on the terminal inventory dictates the depth of LO placement on both sides of the book. In the left panel of the middle row in Figure 4.4, when the inventory level equals the threshold, the optimal bid delves even deeper into the book at the next time step, while the ask spread remains constant, regardless of the penalty's magnitude. For inventory levels surpassing the threshold, the HFM narrows the ask spread to execute more ask LOs and, consequently, reduce the inventory level. The higher the inventory, the closer she positions the ask quote to the price S_{t_k}. Analogous analyses are applicable to negative inventory positions, as demonstrated in the right panel of Figure 4.4.



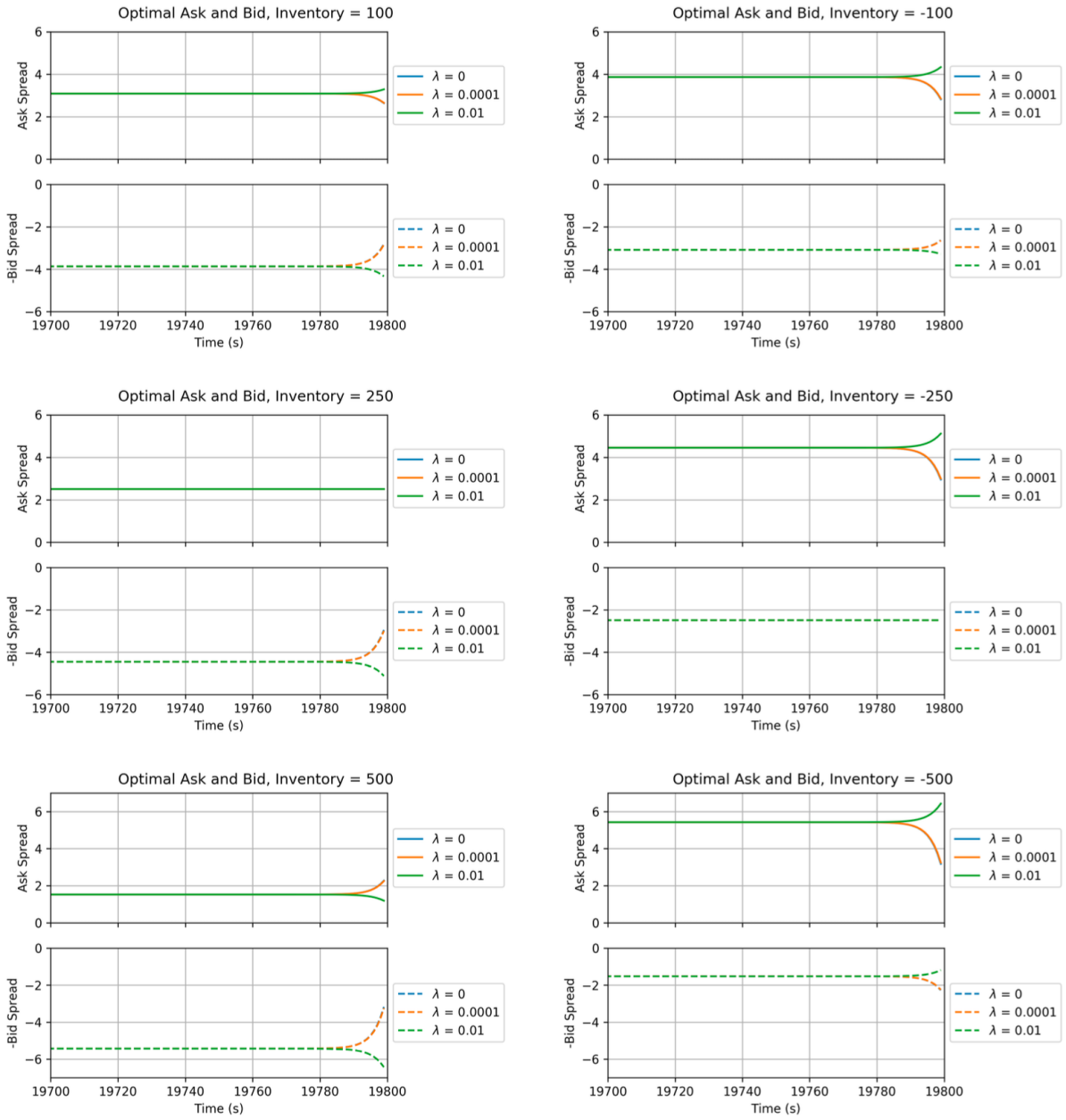
(a) $\phi = 0$

Figure 4.4: Optimal trading strategies in the final 100 seconds: impact of inventory, terminal inventory penalty λ, and running inventory penalty ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1) = 0$ and $\Delta t_k = 0$ for $k = 0, \ldots, N$.



(b) $\phi = 1 \times 10^{-6}$

Figure 4.4: Optimal trading strategies in the final 100 seconds: impact of inventory, terminal inventory penalty λ, and running inventory penalty ϕ. (cont.)

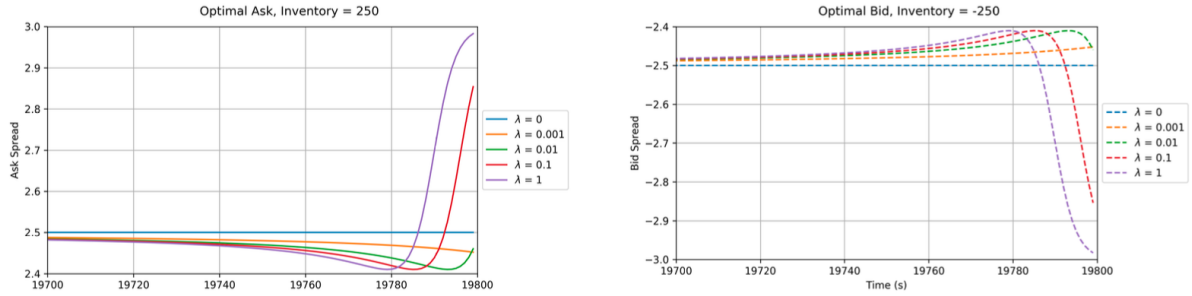


(c) $\phi = 1 \times 10^{-3}$

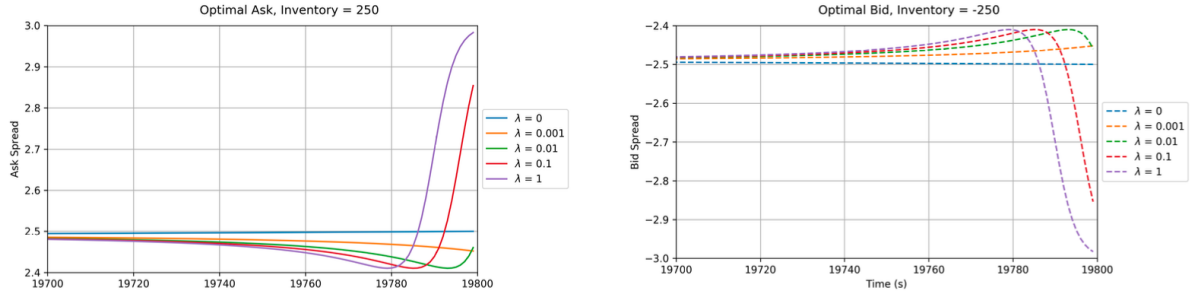
Figure 4.4: Optimal trading strategies in the final 100 seconds: impact of inventory, terminal inventory penalty λ, and running inventory penalty ϕ. (cont.)

The optimal strategies exhibit significant differences when joint arrivals are permitted. The insights from Figure 4.5 underscore that when $\pi_{t_k}(1,1) > 0$, there exists no specific inventory threshold yielding a constant optimal bid or ask spread across various penalty levels, irrespective of the value of running inventory penalty ϕ. Moreover, at certain intermediate inventory levels, the optimal spreads under a substantial terminal inventory penalty λ deviate from a monotonic temporal pattern.

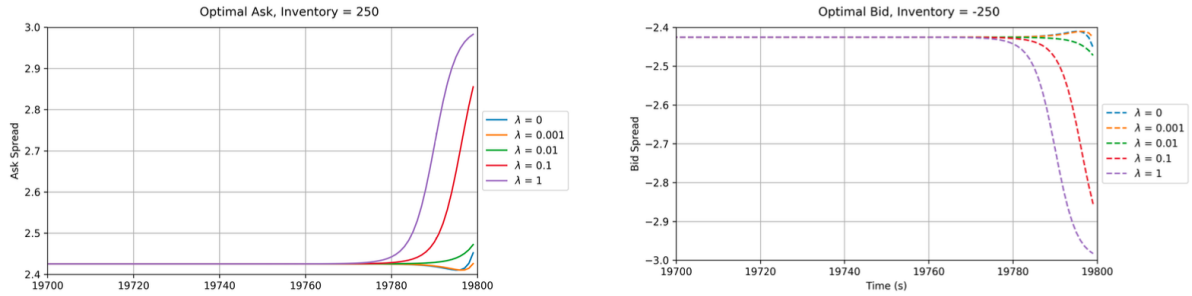
To elucidate the non-monotonicity observed, let's consider the left panel of Figure 4.5. When a higher penalty is imposed on the terminal inventory position, a logical response is for the HFM tends to adopt a narrower ask spread. This adjustment aims to facilitate more sales of her shares as the terminal time approaches, with the goal of attaining a lower inventory position. However, as time approaches the very limit of the terminal time T, a shift in strategy becomes apparent. It becomes more crucial to profit directly from fewer but wider roundtrip transactions. This strategy helps compensate for losses from the terminal inventory cost while simultaneously maintaining the current intermediate inventory level. Such a strategic adaptation is deemed reasonable since $\pi_{t_k}(1,1) > 0$, indicating a higher likelihood of roundtrip transactions occurring.



(a) $\phi = 0$



(b) $\phi = 1 \times 10^{-6}$



(c) $\phi = 1 \times 10^{-3}$

Figure 4.5: Optimal trading strategies in the final 100 seconds when $\pi_{t_k}(1,1) > 0$ for different values of ϕ. The agent's actions are executed every second, spanning from 0 to 19800 seconds. λ ranges from 0 to 1. The plots are generated based on specific parameters ensuring the fulfillment of Conditions (4.21)-(4.23) and (4.26). Specifically, $\mu_c^{\pm} = 100$, $\mu_p^{\pm} = 5$, $\mu_{cp}^{\pm} = 500$, $\mu_{c^2}^{\pm} = 1 \times 10^4$, $\mu_{c^2p}^{\pm} = 5 \times 10^4$, $\pi_{t_k}^{+} = \pi_{t_k}^{-} \equiv 0.2$. $\pi(1,1) = 0.05$ and $\Delta t_k = 0$ for $k = 0, \ldots, N$.

Chapter 5

Conclusion and Future Work

In this dissertation, we address the market making problem by incorporating considerations of latency and running inventory control. Our approach involves leveraging the optimal control framework alongside advanced RL techniques.

For the latency-inclusive market making scenario, we initially study a stochastic control problem. Herein, we reformulate the problem into a finite MDP, providing explicit characterizations of immediate rewards for each temporal period and subsequently numerically solve the formulated problem. Our contributions extend to dimensionality reduction, wherein we effectively diminish the number of state variables from 7 to 5. Furthermore, we empirically assess the efficacy of optimal placement strategies. Throughout this investigation, we introduce and emphasize the significance of two concepts: the number of filled shares and the value of order with partial fill. These concepts prove instrumental in articulating demand functions and immediate rewards. As a complementary avenue, we adopt RL techniques, systematically evaluating the performance of various algorithms.

In the realm of market making with running inventory, this work serves as an extension of the study [12]. Our contribution builds upon the existing framework by introducing a new dimension: running inventory penalties. This addition, extending beyond terminal

inventory control, is strategically designed to enhance the management of inventory risk, fostering greater efficiency in market making strategies.

Future research avenues within the optimal control framework for addressing market making with latency include:

- Investigate a stochastic delay problem, introducing variability in latency values rather than assuming constant latency.
- Extend beyond the confines of a random linear demand function and explore alternative forms of demand functions.
- Develop an analytical expression for the value function under specific assumptions, potentially utilizing techniques like quadratic approximation.
- Explore the efficacy of neighborhood search method to derive suboptimal strategies, offering insights into robustness and adaptability of market making algorithms.

For future investigations into the application of RL methods in addressing market making with latency, the following pivotal areas could be explored:

- Increase training and testing episodes in the current simulation. The limitations in time and computational resources have constrained the extent of agent training. Expanding these episodes would allow for a more comprehensive exploration of the agent's learning capabilities and convergence patterns.
- Experiment with more sophisticated RL algorithms to assess their efficacy.
- Evaluate the performance of these RL methods using authentic LOB data.
- Implement nonlinear VFA techniques, particularly leveraging deep neural networks.

References

[1] T. Adrian, A. Capponi, M. Fleming, E. Vogt, and H. Zhang. Intraday market making with overnight inventory costs. *Journal of Financial Markets*, 50:100564, 2020.

[2] M. M. Afsar, T. Crump, and B. Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7):1–38, 2022.

[3] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

[4] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

[5] K. Bechler and M. Ludkovski. Order flows and limit order book resiliency on the meso-scale. *Market Microstructure and Liquidity*, 3(03n04):1850006, 2017.

[6] R. Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[7] D. Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

[8] T. Beysolow II and T. Beysolow II. Market making via reinforcement learning. *Applied Reinforcement Learning with Python: With OpenAI Gym, Tensorflow, and Keras*, pages 77–94, 2019.

[9] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

[10] J. Brogaard, T. Hendershott, and R. Riordan. High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306, 2014.

[11] B. Bruder and H. Pham. Impulse control problem on finite horizon with execution delay. *Stochastic Processes and their Applications*, 119(5):1436–1469, 2009.

[12] A. Capponi, J. E. Figueroa-López, and C. Yu. Market making with stochastic liquidity demand: Simultaneous order arrival and price change forecasts. *arXiv preprint arXiv:2101.03086*, 2021.

[13] A. Cartea, R. Donnelly, and S. Jaimungal. Enhancing trading strategies with order book signals. *Applied Mathematical Finance*, 25(1):1–35, 2018.

[14] Á. Cartea and S. Jaimungal. Risk metrics and fine tuning of high-frequency trading strategies. *Mathematical Finance*, 25(3):576–611, 2015.

[15] Á. Cartea, S. Jaimungal, and J. Penalva. *Algorithmic and high-frequency trading*. Cambridge University Press, 2015.

[16] Á. Cartea, S. Jaimungal, and J. Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014.

[17] Á. Cartea, S. Jaimungal, and L. Sánchez-Betancourt. Latency and liquidity risk. *International Journal of Theoretical and Applied Finance*, 24(06n07):2150035, 2021.

[18] Á. Cartea and L. Sanchez-Betancourt. The shadow price of latency: Improving intraday fill ratios in foreign exchange markets. *SIAM Journal on Financial Mathematics*, 12(1):254–294, 2021.

[19] Á. Cartea and L. Sánchez-Betancourt. Optimal execution with stochastic delay. *Finance and Stochastics*, 27(1):1–47, 2023.

[20] N. T. Chan and C. Shelton. An electronic market-maker. 2001.

[21] R. Cont. Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine*, 28(5):16–25, 2011.

[22] R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of financial econometrics*, 12(1):47–88, 2014.

[23] R. Donnelly. Optimal execution: A review. *Applied Mathematical Finance*, 29(3):181–212, 2022.

[24] X. Gao and Y. Wang. Optimal market making in the presence of latency. *Quantitative Finance*, 20(9):1495–1512, 2020.

[25] B. Gašperov and Z. Kostanjčar. Market making with signals through deep reinforcement learning. *IEEE Access*, 9:61611–61622, 2021.

[26] L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

[27] O. Guéant. *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, volume 33. CRC Press, 2016.

[28] O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Dealing with the inventory risk: a solution to the market making problem. *Mathematics and financial economics*, 7:477–507, 2013.

[29] F. Guilbaud and H. Pham. Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, 13(1):79–94, 2013.

[30] T. Hendershott, C. M. Jones, and A. J. Menkveld. Does algorithmic trading improve liquidity? *The Journal of finance*, 66(1):1–33, 2011.

[31] T. Ho and H. R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47–73, 1981.

[32] A. J. Kim, C. R. Shelton, and T. Poggio. Modeling stock order flows and learning market-making from data. 2002.

[33] J. Kober, J. A. Bagnell, and J. Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[34] P. Kumar. Deep reinforcement learning for market making. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1892–1894, 2020.

[35] C.-A. Lehalle and O. Mounjid. Limit order strategic placement with adverse selection risk and the role of latency. *Market Microstructure and Liquidity*, 3(01):1750009, 2017.

[36] Y.-S. Lim and D. Gorse. Reinforcement learning for high-frequency market making. In *ESANN 2018-Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 521–526. Esann, 2018.

[37] M. Mani, S. Phelps, and S. Parsons. Applications of reinforcement learning in automated market-making. In *Proceedings of the GAIW: Games, Agents and Incentives Workshops, Montreal, Canada*, pages 13–14, 2019.

[38] O. Mihatsch and R. Neuneier. Risk-sensitive reinforcement learning. *Machine learning*, 49:267–290, 2002.

[39] C. C. Moallemi and M. Sağlam. Or forum—the cost of latency in high-frequency trading. *Operations Research*, 61(5):1070–1086, 2013.

[40] B. Øksendal and A. Sulem. Optimal stochastic impulse control with delayed reaction. *Applied Mathematics and Optimization*, 58:243–255, 2008.

[41] M. L. Puterman. Markov decision processes: Discrete stochastic dynamic programming. 1994.

[42] A. Ranaldo. Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74, 2004.

[43] G. A. Rummery and M. Niranjan. *On-line Q-learning using connectionist systems*, volume 37. University of Cambridge, Department of Engineering Cambridge, UK, 1994.

[44] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[45] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. Pmlr, 2014.

[46] J. A. Sirignano. Deep learning for limit order books. *Quantitative Finance*, 19(4):549–570, 2019.

[47] S. Stoikov and R. Waeber. Reducing transaction costs with low-latency trading algorithms. *Quantitative Finance*, 16(9):1445–1451, 2016.

[48] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[49] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

[50] V. Uc-Cetina, N. Navarro-Guerrero, A. Martin-Gonzalez, C. Weber, and S. Wermter. Survey on reinforcement learning for language processing. *Artificial Intelligence Review*, 56(2):1543–1575, 2023.

[51] C. J. C. H. Watkins. Learning from delayed rewards. 1989.

[52] C. Yu. *Market Making in a Limit Order Book: Classical Optimal Control and Reinforcement Learning Approaches*. PhD thesis, Washington University in St. Louis, 2021.

[53] C. Yu, J. Liu, S. Nemati, and G. Yin. Reinforcement learning in healthcare: A survey. *ACM Computing Surveys (CSUR)*, 55(1):1–36, 2021.

Appendix A

Proofs and Simulation Configuration of Chapter 2

A.1 Proof of Proposition 2.2.1

Proof. We begin by proving Eq. (2.29). Note that

$$\begin{aligned}\mathbb{E}\big[VO_1^{\pm}(k) \mid \mathcal{F}_{t_k}\big] = r_{t_k}^{\pm}\mathbb{E}\big[\mathbb{1}_{fill_k^{\pm}} c_{t_k}^{\pm}(p_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm} \mid \mathcal{F}_{t_k}\big] \\ \mp \mathbb{E}\big[\Delta S_{t_k} \cdot \mathbb{1}_{fill_k^{\pm}} c_{t_k}^{\pm}(p_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm} \mid \mathcal{F}_{t_k}\big].\end{aligned}$$

By Assumptions 2.1.2.1 and 2.1.2.2, the second term is 0. For the first term, by Assumption 2.1.1.4, we have

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}_{fill_k^{\pm}} c_{t_k}^{\pm}(p_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm} \mid \mathcal{F}_{t_k}\big] \\ &= \mathbb{E}\big[\mathbb{1}_{fill_k^{\pm}} \mathbb{E}[c_{t_k}^{\pm}(p_{t_k}^{\pm} - r_{t_k}^{\pm})_+ \wedge Q_{out_{t_k}}^{\pm} \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill_k^{\pm}}] \mid \mathcal{F}_{t_k}\big] \\ &= \mathbb{E}\big[\mathbb{1}_{fill_k^{\pm}} h_{\Delta\tau}^{\pm}(\mathbb{1}_{fill_k^{\pm}}, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}) \mid \mathcal{F}_{t_k}\big] = \pi_{t_k}^{\pm} h_{\Delta\tau}^{\pm}(1, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}),\end{aligned}$$

where we used the notation (2.32). We then conclude that

$$\mathbb{E}\big[VO_1^{\pm}(k) \mid \mathcal{F}_{t_k}\big] = \pi_{t_k}^{\pm} h_{\Delta\tau}^{\pm}(1, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}).$$

In particular, we have

$$\begin{aligned}\mathbb{E}\big[VO_1^{\pm}(k) \mid X_{t_k}\big] &= \mathbb{E}\big[\mathbb{E}[VO_1^{\pm}(k) \mid \mathcal{F}_{t_k}] \mid X_{t_k}\big] \\ &= \mathbb{E}\big[\pi_{t_k}^{\pm} h_{\Delta\tau}^{\pm}(1, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}) \mid X_{t_k}\big] \\ &= \pi_{t_k}^{\pm} h_{\Delta\tau}^{\pm}(1, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}) = \mathbb{E}\big[VO_1^{\pm}(k) \mid \mathcal{F}_{t_k}\big].\end{aligned}$$

The proof of (2.29) is now complete.

Note that $VO_3^{\pm}(k)$ takes the same form as $VO_1^{\pm}(k)$ and, thus, using the same arguments as above and Assumptions 2.1.2.1, 2.1.2.3, and 2.1.1.6, we can conclude that

$$\mathbb{E}\big[VO_3^{\pm}(k) \mid X_{t_k}\big] = \mathbb{E}\big[VO_3^{\pm}(k) \mid \mathcal{F}_{t_k}\big] = \tilde{\pi}_{t_k}^{\pm} \tilde{h}_{\Delta t}^{\pm}(1, r_{t_k}^{\pm}, Q_{out_{t_k}}^{\pm}),$$

where $\tilde{\pi}_{t_k}^{\pm}$ is defined in Eq. (2.15) and we used the notation (2.32).

Next, we proceed to prove (2.30), for which we need to evaluate:

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L_{t_k}^{\pm}\}} \cdot (L_{t_k}^{\pm} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot [\mathbb{1}_{fill_{k+}^{\pm}} c_{t_{k+}}^{\pm}(p_{t_{k+}}^{\pm} - L_{t_k}^{\pm})_+ \wedge Q_{\max}^{\pm}] \\ &\quad + \mathbb{1}_{\{\pm\Delta S_{t_k} \geq L_{t_k}^{\pm}\}} \cdot (\mp\Delta S_{t_{k+}}) \cdot Q_{\max}^{\pm} \mid \mathcal{F}_{t_k}\big].\end{aligned} \tag{A.1}$$

For the first term of (A.1), using Assumptions 2.1.1.5 and 2.1.2.2, we have

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L_{t_k}^{\pm}\}} L_{t_k}^{\pm} [\mathbb{1}_{fill_{k+}^{\pm}} c_{t_{k+}}^{\pm}(p_{t_{k+}}^{\pm} - L_{t_k}^{\pm})_+ \wedge Q_{\max}^{\pm}] \mid \mathcal{F}_{t_k}\big] \\ &= L_{t_k}^{\pm} \mathbb{P}(\pm\Delta S_{t_k} < L_{t_k}^{\pm} \mid \mathcal{F}_{t_k}) \mathbb{E}\big[\mathbb{1}_{fill_{k+}^{\pm}} c_{t_{k+}}^{\pm}(p_{t_{k+}}^{\pm} - L_{t_k}^{\pm})_+ \wedge Q_{\max}^{\pm} \mid \mathcal{F}_{t_k}\big] \\ &= L_{t_k}^{\pm} \mathbb{P}(\pm\Delta S_{t_k} < L_{t_k}^{\pm} \mid \mathcal{F}_{t_k}) \mathbb{E}\big[\mathbb{E}[\mathbb{1}_{fill_{k+}^{\pm}} c_{t_{k+}}^{\pm}(p_{t_{k+}}^{\pm} - L_{t_k}^{\pm})_+ \wedge Q_{\max}^{\pm} \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] \\ &= L_{t_k}^{\pm} \mathbb{P}(\pm\Delta S_{t_k} < L_{t_k}^{\pm} \mid \mathcal{F}_{t_k}) \mathbb{E}\big[\mathbb{P}(\mathbb{1}_{fill_{k+}^{\pm}} = 1 \mid \mathcal{F}_{t_{k+}}) \check{h}_{\Delta\tau,\Delta t}^{\pm}(1, L_{t_k}^{\pm}, Q_{\max}^{\pm}) \mid \mathcal{F}_{t_k}\big] \\ &= L_{t_k}^{\pm} \mathbb{P}(\pm\Delta S_{t_k} < L_{t_k}^{\pm} \mid \mathcal{F}_{t_k}) \mathbb{E}\big[\mathbb{P}(\mathbb{1}_{fill_{k+}^{\pm}} = 1 \mid \mathcal{F}_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] \check{h}_{\Delta\tau,\Delta t}^{\pm}(1, L_{t_k}^{\pm}, Q_{\max}^{\pm}) \\ &= L_{t_k}^{\pm} \mathbb{P}(\pm\Delta S_{t_k} < L_{t_k}^{\pm} \mid \mathcal{F}_{t_k}) \pi_{t_{k+}}^{\pm} \check{h}_{\Delta\tau,\Delta t}^{\pm}(1, L_{t_k}^{\pm}, Q_{\max}^{\pm}),\end{aligned}$$

where we used (2.32). Similarly, for the second term of (A.1), by Assumption 2.1.2.2, we have

$$\begin{aligned}
&\mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \Delta S_{t_k} [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})_+ \wedge Q^{\pm}_{\max}] \mid \mathcal{F}_{t_k}\big] \\
&= \mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \Delta S_{t_k} \mid \mathcal{F}_{t_k}\big] \pi^{\pm}_{t_{k+}} \check{h}^{\pm}_{\Delta\tau,\Delta t}(1, L^{\pm}_{t_k}, Q^{\pm}_{\max}).
\end{aligned}$$

The sum of the first and second terms of (A.1) is then $\pi^{\pm}_{t_{k+}} \check{h}^{\pm}_{\Delta\tau,\Delta t}(1, L^{\pm}_{t_k}, Q^{\pm}_{\max}) \mathbb{E}\big[(L^{\pm}_{t_k} \mp \Delta S_{t_k})_+ \mid \mathcal{F}_{t_k}\big]$.

For the third term of (A.1), by Assumptions 2.1.2.1 and 2.1.2.4, we have

$$\begin{aligned}
&\mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \Delta S_{t_{k+}} [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})_+ \wedge Q^{\pm}_{\max}] \mid \mathcal{F}_{t_k}\big] \\
&= \mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \mathbb{E}[\Delta S_{t_{k+}} [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})_+ \wedge Q^{\pm}_{\max}] \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] \\
&= \mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \mathbb{E}[\Delta S_{t_{k+}} \mid \mathcal{F}_{t_{k+}}] \mathbb{E}[\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})_+ \wedge Q^{\pm}_{\max} \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] = 0.
\end{aligned}$$

For the last term of (A.1), again by Assumption 2.1.2.1,

$$\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\}} \cdot (\mp\Delta S_{t_{k+}}) \cdot Q^{\pm}_{\max} \mid \mathcal{F}_{t_k}\big] = Q^{\pm}_{\max} \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\}} \cdot \mathbb{E}[\mp\Delta S_{t_{k+}} \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] = 0.$$

In summary, Eq. (A.1) becomes

$$\begin{aligned}
&\pi^{\pm}_{t_{k+}} \check{h}^{\pm}_{\Delta\tau,\Delta t}(1, L^{\pm}_{t_k}, Q^{\pm}_{\max}) \mathbb{E}\big[(L^{\pm}_{t_k} \mp \Delta S_{t_k})_+ \mid \mathcal{F}_{t_k}\big] \\
&\quad = \pi^{\pm}_{t_{k+}} \check{h}^{\pm}_{\Delta\tau,\Delta t}(1, L^{\pm}_{t_k}, Q^{\pm}_{\max}) \, \mathbb{E}[(l \mp \Delta S_{t_k})_+]|_{l=L^{\pm}_{t_k}},
\end{aligned}$$

where, in the last equality, we used Assumption 2.1.2.5. Note that by Assumption 2.1.2.6, $\mathbb{E}[(l \mp \Delta S_{t_k})_+]$ depends only on $\Delta\tau$ and not k.

□

A.2 Proof of Theorem 2.2.2

Proof. For $k = N$, by the Bellman equation (2.20), we have

$$\begin{aligned} V_{t_N} &= \mathbb{E}\big[W_{t_N} + P^+_{fill_{t_N}} Q_{fill^+_{t_N}} - P_{fill^-_{t_N}} Q_{fill^-_{t_N}} + (S_{t_N} + \Delta S_{t_N})(I_{t_N} - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}}) \\ &\qquad - \lambda(I_{t_N} - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}})^2 \mid \mathcal{F}_{t_N}\big] \\ &= W_{t_N} + S_{t_N} I_{t_N} + \mathbb{E}\big[Q_{fill^+_{t_N}}(P_{fill^+_{t_N}} - S_{t_N} - \Delta S_{t_N}) + Q_{fill^-_{t_N}}(S_{t_N} + \Delta S_{t_N} - P_{fill^-_{t_N}}) \\ &\qquad\qquad + I_{t_N}\Delta S_{t_N} - \lambda(I_{t_N} - Q_{fill^+_{t_N}} + Q_{fill^-_{t_N}})^2 \mid \mathcal{F}_{t_N}\big]. \end{aligned}$$

By Assumption 2.1.2.1, we have $\mathbb{E}[I_{t_N}\Delta S_{t_N} \mid \mathcal{F}_{t_N}] = 0$. It remains to demonstrate that for all admissible values of $(r^+_{t_N}, r^-_{t_N})$, the following expression holds:

$$\begin{aligned} H^+_{t_N}(\Delta\tau, r^+_{t_N}, Q^+_{out_{t_N}}) &= \mathbb{E}\big[Q_{fill^+_{t_N}}(P_{fill^+_{t_N}} - S_{t_N} - \Delta S_{t_N}) \mid \mathcal{F}_{t_N}\big], \\ H^-_{t_N}(\Delta\tau, r^-_{t_N}, Q^-_{out_{t_N}}) &= \mathbb{E}\big[Q_{fill^-_{t_N}}(S_{t_N} + \Delta S_{t_N} - P_{fill^-_{t_N}}) \mid \mathcal{F}_{t_N}\big]. \end{aligned} \tag{A.2}$$

We prove the first equation in (A.2), the ask side of the market (the other equation is proved similarly).

When $r^+_{t_N} = \infty$, there is no outstanding ask orders at time t_N, therefore $Q_{fill^+_{t_N}} = 0$, same as $H^+_{t_N}(\Delta\tau, r^+_{t_N}, Q^+_{out_{t_N}})$.

When $r^+_{t_N} \in \mathbb{Z}_+$, by the definition (2.26) with $k = N$ and Proposition 2.2.1, we have

$$\begin{aligned} H^+_{t_N}(\Delta\tau, r^+_{t_N}, Q^+_{out_{t_N}}) &= \mathbb{E}\big[(\mathbb{1}_{fill^+_N} \cdot c^+_{t_N}\left[p^+_{t_N} - r^+_{t_N}\right]_+ \wedge Q^+_{out_{t_N}})(r^+_{t_N} - \Delta S_{t_N}) \mid X_{t_N}\big] \\ &= \mathbb{E}\big[(\mathbb{1}_{fill^+_N} \cdot c^+_{t_N}\left[p^+_{t_N} - r^+_{t_N}\right]_+ \wedge Q^+_{out_{t_N}})(r^+_{t_N} - \Delta S_{t_N}) \mid \mathcal{F}_{t_N}\big] \\ &= \mathbb{E}\big[Q_{fill^+_{t_N}}(P_{fill^+_{t_N}} - S_{t_N} - \Delta S_{t_N}) \mid \mathcal{F}_{t_N}\big], \end{aligned}$$

where in the last equality above, we used the dynamics in Eqs. (2.6) and (2.9). This shows the first equation in (A.2).

Next, let us assume that, for some $k < N$, $V_{t_{k+1}} = W_{t_{k+1}} + S_{t_{k+1}} I_{t_{k+1}} + g_{k+1}(X_{t_{k+1}})$ holds and aim to show the corresponding equation for V_{t_k}. By the Bellman equation (2.20) and the dynamics of $W_{t_{k+1}}$, $S_{t_{k+1}}$ and $I_{t_{k+1}}$ in Eqs. (2.3), (2.4), and (2.5), we can write

$$
\begin{aligned}
V_{t_k} &= \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\left[V_{t_{k+1}} \mid \mathcal{F}_{t_k} \right] \\
&= \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}[W_{t_k} + P_{fill^+_{t_k}} Q_{fill^+_{t_k}} + P_{fill^+_{t_{k+}}} Q_{fill^+_{t_{k+}}} - P_{fill^-_{t_k}} Q_{fill^-_{t_k}} - P_{fill^-_{t_{k+}}} Q_{fill^-_{t_{k+}}} \\
&\qquad + (S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}})(I_{t_k} - Q_{fill^+_{t_k}} - Q_{fill^+_{t_{k+}}} + Q_{fill^-_{t_k}} + Q_{fill^-_{t_{k+}}}) \\
&\qquad + g_{k+1}(X_{t_{k+1}}) \mid \mathcal{F}_{t_k}] \\
&= W_{t_k} + S_{t_k} I_{t_k} + \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[Q_{fill^+_{t_k}} (P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \\
&\qquad\qquad + Q_{fill^+_{t_{k+}}} (P_{fill^+_{t_{k+}}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \\
&\qquad\qquad + Q_{fill^-_{t_k}} (S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_k}}) \\
&\qquad\qquad + Q_{fill^-_{t_{k+}}} (S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_{k+}}}) \\
&\qquad\qquad + I_{t_k} (\Delta S_{t_k} + \Delta S_{t_{k+}}) + g_{k+1}(X_{t_{k+1}}) \mid \mathcal{F}_{t_k} \big]. \qquad \text{(A.3)}
\end{aligned}
$$

By Assumption 2.1.2.1, the second to last term, $\mathbb{E}[I_{t_k}(\Delta S_{t_k} + \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}]$, vanishes.

We consider all possible values for $L^+_{t_k}$ and $L^-_{t_k}$ (nine in total). For illustrations, we only show the details for the cases **(a)** $L^+_{t_k}, L^-_{t_k} \in \mathbb{Z}_+$, **(b)** $L^+_{t_k} = o$, $L^-_{t_k} \in \mathbb{Z}_+$, (c) $L^+_{t_k} = \infty$, $L^-_{t_k} \in \mathbb{Z}_+$ (the other cases will then be clear).

(a) By the definitions of the function G_k, and $H^{\pm}_{act_{t_k}}(r^{\pm}, L^{\pm}, q^{\pm})$ in (2.33), the result will follow from (A.3), if we have, for any admissible $(r^+_{t_k}, r^-_{t_k}, L^+_{t_k}, L^-_{t_k})$,

$$\mathbb{E}\big[Q_{fill^+_{t_k}}(P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] = H^+_{t_k}(\Delta\tau, r^+_{t_k}, Q^+_{out_{t_k}}), \tag{A.4}$$

$$\mathbb{E}\big[Q_{fill^-_{t_k}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_k}}) \mid \mathcal{F}_{t_k}\big] = H^-_{t_k}(\Delta\tau, r^-_{t_k}, Q^-_{out_{t_k}}), \tag{A.5}$$

and

$$\mathbb{E}\big[Q_{fill^+_{t_{k+}}}(P_{fill^+_{t_{k+}}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] = H^+_{t_{k+}}(\Delta\tau, \Delta t, L^+_{t_k}), \tag{A.6}$$

$$\mathbb{E}\big[Q_{fill^-_{t_{k+}}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_{k+}}}) \mid \mathcal{F}_{t_k}\big] = H^-_{t_{k+}}(\Delta\tau, \Delta t, L^-_{t_k}). \tag{A.7}$$

We establish Eqs. (A.4) and (A.6) for the ask side, noting that the bid side can be demonstrated analogously. In the case where $r^+_{t_k} = \infty$, both sides in Eq. (A.4) vanish and it follows directly. For $r^+_{t_k} \in \mathbb{Z}_+$, by the definition of $H^+_{t_k}(\Delta\tau, r^+_{t_k}, Q^+_{out_{t_k}})$ in Eq. (2.26) and Proposition 2.2.1,

$$\begin{aligned} H^+_{t_k}(\Delta\tau, r^+_{t_k}, Q^+_{out_{t_k}}) &= \mathbb{E}\big[(\mathbb{1}_{fill^+_k} c^+_{t_k} \left[p^+_{t_k} - r^+_{t_k}\right]_+ \wedge Q^+_{out_{t_k}})(r^+_{t_k} - \Delta S_{t_k}) \mid X_{t_k}\big] \\ &= \mathbb{E}\big[(\mathbb{1}_{fill^+_k} c^+_{t_k} \left[p^+_{t_k} - r^+_{t_k}\right]_+ \wedge Q^+_{out_{t_k}})(r^+_{t_k} - \Delta S_{t_k}) \mid \mathcal{F}_{t_k}\big] \\ &= \mathbb{E}\big[Q_{fill^+_{t_k}}(P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big], \end{aligned} \tag{A.8}$$

where the last equality follows from the dynamics of $P_{fill^+_{t_k}}$ and $Q_{fill^+_{t_k}}$ in Eqs. (2.7) and (2.9), and the fact that

$$\mathbb{E}[Q_{fill^+_{t_k}} \Delta S_{t_{k+}} \mid \mathcal{F}_{t_k}] = \mathbb{E}[Q_{fill^+_{t_k}} \mathbb{E}\Delta S_{t_{k+}} \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}] = 0,$$

because of Assumption 2.1.2.1.

We proceed to establish the validity of (A.6) (the proof of (A.7) is similar). Recall that we are assuming that $L^+_{t_k} \in \mathbb{Z}_+$. Then, using Eqs. (2.7) and (2.10), we have

$$\begin{aligned}
&\mathbb{E}\big[Q_{fill^+_{t_{k+}}}(P_{fill^+_{t_{k+}}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] \\
&= \mathbb{E}\big[(\mathbb{1}_{\{\Delta S_{t_k} \geq L^+_{t_k}\}} Q^+_{\max} + \mathbb{1}_{\{\Delta S_{t_k} < L^+_{t_k}\}} \mathbb{1}_{fill^+_{k+}} c^+_{t_{k+}}[p^+_{t_{k+}} - L^+_{t_k}]_+ \wedge Q^+_{\max}) \\
&\qquad \cdot (\mathbb{1}_{\{\Delta S_{t_k} \geq L^+_{t_k}\}}[S_{t_k} + \Delta S_{t_k}] + \mathbb{1}_{\{\Delta S_{t_k} < L^+_{t_k}\}}[S_{t_k} + L^+_{t_k}] - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] \\
&= \mathbb{E}\big[\mathbb{1}_{\{\Delta S_{t_k} < L^+_{t_k}\}} \cdot (\mathbb{1}_{fill^+_{k+}} c^+_{t_{k+}}[p^+_{t_{k+}} - L^+_{t_k}]_+ \wedge Q^+_{\max})(L^+_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \\
&\qquad + \mathbb{1}_{\{\Delta S_{t_k} \geq L^+_{t_k}\}} \cdot Q^+_{\max}(-\Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] \\
&= \mathbb{E}\big[VO^+_2(k) \mid \mathcal{F}_{t_k}\big] = H^+_{t_{k+}}(\Delta\tau, \Delta t, L^+_{t_k}),
\end{aligned}$$

where in the last equality, we simply use (2.27) and Proposition 2.2.1.

(b) Now, we assume that $L^+_{t_k} = o$ and $L^-_{t_k} \in \mathbb{Z}_+$. Since $L^+_{t_k} = o$, we have $P_{fill^\pm_{t_k}} = P_{fill^\pm_{t_{k+}}} = S_{t_k} \pm r^\pm_{t_k}$ and, thus, recalling that $\widetilde{Q}_{fill^+_{t_k}} = Q_{fill^+_{t_k}} + Q_{fill^+_{t_{k+}}}$, we can reorganize (A.3) as follows:

$$\begin{aligned}
V_{t_k} = W_{t_k} + S_{t_k} I_{t_k} + \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[&\widetilde{Q}_{fill^+_{t_k}}(P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \\
&+ Q_{fill^-_{t_k}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_k}}) \\
&+ Q_{fill^-_{t_{k+}}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_{k+}}}) \\
&+ I_{t_k}(\Delta S_{t_k} + \Delta S_{t_{k+}}) + g_{k+1}(X_{t_{k+1}}) \mid \mathcal{F}_{t_k}\big].
\end{aligned}$$

Since we already showed (A.5) and (A.7), it remains to prove that

$$\mathbb{E}\big[\widetilde{Q}_{fill^+_{t_k}}(P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \mid \mathcal{F}_{t_k}\big] = \widetilde{H}^+_{t_k}(\Delta t, r^+_{t_k}, Q^+_{out_{t_k}}).$$

This is proved in the same way as (A.8).

(c) Now, we assume that $L^+_{t_k} = \infty$ and $L^-_{t_k} \in \mathbb{Z}_+$. In that case, $Q_{fill^+_{t_{k+}}} = 0$ and (A.3) is written as

$$\begin{aligned} V_{t_k} = W_{t_k} + S_{t_k} I_{t_k} + \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[& Q_{fill^+_{t_k}}(P_{fill^+_{t_k}} - S_{t_k} - \Delta S_{t_k} - \Delta S_{t_{k+}}) \\ & + Q_{fill^-_{t_k}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_k}}) \\ & + Q_{fill^-_{t_{k+}}}(S_{t_k} + \Delta S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_{k+}}}) \\ & + I_{t_k}(\Delta S_{t_k} + \Delta S_{t_{k+}}) + g_{k+1}(X_{t_{k+1}}) \mid \mathcal{F}_{t_k} \big]. \end{aligned}$$

The result then follows directly from (A.4), (A.5), and (A.7). □

A.3 Proof of Proposition 2.3.1

Proof. We begin by proving Eq. (2.42). Note that, by Assumptions 2.1.2.1 and 2.1.2.2, we have

$$\begin{aligned} H^{\pm}_{t_k}(\Delta\tau, r^{\pm}_{t_k}) &= \mathbb{E}\big[(r^{\pm}_{t_k} \mp \Delta S_{t_k}) \cdot \mathbb{1}_{fill^{\pm}_k} c^{\pm}_{t_k}(p^{\pm}_{t_k} - r^{\pm}_{t_k}) \mid \mathcal{F}_{t_k}\big] \\ &= \mathbb{E}\big[r^{\pm}_{t_k} \mp \Delta S_{t_k} \mid \mathcal{F}_{t_k}\big] \cdot \mathbb{E}\big[\mathbb{1}_{fill^{\pm}_k} c^{\pm}_{t_k}(p^{\pm}_{t_k} - r^{\pm}_{t_k}) \mid \mathcal{F}_{t_k}\big] \\ &= r^{\pm}_{t_k} \cdot \mathbb{E}\big[c^{\pm}_{t_k}(p^{\pm}_{t_k} - r^{\pm}) \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill^{\pm}_k} = 1\big] \mathbb{P}(\mathbb{1}_{fill^{\pm}_k} = 1 \mid \mathcal{F}_{t_k}) \\ &= r^{\pm}_{t_k} \pi^{\pm}_{t_k} \cdot \big\{ \mathbb{E}\big[c^{\pm}_{t_k} p^{\pm}_{t_k} \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill^{\pm}_k} = 1\big] - r^{\pm} \mathbb{E}\big[c^{\pm}_{t_k} \mid \mathcal{F}_{t_k}, \mathbb{1}_{fill^{\pm}_k} = 1\big] \big\} \\ &= r^{\pm}_{t_k} \pi^{\pm}_{t_k} (\mu^{\pm}_{cp} - \mu^{\pm}_{c} r^{\pm}_{t_k}). \end{aligned}$$

The proof of (2.42) is now complete.

Note that $VO^{\pm}_3(k)$ in Eq. (2.41) takes the same form as $VO^{\pm}_1(k)$ in Eq. (2.39) and, thus, using the same arguments as above and Assumptions 2.1.2.1, 2.1.2.3, and 2.1.1.6, we can

conclude that

$$\widetilde{H}^{\pm}_{t_k}(\Delta t, r^{\pm}_{t_k}) = \tilde{\pi}^{\pm}_{t_k} r^{\pm}_{t_k}(\tilde{\mu}^{\pm}_{cp} - \tilde{\mu}^{\pm}_{c} r^{\pm}_{t_k}).$$

Next, we proceed to show Eq. (2.43), for which we need to evaluate the following term:

$$\begin{aligned}\mathbb{E}\big[&\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \cdot (L^{\pm}_{t_k} \mp \Delta S_{t_k} \mp \Delta S_{t_{k+}}) \cdot \mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \\ &+ \mathbb{1}_{\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\}} \cdot (\mp\Delta S_{t_{k+}}) \cdot Q^{\pm}_{\max} \mid \mathcal{F}_{t_k}\big].\end{aligned} \tag{A.9}$$

For the first term of (A.9), by Assumptions 2.1.1.5 and 2.1.2.2, we have

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} L^{\pm}_{t_k} \cdot [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})] \mid \mathcal{F}_{t_k}\big] \\ &= L^{\pm}_{t_k}\mathbb{P}(\pm\Delta S_{t_k} < L^{\pm}_{t_k} \mid \mathcal{F}_{t_k})\mathbb{E}\big[\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_k}\big] \\ &= L^{\pm}_{t_k}\mathbb{P}(\pm\Delta S_{t_k} < L^{\pm}_{t_k} \mid \mathcal{F}_{t_k})\mathbb{E}\big[\mathbb{E}[\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] \\ &= L^{\pm}_{t_k}\mathbb{P}(\pm\Delta S_{t_k} < L^{\pm}_{t_k} \mid \mathcal{F}_{t_k})\mathbb{E}\big[\mathbb{P}(\mathbb{1}_{fill^{\pm}_{k+}} = 1 \mid \mathcal{F}_{t_{k+}})\mathbb{E}[c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{\pm}_{k+}} = 1] \mid \mathcal{F}_{t_k}\big] \\ &= L^{\pm}_{t_k}\mathbb{P}(\pm\Delta S_{t_k} < L^{\pm}_{t_k} \mid \mathcal{F}_{t_k})\mathbb{E}\big[\mathbb{P}(\mathbb{1}_{fill^{\pm}_{k+}} = 1 \mid \mathcal{F}_{t_{k+}}) \mid \mathcal{F}_{t_k}\big]\mathbb{E}\big[\mathbb{E}[c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_{k+}}, \mathbb{1}_{fill^{\pm}_{k+}} = 1] \mid \mathcal{F}_{t_k}\big] \\ &= L^{\pm}_{t_k}\mathbb{P}(\pm\Delta S_{t_k} < L^{\pm}_{t_k} \mid \mathcal{F}_{t_k})\pi^{\pm}_{t_{k+}}(\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} L^{\pm}_{t_k}).\end{aligned}$$

Similarly, for the second term of (A.9), by Assumption 2.1.2.2, we have

$$\begin{aligned}&\mp\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}}\Delta S_{t_k} \cdot [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})] \mid \mathcal{F}_{t_k}\big] \\ &= \mp\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}}\Delta S_{t_k} \mid \mathcal{F}_{t_k}\big] \cdot \mathbb{E}\big[\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}}(p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_k}\big] \\ &= \mp\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}}\Delta S_{t_k} \mid \mathcal{F}_{t_k}\big]\pi^{\pm}_{t_{k+}}(\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} L^{\pm}_{t_k}).\end{aligned}$$

The sum of the first and second terms of (A.9) is $\pi^{\pm}_{t_{k+}}(\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} L^{\pm}_{t_k})\mathbb{E}[(L^{\pm}_{t_k} \mp \Delta S_{t_k})_+ \mid \mathcal{F}_{t_k}]$.

For the third term of (A.9), by Assumptions 2.1.2.1 and 2.1.2.4, we have

$$\begin{aligned}
&\mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \Delta S_{t_{k+}} [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})] \mid \mathcal{F}_{t_k}\big] \\
&= \mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \mathbb{E}[\Delta S_{t_{k+}} [\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k})] \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] \\
&= \mp \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} < L^{\pm}_{t_k}\}} \mathbb{E}[\Delta S_{t_{k+}} \mid \mathcal{F}_{t_{k+}}] \mathbb{E}[\mathbb{1}_{fill^{\pm}_{k+}} c^{\pm}_{t_{k+}} (p^{\pm}_{t_{k+}} - L^{\pm}_{t_k}) \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] = 0.
\end{aligned}$$

For the last term of (A.9), by Assumption 2.1.2.1, we have

$$\mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\}} (\mp \Delta S_{t_{k+}}) Q^{\pm}_{\max} \mid \mathcal{F}_{t_k}\big] = Q^{\pm}_{\max} \mathbb{E}\big[\mathbb{1}_{\{\pm\Delta S_{t_k} \geq L^{\pm}_{t_k}\}} \mathbb{E}[\mp \Delta S_{t_{k+}} \mid \mathcal{F}_{t_{k+}}] \mid \mathcal{F}_{t_k}\big] = 0.$$

In summary, by Assumptions 2.1.2.5 and 2.1.2.6, Eq. (A.9) becomes

$$\pi^{\pm}_{t_{k+}} (\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} L^{\pm}_{t_k}) \mathbb{E}[(L^{\pm}_{t_k} \mp \Delta S_{t_k})_+ \mid \mathcal{F}_{t_k}] = \pi^{\pm}_{t_{k+}} (\mu^{\pm}_{cp_+} - \mu^{\pm}_{c_+} L^{\pm}_{t_k}) \, \mathbb{E}[(l \mp \Delta S_{t_k})_+]\big|_{l = L^{\pm}_{t_k}} .$$

□

A.4 Proof of Theorem 2.3.2

Proof. For $k = N$, by the Bellman equation (2.54), we have

$$\begin{aligned}
V_{t_N} = \max_{(L^+_{t_N}, L^-_{t_N}) \in \mathcal{A}_N} \mathbb{E}\big[& W_{t_N} + P_{fill^+_{t_{N+}}} Q_{fill^+_{t_{N+}}} - P_{fill^-_{t_{N+}}} Q_{fill^-_{t_{N+}}} \\
& + (S_{t_N} + \Delta S_{t_{N+}})(I_{t_N} - Q_{fill^+_{t_{N+}}} + Q_{fill^-_{t_{N+}}}) \\
& - \lambda (I_{t_N} - Q_{fill^+_{t_{N+}}} + Q_{fill^-_{t_{N+}}})^2 \mid \mathcal{F}_{t_N} \big] \\
= W_{t_N} + S_{t_N} I_{t_N} + \max_{(L^+_{t_N}, L^-_{t_N}) \in \mathcal{A}_N} \mathbb{E}\big[& Q_{fill^+_{t_{N+}}} (P_{fill^+_{t_{N+}}} - S_{t_N} - \Delta S_{t_{N+}}) \\
& + Q_{fill^-_{t_{N+}}} (S_{t_N} + \Delta S_{N+} - P_{fill^-_{t_{N+}}}) \\
& + I_{t_N} \Delta S_{N+} - \lambda (I_{t_N} - Q_{fill^+_{t_{N+}}} + Q_{fill^-_{t_{N+}}})^2 \mid \mathcal{F}_{t_N} \big].
\end{aligned}$$

By Assumption 2.3.2.1, it follows that

$$\mathbb{E}\big[I_{t_N} \Delta S_{N+} \mid \mathcal{F}_{t_N} \big] = 0.$$

To complete the proof, it is necessary to demonstrate that for all admissible values of $(r^+_{t_N}, r^-_{t_N}, L^+_{t_N}, L^-_{t_N})$, the following expression holds.

$$H^+_{act}(r^+_{t_N}, L^+_{t_N}, Q^+_{out_{t_N}}) = \mathbb{E}\big[Q_{fill^+_{t_{N+}}} (P_{fill^+_{t_{N+}}} - S_{t_N} - \Delta S_{t_{N+}}) \mid \mathcal{F}_{t_N} \big], \tag{A.10}$$

$$H^-_{act}(r^-_{t_N}, L^-_{t_N}, Q^-_{out_{t_N}}) = \mathbb{E}\big[Q_{fill^-_{t_{N+}}} (S_{t_N} + \Delta S_{N+} - P_{fill^-_{t_{N+}}}) \mid \mathcal{F}_{t_N} \big]. \tag{A.11}$$

We provide a proof for Eq. (A.10), the ask side of the market (the other equation is proved similarly). When $L^+_{t_N} \in \mathbb{Z}_+$, we have

$$\begin{aligned} H^+_{act}(r^+_{t_N}, L^+_{t_N}, Q^+_{out_{t_N}}) &= \mathbb{E}\big[(\mathbb{1}_{fill^+_{N+}} c^+_{t_{N+}} (p^+_{t_{N+}} - L^+_{t_N})_+ \wedge Q^+_{\max})(L^+_{t_N} - \Delta S_{t_{N+}}) \mid X_{t_N}\big] \\ &= \mathbb{E}\big[(\mathbb{1}_{fill^+_{N+}} c^+_{t_{N+}} (p^+_{t_{N+}} - L^+_{t_N})_+ \wedge Q^+_{\max})(L^+_{t_N} - \Delta S_{t_{N+}}) \mid \mathcal{F}_{t_N}\big] \\ &= \mathbb{E}\big[Q_{fill^+_{t_{N+}}} (P_{fill^+_{t_{N+}}} - S_{t_N} - \Delta S_{t_{N+}}) \mid \mathcal{F}_{t_N}\big], \end{aligned}$$

where in the last equality above, we used the dynamics in Eqs. (2.49) and (2.50).

When $L^+_{t_N} = o$, again, by the dynamics in Eqs. (2.49) and (2.50), we have

$$\begin{aligned} H^+_{act}(r^+_{t_N}, L^+_{t_N}, Q^+_{out_{t_N}}) &= \mathbb{E}\big[(\mathbb{1}_{fill^+_{N+}} c^+_{t_{N+}} (p^+_{t_{N+}} - r^+_{t_N})_+ \wedge Q^+_{out_{t_N}})(r^+_{t_{N+}} - \Delta S_{t_{N+}}) \mid X_{t_N}\big] \\ &= \mathbb{E}\big[(\mathbb{1}_{fill^+_{N+}} c^+_{t_{N+}} (p^+_{t_{N+}} - r^+_{t_N})_+ \wedge Q^+_{out_{t_N}})(r^+_{t_{N+}} - \Delta S_{t_{N+}}) \mid \mathcal{F}_{t_N}\big] \\ &= \mathbb{E}\big[Q_{fill^+_{t_{N+}}} (P_{fill^+_{t_{N+}}} - S_{t_N} - \Delta S_{t_{N+}}) \mid \mathcal{F}_{t_N}\big]. \end{aligned}$$

When $L^+_{t_N} = \infty$, the result is trivial because both sides of Eq. (A.10) are equal to 0. This shows the Eq. (A.10).

Next, for some $k < N$, let us assume that $V_{t_{k+1}} = W_{t_{k+1}} + S_{t_{k+1}} I_{t_{k+1}} + g_{k+1}(X_{t_{k+1}})$ holds and aim to show the corresponding equation for V_{t_k}. Indeed, by the Bellman equation (2.54) and the dynamics of $W_{t_{k+1}}$, $S_{t_{k+1}}$ and $I_{t_{k+1}}$ in Eqs. (2.47), (2.46) and (2.48), we

have

$$
\begin{aligned}
V_{t_k} &= \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[V_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] \\
&= \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[W_{t_k} + P_{fill^+_{t_{k+}}} Q_{fill^+_{t_{k+}}} - P_{fill^-_{t_{k+}}} Q_{fill^-_{t_{k+}}} + g_{k+1}(X_{t_{k+1}}) \\
&\qquad\qquad + (S_{t_k} + \Delta S_{t_{k+}})(I_{t_k} - Q_{fill^+_{t_{k+}}} + Q_{fill^-_{t_{k+}}}) \mid \mathcal{F}_{t_k}\big] \\
&= W_{t_k} + S_{t_k} I_{t_k} + \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \mathbb{E}\big[I_{t_k} \Delta S_{t_{k+}} + g_{k+1}(X_{t_{k+1}}) + Q_{fill^+_{t_{k+}}} (P_{fill^+_{t_{k+}}} - S_{t_k} - \Delta S_{t_{k+}}) \\
&\qquad\qquad + Q_{fill^-_{t_{k+}}} (S_{t_k} + \Delta S_{t_{k+}} - P_{fill^-_{t_{k+}}}) \mid \mathcal{F}_{t_k}\big].
\end{aligned}
$$

Similar to our approach for the $k = N$ scenario and the reasoning in proof of Theorem 2.2.2, we can employ the same logic to the current case. Therefore, we have

$$
\begin{aligned}
V_{t_k} = W_{t_k} + S_{t_k} I_{t_k} + \max_{(L^+_{t_k}, L^-_{t_k}) \in \mathcal{A}_k} \big\{ & H^+_{act}(r^+_{t_k}, L^+_{t_k}, Q^+_{out_{t_k}}) + H^-_{act}(r^-_{t_k}, L^-_{t_k}, Q^-_{out_{t_k}}) \\
& + \mathbb{E}\big[g_{k+1}(X_{t_{k+1}}) \mid \mathcal{F}_{t_k}\big]\big\}.
\end{aligned}
$$

Similar arguments follow for the bid side. □

A.5 Parameters under 0.20 seconds Latency

Table A.1: Average arrival probability of market orders over 252 trading days in 2019. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second.

Stock	$\widehat{\pi}^{+}_{\Delta\tau}$	$\widehat{\pi}^{-}_{\Delta\tau}$	$\widehat{\pi}^{+}_{\Delta\tau,\Delta t}$	$\widehat{\pi}^{-}_{\Delta\tau,\Delta t}$
AAPL	0.072558	0.072784	0.215578	0.221396
AMZN	0.024995	0.025629	0.080353	0.082441
MSFT	0.054371	0.054845	0.167368	0.169665

Table A.2: Average values of $\hat{\mu}^{\pm}_{\{c,p\}}$ over 252 trading days for 2019 AAPL. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second.

$\bar{\mu}^{+}_{c} = 38.39$	$\bar{\mu}^{-}_{c} = 42.61$
$\bar{\mu}^{+}_{p} = 3.473$	$\bar{\mu}^{-}_{p} = 3.542$
$\bar{\mu}^{+}_{cp} = 131.79$	$\bar{\mu}^{-}_{cp} = 148.74$
$\bar{\mu}^{+}_{c^2} = 1521.62$	$\bar{\mu}^{-}_{c^2} = 1882.37$
$\bar{\mu}^{+}_{p^2} = 12.23$	$\bar{\mu}^{-}_{p^2} = 12.74$

(a) Latency Period

$\bar{\mu}^{+}_{c} = 45.75$	$\bar{\mu}^{-}_{c} = 51.15$
$\bar{\mu}^{+}_{p} = 3.330$	$\bar{\mu}^{-}_{p} = 3.375$
$\bar{\mu}^{+}_{cp} = 151.41$	$\bar{\mu}^{-}_{cp} = 171.27$
$\bar{\mu}^{+}_{c^2} = 2142.96$	$\bar{\mu}^{-}_{c^2} = 2689.93$
$\bar{\mu}^{+}_{p^2} = 11.20$	$\bar{\mu}^{-}_{p^2} = 11.52$

(b) Non-latency Period

Table A.3: Parameter values for c and p in simulation for 2019 AAPL. $\Delta\tau = 0.20$ seconds and $\Delta t = 1$ second.

Parameter	**Value**	**Unit**
Slope of demand function $c^{+}_{\Delta\tau}$	$\Gamma(shape = 30.81, scale = 1.245)$	share/tick
Slope of demand function $c^{-}_{\Delta\tau}$	$\Gamma(shape = 27.24, scale = 1.564)$	share/tick
Slope of demand function $c^{+}_{\Delta\tau,\Delta t}$	$\Gamma(shape = 42.54, scale = 1.075)$	share/tick
Slope of demand function $c^{-}_{\Delta\tau,\Delta t}$	$\Gamma(shape = 35.74, scale = 1.430)$	share/tick
Reservation price $p^{+}_{\Delta\tau}$	$\Gamma(shape = 69.62, scale = 0.049)$	tick
Reservation price $p^{-}_{\Delta\tau}$	$\Gamma(shape = 63.66, scale = 0.055)$	tick
Reservation price $p^{+}_{\Delta\tau,\Delta t}$	$\Gamma(shape = 92.48, scale = 0.036)$	tick
Reservation price $p^{-}_{\Delta\tau,\Delta t}$	$\Gamma(shape = 85.79, scale = 0.039)$	tick

Appendix B

Proofs of Chapter 4

The subsequent proofs follow a similar framework as presented in [12], and is included herein for the sake of completeness. Readers familiar with [12] will recognize the underlying structure and can refer to that work for further details.

B.1 Proof of Proposition 4.2.1 and Theorem 4.2.4

Proof. We prove Proposition 4.2.1 and Theorem 4.2.4 by conducting a four-step approach.

Step 1. To begin, we put forth a proposed ansatz for the value function V_{t_k}:

$$V_{t_k} = \widetilde{v}\left(t_k, S_{t_k}, W_{t_k}, I_{t_k}\right) := W_{t_k} + \alpha_{t_k} I_{t_k}^2 + S_{t_k} I_{t_k} + \widetilde{h}_{t_k} I_{t_k} + \widetilde{g}_{t_k}. \tag{B.1}$$

The function $\alpha : \mathcal{T} \to \mathbb{R}$ is a deterministic function defined on the time set $\mathcal{T} = \{t_0, t_1, \ldots, t_{N+1}\}$. Additionally, we introduce the processes $\{\widetilde{h}_t\}_{t\in\mathcal{T}}$, $\{\widetilde{g}_t\}_{t\in\mathcal{T}}$, which are adapted to the filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$. By considering the terminal conditions $V_T = W_T + S_T I_T - \lambda I_T^2$, we can deduce that $\alpha_T = -\lambda, \widetilde{g}_T = 0$, and $\widetilde{h}_T = 0$. To facilitate our subsequent discussions, we introduce the following notation:

$$\widetilde{h}_{t_{k+1}}^{t_k} := \mathbb{E}[\widetilde{h}_{t_{k+1}} \mid \mathcal{F}_{t_k}], \quad \widetilde{g}_{t_{k+1}}^{t_k} := \mathbb{E}[\widetilde{g}_{t_{k+1}} \mid \mathcal{F}_{t_k}]. \tag{B.2}$$

By substituting the expressions given in Eq. (4.4), (4.5), and (B.1) into the right-hand side of the Bellman equation (4.6), we obtain the following expression:

$$
\begin{aligned}
V_{t_k} &= \sup_{(L^+_{t_k},L^-_{t_k})\in\mathcal{A}} \mathbb{E}[V_{t_{k+1}} - \phi I^2_{t_{k+1}} \mid \mathcal{F}_{t_k}] \\
&= \sup_{(L^+_{t_k},L^-_{t_k})\in\mathcal{A}} \mathbb{E}[W_{t_{k+1}} + \alpha_{t_{k+1}} I^2_{t_{k+1}} + S_{t_{k+1}} I_{t_{k+1}} + \widetilde{h}_{t_{k+1}} I_{t_{k+1}} + \widetilde{g}_{t_{k+1}} - \phi I^2_{t_{k+1}} \mid \mathcal{F}_{t_k}] \\
&= \sup_{(L^+_{t_k},L^-_{t_k})\in\mathcal{A}} \mathbb{E}\Big\{ W_{t_k} + \sum_{\delta=\pm} (S_{t_k} + \delta L^\delta_{t_k}) \delta \mathbb{1}^\delta_{t_{k+1}} c^\delta_{t_{k+1}} (p^\delta_{t_{k+1}} - L^\delta_{t_k}) \\
&\qquad + \alpha_{t_{k+1}} \big[I_{t_k} - \sum_{\delta=\pm} \delta \mathbb{1}^\delta_{t_{k+1}} c^\delta_{t_{k+1}} (p^\delta_{t_{k+1}} - L^\delta_{t_k}) \big]^2 \\
&\qquad + S_{t_{k+1}} \big[I_{t_k} - \sum_{\delta=\pm} \delta \mathbb{1}^\delta_{t_{k+1}} c^\delta_{t_{k+1}} (p^\delta_{t_{k+1}} - L^\delta_{t_k}) \big] \\
&\qquad + \widetilde{h}_{t_{k+1}} \big[I_{t_k} - \sum_{\delta=\pm} \delta \mathbb{1}^\delta_{t_{k+1}} c^\delta_{t_{k+1}} (p^\delta_{t_{k+1}} - L^\delta_{t_k}) \big] \\
&\qquad + \widetilde{g}_{t_{k+1}} - \phi I^2_{t_{k+1}} \mid \mathcal{F}_{t_k} \Big\}.
\end{aligned} \tag{B.3}
$$

Expanding the squares, we have

$$
W_{t_k} + \sum_{\delta=\pm} \mathbb{1}^\delta_{t_{k+1}} \big[- c^\delta_{t_{k+1}} (L^\delta_{t_k})^2 + (c^\delta_{t_{k+1}} p^\delta_{t_{k+1}} - \delta c^\delta_{t_{k+1}} S_{t_k}) L^\delta_{t_k} + \delta c^\delta_{t_{k+1}} p^\delta_{t_{k+1}} S_{t_k} \big] \tag{B.4}
$$

$$
\begin{aligned}
&+ (\alpha_{t_{k+1}} - \phi) \Big\{ I^2_{t_k} + \sum_{\delta=\pm} \mathbb{1}^\delta_{t_{k+1}} \Big\{ (c^\delta_{t_{k+1}})^2 (L^\delta_{t_k})^2 + \big[2\delta I_{t_k} c^\delta_{t_{k+1}} - 2 (c^\delta_{t_{k+1}})^2 p^\delta_{t_{k+1}} \big] L^\delta_{t_k} \\
&\qquad\qquad + (c^\delta_{t_{k+1}} p^\delta_{t_{k+1}})^2 - 2\delta I_{t_k} c^\delta_{t_{k+1}} p^\delta_{t_{k+1}} \Big\}
\end{aligned} \tag{B.5}
$$

$$
\begin{aligned}
&\qquad + 2 \mathbb{1}^+_{t_{k+1}} \mathbb{1}^-_{t_{k+1}} c^+_{t_{k+1}} c^-_{t_{k+1}} (-L^+_{t_k} L^-_{t_k} + p^+_{t_{k+1}} L^-_{t_k} + p^-_{t_{k+1}} L^+_{t_k} - p^+_{t_{k+1}} p^-_{t_{k+1}}) \Big\}
\end{aligned}
$$

$$
+ S_{t_{k+1}} \Big[I_{t_k} + \sum_{\delta=\pm} \mathbb{1}^\delta_{t_{k+1}} (-\delta c^\delta_{t_{k+1}} p^\delta_{t_{k+1}} + \delta c^\delta_{t_{k+1}} L^\delta_{t_k}) \Big] \tag{B.6}
$$

$$
+ \widetilde{h}_{t_{k+1}} I_{t_k} + \sum_{\delta=\pm} \mathbb{1}^\delta_{t_{k+1}} (-\delta \widetilde{h}_{t_{k+1}} c^\delta_{t_{k+1}} p^\delta_{t_{k+1}} + \delta \widetilde{h}_{t_{k+1}} c^\delta_{t_{k+1}} L^\delta_{t_k}) + \widetilde{g}_{t_{k+1}}. \tag{B.7}
$$

The computation of the conditional expectations for most of the terms above is straightforward, utilizing the conditions stated in Assumption 4.1.1 and the adaptability of the controls $\left\{L^{\pm}_{t_k}\right\}, \{S_{t_k}\}$, and $\{I_{t_k}\}$. For instance,

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}^{+}_{t_{k+1}}\mathbb{1}^{-}_{t_{k+1}}c^{+}_{t_{k+1}}c^{-}_{t_{k+1}}p^{+}_{t_{k+1}}p^{-}_{t_{k+1}} \mid \mathcal{F}_{t_k}] \\
&= \mathbb{E}\big[\mathbb{1}^{+}_{t_{k+1}}\mathbb{1}^{-}_{t_{k+1}}\mathbb{E}[c^{+}_{t_{k+1}}p^{+}_{t_{k+1}} \mid \mathcal{F}_{t_k}, \mathbb{1}^{+}_{t_{k+1}}\mathbb{1}^{-}_{t_{k+1}}]\mathbb{E}[c^{-}_{t_{k+1}}p^{-}_{t_{k+1}} \mid \mathcal{F}_{t_k}, \mathbb{1}^{+}_{t_{k+1}}\mathbb{1}^{-}_{t_{k+1}}] \mid \mathcal{F}_{t_k}\big] \\
&= \mu^{+}_{cp}\mu^{-}_{cp}\mathbb{E}[\mathbb{1}^{+}_{t_{k+1}}\mathbb{1}^{-}_{t_{k+1}} \mid \mathcal{F}_{t_k}] \\
&= \mu^{+}_{cp}\mu^{-}_{cp}\pi_{t_{k+1}}(1,1).
\end{aligned}$$

For the terms in (B.6), using the conditional independence of $S_{t_{k+1}} - S_{t_k}$ and $(\mathbb{1}^{+}_{t_{k+1}}, \mathbb{1}^{-}_{t_{k+1}}, c^{+}_{t_{k+1}}, p^{+}_{t_{k+1}}, c^{-}_{t_{k+1}}, p^{-}_{t_{k+1}})$, given $\mathcal{F}_{t_k}$, we have:

$$\begin{aligned}
&\mathbb{E}\big[(S_{t_{k+1}} - S_{t_k})\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] \\
&= \mathbb{E}[S_{t_{k+1}} - S_{t_k} \mid \mathcal{F}_{t_k}]\mathbb{E}[\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}] \\
&= \Delta_{t_k}\pi^{\delta}_{t_{k+1}}\mu^{\delta}_{cp},
\end{aligned} \tag{B.8}$$

and, thus, $\mathbb{E}[S_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}] = (S_{t_k} + \Delta_{t_k})\pi^{\delta}_{t_{k+1}}\mu^{\delta}_{cp}$. Likewise, we have $\mathbb{E}[S_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}] = (S_{t_k} + \Delta_{t_k})\pi^{\delta}_{t_{k+1}}\mu^{\delta}_{c}$. For the expressions in (B.7), let's, for the time being, assume that:

$$\mathbb{E}[\widetilde{h}_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}] = \widetilde{h}^{t_k}_{t_{k+1}}\mathbb{E}[\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}], \tag{B.9}$$

$$\mathbb{E}[\widetilde{h}_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}] = \widetilde{h}^{t_k}_{t_{k+1}}\mathbb{E}[\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}]. \tag{B.10}$$

The above identities will be established in detail in Step 4. Building upon the preceding arguments, we can evaluate the conditional expectation $\mathbb{E}[\cdot \mid \mathcal{F}_{t_k}]$ for the terms appearing in Eqs. (B.4)-(B.7). Subsequently, we substitute these conditional expectations into the

right-hand side of Eq. (B.3), yielding the following expression:

$$
\begin{aligned}
&\alpha_{t_k} I_{t_k}^2 + S_{t_k} I_{t_k} + \widetilde{h}_{t_k} I_{t_k} + \widetilde{g}_{t_k} \\
&= \sup_{(L_{t_k}^+, L_{t_k}^-) \in \mathcal{A}} \sum_{\delta=\pm} \pi_{t_{k+1}}^{\delta} \Big\{ \big((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{\delta} - \mu_c^{\delta}\big)(L_{t_k}^{\delta})^2 \\
&\qquad\qquad + \big[\mu_{cp}^{\delta} + \delta \widetilde{h}_{t_{k+1}}^{t_k} \mu_c^{\delta} + (\alpha_{t_{k+1}} - \phi)(2\delta \mu_c^{\delta} I_{t_k} - 2\mu_{c^2p}^{\delta}) + \delta \mu_c^{\delta} \Delta_{t_k}\big] L_{t_k}^{\delta} \\
&\qquad\qquad + (\alpha_{t_{k+1}} - \phi)(\mu_{c^2p^2}^{\delta} - 2\delta \mu_{cp}^{\delta} I_{t_k}) - \delta \widetilde{h}_{t_{k+1}}^{t_k} \mu_{cp}^{\delta} - \delta \mu_{cp}^{\delta} \Delta_{t_k} \Big\} \\
&\quad + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)(-\mu_c^+ \mu_c^- L_{t_k}^+ L_{t_k}^- + \mu_{cp}^+ \mu_c^- L_{t_k}^- + \mu_c^+ \mu_{cp}^- L_{t_k}^+ - \mu_{cp}^+ \mu_{cp}^-) \\
&\quad + (\alpha_{t_{k+1}} - \phi) I_{t_k}^2 + I_{t_k}(S_{t_k} + \Delta_{t_k}) + \widetilde{h}_{t_{k+1}}^{t_k} I_{t_k} + \widetilde{g}_{t_{k+1}}^{t_k}.
\end{aligned}
\tag{B.11}
$$

Let us denote the right-hand side of the above equation as $\sup_{(L_{t_k}^+, L_{t_k}^-) \in \mathcal{A}} \widetilde{f}\left(L_{t_k}^+, L_{t_k}^-\right)$. It is evident that $\widetilde{f}\left(L_{t_k}^+, L_{t_k}^-\right)$ represents a quadratic function with respect to $L_{t_k}^+$ and $L_{t_k}^-$. By setting the partial derivatives with respect to $L_{t_k}^+$ and $L_{t_k}^-$ equal to zero, we obtain the following system of equations:

$$
\begin{aligned}
\partial_{L_{t_k}^+} \widetilde{f} &= 2\pi_{t_{k+1}}^+ \big((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^+ - \mu_c^+\big) L_{t_k}^+ \\
&\quad + \pi_{t_{k+1}}^+ \big[\mu_{cp}^+ + \widetilde{h}_{t_{k+1}}^{t_k} \mu_c^+ + (\alpha_{t_{k+1}} - \phi)(2\mu_c^+ I_{t_k} - 2\mu_{c^2p}^+) + \mu_c^+ \Delta_{t_k}\big] \\
&\quad - 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^+ \mu_c^- L_{t_k}^- + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^+ \mu_{cp}^- = 0, \\
\partial_{L_{t_k}^-} \widetilde{f} &= 2\pi_{t_{k+1}}^- \big((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^- - \mu_c^-\big) L_{t_k}^- \\
&\quad + \pi_{t_{k+1}}^- \big[\mu_{cp}^- - \widetilde{h}_{t_{k+1}}^{t_k} \mu_c^- + (\alpha_{t_{k+1}} - \phi)(-2\mu_c^- I_{t_k} - 2\mu_{c^2p}^-) - \mu_c^- \Delta_{t_k}\big] \\
&\quad - 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^+ \mu_c^- L_{t_k}^+ + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^- \mu_{cp}^+ = 0.
\end{aligned}
\tag{B.12}
$$

By solving for $L_{t_k}^+$and $L_{t_k}^-$, we obtain

$$
\widetilde{L}_{t_k}^{+,*} = {}^{(1)}A_{t_k}^+ I_{t_k} + {}^{(2)}\widetilde{A}_{t_k}^+ + {}^{(3)}\widetilde{A}_{t_k}^+, \quad \widetilde{L}_{t_k}^{-,*} = -{}^{(1)}A_{t_k}^- I_{t_k} - {}^{(2)}\widetilde{A}_{t_k}^- + {}^{(3)}\widetilde{A}_{t_k}^-, \tag{B.13}
$$

where

$$
{}^{(1)}A^{\pm}_{t_k} = \frac{\beta^{\pm}_{t_k}(\alpha_{t_{k+1}} - \phi)}{\gamma_{t_k}}, \quad {}^{(2)}\widetilde{A}^{\pm}_{t_k} = \frac{\beta^{\pm}_{t_k}\widetilde{h}^{t_k}_{t_{k+1}}}{2\gamma_{t_k}}, \tag{B.14}
$$

$$
\begin{aligned}
{}^{(3)}\widetilde{A}^{\pm}_{t_k} = \frac{\pi^{\mp}_{t_{k+1}}}{2\gamma_{t_k}}\big((\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2} - \mu^{\mp}_{c}\big)\Big[\pi^{\pm}_{t_{k+1}}\big(\mu^{\pm}_{cp} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\pm}_{c^2p}\big) \\
+ 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{\pm}_{c}\mu^{\mp}_{cp} \pm \pi^{\pm}_{t_{k+1}}\Delta_{t_k}\mu^{\pm}_{c}\Big] \\
+ \pi_{t_{k+1}}(1,1)\frac{(\alpha_{t_{k+1}} - \phi)}{2\gamma_{t_k}}\mu^{+}_{c}\mu^{-}_{c}\Big[\pi^{\mp}_{t_{k+1}}\big(\mu^{\mp}_{cp} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2p}\big) \\
+ 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{\mp}_{c}\mu^{\pm}_{cp} \mp \pi^{\mp}_{t_{k+1}}\Delta_{t_k}\mu^{\mp}_{c}\Big].
\end{aligned} \tag{B.15}
$$

Upon substituting $\widetilde{L}^{\pm,*}_{t_k}$ into Eq. (B.11) and equating terms with respect to I_{t_k}, we derive the following recursive expressions for $\alpha_{t_k}, \tilde{h}_{t_k}$, and $\tilde{g}_{t_k}$:

$$
\begin{aligned}
\alpha_{t_k} = \alpha_{t_{k+1}} - \phi + \sum_{\delta=\pm}\pi^{\delta}_{t_{k+1}}\Big[\big((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\big)({}^{(1)}A^{\delta}_{t_k})^2 + 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c}({}^{(1)}A^{\delta}_{t_k})\Big] \\
+ 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}({}^{(1)}A^{+}_{t_k}\,{}^{(1)}A^{-}_{t_k}),
\end{aligned} \tag{B.16}
$$

$$
\begin{aligned}
\widetilde{h}_{t_k} =& \widetilde{h}^{t_k}_{t_{k+1}} + \sum_{\delta=\pm} \pi^{\delta}_{t_{k+1}} \Big\{ 2\big((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\big)\big[{}^{(1)}A^{\delta}_{t_k}\big((\delta\,{}^{(3)}\widetilde{A}^{\delta}_{t_k}) + {}^{(2)}\widetilde{A}^{\delta}_{t_k}\big)\big] \\
&\qquad + 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c}\big((\delta\,{}^{(3)}\widetilde{A}^{\delta}_{t_k}) + {}^{(2)}\widetilde{A}^{\delta}_{t_k}\big) - 2(\alpha_{t_{k+1}} - \phi)(\delta\mu^{\delta}_{cp}) \\
&\qquad + (\delta^{(1)}A^{\delta}_{t_k})\big(\mu^{\delta}_{cp} + (\delta\widetilde{h}^{t_k}_{t_{k+1}})\mu^{\delta}_{c} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2p}\big)\Big\} \\
&- 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}\Big[{}^{(1)}A^{+}_{t_k}({}^{(3)}\widetilde{A}^{-}_{t_k} - {}^{(2)}\widetilde{A}^{-}_{t_k}) - {}^{(1)}A^{-}_{t_k}({}^{(2)}\widetilde{A}^{+}_{t_k} + {}^{(3)}\widetilde{A}^{+}_{t_k}) \\
&\qquad + \frac{\mu^{+}_{cp}}{\mu^{+}_{c}}({}^{(1)}A^{-}_{t_k}) - \frac{\mu^{-}_{cp}}{\mu^{-}_{c}}({}^{(1)}A^{+}_{t_k})\Big] \\
&+ \Delta_{t_k}\big[{}^{(1)}A^{+}_{t_k}\pi^{+}_{t_{k+1}}\mu^{+}_{c} + {}^{(1)}A^{-}_{t_k}\pi^{-}_{t_{k+1}}\mu^{-}_{c} + 1\big],
\end{aligned}
\tag{B.17}
$$

$$
\begin{aligned}
\widetilde{g}_{t_k} =& \widetilde{g}^{t_k}_{t_{k+1}} + \sum_{\delta=\pm} \pi^{\delta}_{t_{k+1}} \Big[\big((\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2} - \mu^{\delta}_{c}\big)\big({}^{(3)}\widetilde{A}^{\delta}_{t_k} + (\delta\,{}^{(2)}\widetilde{A}^{\delta}_{t_k})\big)^2 \\
&\qquad + (\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2p^2} - (\delta\widetilde{h}^{t_k}_{t_{k+1}})\mu^{\delta}_{cp} \\
&\qquad + \big(\mu^{\delta}_{cp} + (\delta\widetilde{h}^{t_k}_{t_{k+1}})\mu^{\delta}_{c} - 2(\alpha_{t_{k+1}} - \phi)\mu^{\delta}_{c^2p}\big)\big({}^{(3)}\widetilde{A}^{\delta}_{t_k} + (\delta\,{}^{(2)}\widetilde{A}^{\delta}_{t_k})\big)\Big] \\
&- 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu^{+}_{c}\mu^{-}_{c}\Big[({}^{(2)}\widetilde{A}^{+}_{t_k} + {}^{(3)}\widetilde{A}^{+}_{t_k})({}^{(3)}\widetilde{A}^{-}_{t_k} - {}^{(2)}\widetilde{A}^{-}_{t_k}) \\
&\qquad - \frac{\mu^{+}_{cp}}{\mu^{+}_{c}}({}^{(3)}\widetilde{A}^{-}_{t_k} - {}^{(2)}\widetilde{A}^{-}_{t_k}) - \frac{\mu^{-}_{cp}}{\mu^{-}_{c}}({}^{(2)}\widetilde{A}^{+}_{t_k} + {}^{(3)}\widetilde{A}^{+}_{t_k}) + \frac{\mu^{+}_{cp}\mu^{-}_{cp}}{\mu^{-}_{c}\mu^{+}_{c}}\Big] \\
&+ \Delta_{t_k}\Big[({}^{(3)}\widetilde{A}^{\delta}_{t_k} + {}^{(2)}\widetilde{A}^{\delta}_{t_k})\pi^{+}_{t_{k+1}}\mu^{+}_{c} - ({}^{(3)}\widetilde{A}^{\delta}_{t_k} - {}^{(2)}\widetilde{A}^{\delta}_{t_k})\pi^{-}_{t_{k+1}}\mu^{-}_{c} - \pi^{+}_{t_{k+1}}\mu^{+}_{cp} + \pi^{-}_{t_{k+1}}\mu^{-}_{cp}\Big].
\end{aligned}
\tag{B.18}
$$

Step 2. We proceed to establish the optimality of $\widetilde{L}^{\pm,*}_{t_k}$ as the maximum points of the function $\tilde{f}\left(L^{+}_{t_k}, L^{-}_{t_k}\right)$. To accomplish this, we will rely on the utilization of Lemma 4.2.2, which asserts that $\alpha_{t_k} < 0$. Specifically, for any given t_k, we can demonstrate the following

inequality:

$$\begin{aligned} D &= (\partial^2_{L^+_{t_k}} \tilde{f})(\partial^2_{L^-_{t_k}} \tilde{f}) - (\partial_{L^+_{t_k} L^-_{t_k}} \tilde{f})^2 \\ &= 4\pi^+_{t_{k+1}} \pi^-_{t_{k+1}} \big((\alpha_{t_{k+1}} - \phi)\mu^+_{c^2} - \mu^+_c\big)\big((\alpha_{t_{k+1}} - \phi)\mu^-_{c^2} - \mu^-_c\big) \\ &\quad - 4\big[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^+_c \mu^-_c\big]^2 > 0, \\ \partial^2_{L^+_{t_k}} \tilde{f} &= (\alpha_{t_{k+1}} - \phi)\mu^+_{c^2} - \mu^+_c < 0. \end{aligned} \tag{B.19}$$

Applying the second derivative test, we find that $\tilde{f}\left(L^+_{t_k}, L^-_{t_k}\right)$ reaches its maximum value at $\tilde{L}^{\pm,*}_{t_k}$.

Step 3. Now, we proceed to show that (4.17) holds. Plugging the expressions of ${}^{(2)}\widetilde{A}^{\pm}_{t_k}$ and ${}^{(3)}\widetilde{A}^{\pm}_{t_k}$ given in (B.14)-(B.15) into (B.17), we have

$$\widetilde{h}_{t_k} = d_k + \xi_k(\widetilde{h}^{t_k}_{t_{k+1}} + \Delta_{t_k}), \tag{B.20}$$

for some deterministic constant d_k and

$$\begin{aligned} \xi_k = 1 + \frac{\alpha_{t_{k+1}} - \phi}{\gamma_{t_k}} \sum_{\delta=\pm} \pi^\delta_{t_{k+1}} \beta^\delta_{t_k} \left[\frac{\beta^\delta_{t_k}}{\gamma_{t_k}}\left((\alpha_{t_{k+1}} - \phi)\mu^\delta_{c^2} - \mu^\delta_c\right) + 2\mu^\delta_c\right] \\ + 2\frac{(\alpha_{t_{k+1}} - \phi)^2}{\gamma^2_{t_k}} \pi_{t_{k+1}}(1,1)\mu^+_c \mu^-_c \beta^+_{t_k} \beta^-_{t_k}. \end{aligned}$$

Also, h_{t_k} in Eq. (4.15) can be written as

$$h_{t_k} = d_k + \xi_k h_{t_{k+1}}, \tag{B.21}$$

where d_k, ξ_k are the same as those in (B.20). Since $h_{t_{N+1}} = 0$ and $\widetilde{h}_{t_{N+1}} = 0$, we have that $\widetilde{h}_{t_N} = d_N + \xi_N \Delta_{t_N}$ and $h_{t_N} = d_N$ at time t_N. By induction, we get

$$h_{t_k} = \sum_{j=k}^{N} \prod_{\ell=k}^{j-1} \xi_\ell d_j,$$

where $\prod_{\ell=k}^{k-1} \xi_\ell := 1$, and

$$\widetilde{h}_{t_k} = \sum_{j=k}^{N} \prod_{\ell=k}^{j-1} \xi_\ell (d_j + \xi_j \Delta_{t_j}^{t_k}) = h_{t_k} + \sum_{j=k}^{N} \prod_{\ell=k}^{j} \xi_\ell \Delta_{t_j}^{t_k}. \tag{B.22}$$

In particular,

$$\widetilde{h}_{t_{k+1}}^{t_k} = \mathbb{E}\Big[h_{t_{k+1}} + \sum_{j=k+1}^{N} \prod_{\ell=k+1}^{j} \xi_\ell \Delta_{t_j}^{t_{k+1}}) \mid \mathcal{F}_{t_k}\Big] = h_{t_{k+1}} + \sum_{j=k+1}^{N} \prod_{\ell=k+1}^{j} \xi_\ell \Delta_{t_j}^{t_k}.$$

Substituting the above expression into ${}^{(2)}\widetilde{A}_{t_k}^{\pm}$ as defined in (B.14), and subsequently replacing ${}^{(1)}At_k{}^{\pm}I_{t_k}$, ${}^{(2)}\widetilde{A}_{t_k}^{\pm}$, ${}^{(3)}\widetilde{A}_{t_k}^{\pm}$ into (B.13), we conclude that

$$\begin{aligned}
\widetilde{L}_{t_k}^{+,*} &= L_{t_k}^{+,*} + \frac{\beta_{t_k}^{+}}{2\gamma_{t_k}} \Delta_{t_k} + \left(\frac{\beta_{t_k}^{+}}{2\gamma_{t_k}}\right) \sum_{j=k+1}^{N} \prod_{\ell=k+1}^{j} \xi_\ell \Delta_{t_j}^{t_k}, \\
\widetilde{L}_{t_k}^{-,*} &= L_{t_k}^{-,*} - \frac{\beta_{t_k}^{-}}{2\gamma_{t_k}} \Delta_{t_k} - \left(\frac{\beta_{t_k}^{-}}{2\gamma_{t_k}}\right) \sum_{j=k+1}^{N} \prod_{\ell=k+1}^{j} \xi_\ell \Delta_{t_j}^{t_k}.
\end{aligned} \tag{B.23}$$

This establishes both Proposition 4.2.1 and Theorem 4.2.4 simultaneously.

Step 4. The verification of the identities (B.9) - (B.10) is the final step. It is important to observe that (B.22) can be obtained directly from Eq. (B.14) - (B.17), irrespective of whether (B.13) holds. Utilizing (B.22), we get:

$$\begin{aligned}\mathbb{E}\Big[\widetilde{h}_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}p^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\Big] &= \mathbb{E}\Big[\big(h_{t_{k+1}} + \sum_{j=k+1}^{N}\prod_{\ell=k+1}^{j}\xi_\ell\Delta_{t_j}^{t_{k+1}}\big)\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\Big] \\ &= h_{t_{k+1}}\pi^{\delta}_{t_{k+1}}\mu_c^{\delta} + \sum_{j=k+1}^{N}\prod_{\ell=k+1}^{j}\xi_\ell\mathbb{E}\Big[\Delta_{t_j}^{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\Big].\end{aligned} \tag{B.24}$$

Then, with the conditional independence assumption of $(\mathbb{1}^{\pm}_{t_{k+1}}, c^{\pm}_{t_{k+1}}, p^{\pm}_{t_{k+1}})$ and $\{S_{t_{j+1}} - S_{t_j}\}_{j\geq k}$ given $\mathcal{F}_{t_k}$, for $j \geq k$, we have

$$\begin{aligned}\mathbb{E}\big[\Delta_{t_j}^{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] &= \mathbb{E}\big[(S_{t_{j+1}} - S_{t_j})\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] \\ &= \mathbb{E}\big[S_{t_{j+1}} - S_{t_j} \mid \mathcal{F}_{t_k}\big]\mathbb{E}\big[\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] \\ &= \Delta_{t_j}^{t_k}\pi^{\delta}_{t_{k+1}}\mu_c^{\delta}.\end{aligned} \tag{B.25}$$

Finally, we have $\mathbb{E}\big[\widetilde{h}_{t_{k+1}}\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big] = \widetilde{h}^{t_k}_{t_{k+1}}\mathbb{E}\big[\mathbb{1}^{\delta}_{t_{k+1}}c^{\delta}_{t_{k+1}} \mid \mathcal{F}_{t_k}\big]$. The proof of (B.10) follows a similar reasoning. □

B.2 Proof of Lemma 4.2.2

Proof. From the terminal condition, we have $\alpha_T = -\lambda < 0$. Therefore, we only need to show that $0 < \frac{\alpha_{t_k}}{\alpha_{t_{k+1}} - \phi} < 1$ whenever $\alpha_{t_{k+1}} < 0$. Substituting ${}^{(1)}A^{\pm}_{t_k}$ in Eq. (4.11) into Eq. (4.14), we have $\frac{\alpha_{t_k}}{\alpha_{t_{k+1}} - \phi} = 1 + \frac{N_k}{D_k}$, where

$$\begin{aligned}N_k =& \pi^{+}_{t_{k+1}}\pi^{-}_{t_{k+1}}(\alpha_{t_{k+1}} - \phi)\Big[(\mu_c^{+})^2\pi^{+}_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{-} - \mu_c^{-}\right) + (\mu_c^{-})^2\pi^{-}_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{+} - \mu_c^{+}\right)\Big] \\ & - 2\pi^{+}_{t_{k+1}}\pi^{-}_{t_{k+1}}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)^2(\mu_c^{+}\mu_c^{-})^2, \\ D_k =& \left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^{+}\mu_c^{-}\right]^2 - \pi^{+}_{t_{k+1}}\pi^{-}_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{+} - \mu_c^{+}\right)\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{-} - \mu_c^{-}\right).\end{aligned}$$

Hence, it is adequate to demonstrate that $N_k/D_k \in (-1, 0)$ whenever $\alpha_{t_{k+1}} < 0$. Initially, we establish that $D_k < 0$ and $N_k > 0$.

With the fact that $\pi_{t_{k+1}}(1,1) \leq \pi^+_{t_{k+1}} \wedge \pi^-_{t_{k+1}}$ and $\mu^\pm_{c^2} \geq (\mu^\pm_c)^2$, the first term in D_k satisfies

$$\left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^+_c\mu^-_c\right]^2 \leq \alpha^2_{t_{k+1}}\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left(\mu^+_c\mu^-_c\right)^2 \leq (\alpha_{t_{k+1}} - \phi)^2\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\mu^+_{c^2}\mu^-_{c^2}.$$

Moreover, when combined with the second term in D_k, and considering the assumption $\alpha_{t_{k+1}} < 0$ along with $\mu^\pm_\cdot \geq 0$, we obtain

$$D_k \leq \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left[(\alpha_{t_{k+1}} - \phi)\left(\mu^+_{c^2}\mu^-_c + \mu^+_c\mu^-_{c^2}\right) - \mu^+_c\mu^-_c\right] < 0. \tag{B.26}$$

Now, we proceed to prove that $N_k > 0$. Given $\alpha_{t_{k+1}} < 0$ and $\pi_{t_{k+1}}(1,1) \leq \pi^+_{t_{k+1}} \wedge \pi^-_{t_{k+1}}$, the first term in N_k satisfies

$$\begin{aligned}
&\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}(\alpha_{t_{k+1}} - \phi)\left[(\mu^+_c)^2\pi^+_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu^-_{c^2} - \mu^-_c\right) + (\mu^-_c)^2\pi^-_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu^+_{c^2} - \mu^+_c\right)\right] \\
&\geq \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\left[(\mu^+_c)^2\left((\alpha_{t_{k+1}} - \phi)\mu^-_{c^2} - \mu^-_c\right) + (\mu^-_c)^2\left((\alpha_{t_{k+1}} - \phi)\mu^+_{c^2} - \mu^+_c\right)\right].
\end{aligned}$$

Combining with the second term in N_k, we have

$$\begin{aligned}
N_k &\geq (\alpha_{t_{k+1}} - \phi)\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\pi_{t_{k+1}}(1,1)\big\{(\alpha_{t_{k+1}} - \phi)(\mu^+_c)^2\left[\mu^-_{c^2} - (\mu^-_c)^2\right] - \mu^+_c\mu^-_c(\mu^+_c + \mu^-_c) \\
&\qquad + (\alpha_{t_{k+1}} - \phi)(\mu^-_c)^2\left[\mu^+_{c^2} - (\mu^+_c)^2\right]\big\} \\
&\geq -(\alpha_{t_{k+1}} - \phi)\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\pi_{t_{k+1}}(1,1)\mu^+_c\mu^-_c(\mu^+_c + \mu^-_c) > 0.
\end{aligned}$$

The second inequality holds because $\mu^\pm_{c^2} \geq (\mu^\pm_c)^2$ and $\alpha_{t_{k+1}} < 0$. Therefore, $N_k/D_k < 0$. Next, we show whenever $\alpha_{t_{k+1}} < 0$, $N_k/D_k > -1$ or, equivalently, $D_k + N_k < 0$. Note

that

$$
\begin{aligned}
D_k + N_k = {} & \pi_{t_{k+1}}(1,1)\left((\alpha_{t_{k+1}} - \phi)\mu_c^+\mu_c^-\right)^2 (\pi(1,1) - 2\pi_{t_{k+1}}^+\pi_{t_{k+1}}^-) \\
& + (\alpha_{t_{k+1}} - \phi)(\pi_{t_{k+1}}^+)^2\pi_{t_{k+1}}^-(\mu_c^+)^2\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^- - \mu_c^-\right) \\
& + (\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}^+(\pi_{t_{k+1}}^-)^2(\mu_c^-)^2\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^+ - \mu_c^+\right) \\
& - \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^+ - \mu_c^+\right)\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^- - \mu_c^-\right).
\end{aligned} \tag{B.27}
$$

First, consider $N_k + D_k$ as a linear function of $\mu_{c^2}^+$, we have

$$
\begin{aligned}
\partial_{\mu_{c^2}^+}(N_k + D_k) &= \pi_{t_{k+1}}^+(\pi_{t_{k+1}}^-)^2(\alpha_{t_{k+1}} - \phi)^2(\mu_c^-)^2 - \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)^2\mu_{c^2}^- \\
&\quad + \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)\mu_c^- \\
&\le \pi_{t_{k+1}}^+(\pi_{t_{k+1}}^-)^2(\alpha_{t_{k+1}} - \phi)^2\mu_{c^2}^- - \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)^2\mu_{c^2}^- \\
&\quad + \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)\mu_c^- \qquad \text{(B.28)} \\
&= \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)^2\mu_{c^2}^-(\pi_{t_{k+1}}^- - 1) + \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)\mu_c^- < 0, \qquad \text{(B.29)}
\end{aligned}
$$

where (B.28) holds due to $\mu_{c^2}^- \ge (\mu_c^-)^2$, and (B.29) holds since $\pi_{t_{k+1}}^- < 1$ and $\alpha_{t_{k+1}} < 0$. Thus $N_k + D_k$ decrease with $\mu_{c^2}^+$. Since $\mu_{c^2}^+ \ge (\mu_c^+)^2$, substituting $\mu_{c^2}^+$ with $(\mu_c^+)^2$, we have

$$
\begin{aligned}
& D_k + N_k \\
& \le \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2\left[\pi_{t_{k+1}}^+\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^- - \mu_c^-\right) - \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2\right] \\
& + \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-(\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2\left[\pi_{t_{k+1}}^-\left((\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2 - \mu_c^+\right) - \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\left(\mu_c^+\right)^2\right] \\
& + \left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^+\mu_c^-\right]^2 - \pi_{t_{k+1}}^+\pi_{t_{k+1}}^-\left((\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2 - \mu_c^+\right)\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^- - \mu_c^-\right).
\end{aligned} \tag{B.30}
$$

Similarly, the RHS of (B.30) can be regarded as a linear decreasing function of $\mu_{c^2}^-$ since the coefficient of $\mu_{c^2}^-$ is

$$\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi)^2 (\mu_c^+)^2 (\pi_{t_{k+1}}^+ - 1) + \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi)\mu_c^+ < 0.$$

Given the fact that $\mu_{c^2}^- \geq (\mu_c^-)^2$, substituting $\mu_{c^2}^-$ with $(\mu_c^-)^2$ in the RHS of (B.30), we obtain

$$\begin{aligned}
&D_k + N_k \\
&\leq \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2 \left[\pi_{t_{k+1}}^+ \left((\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2 - \mu_c^-\right) - \pi_{t_{k+1}}(1,1)\alpha_{t_{k+1}}(\mu_c^-)^2\right] \\
&\quad + \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2 \left[\pi_{t_{k+1}}^- \left((\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2 - \mu_c^+\right) - \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2\right] \\
&\quad + \left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^+ \mu_c^-\right]^2 \\
&\quad - \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \left[(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2 - \mu_c^+\right] \left[(\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2 - \mu_c^-\right] \\
&= \mu_c^+ \mu_c^- \Big[(\pi_{t_{k+1}}^+)^2 \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi)\mu_c^+ \left((\alpha_{t_{k+1}} - \phi)\mu_c^- - 1\right) \\
&\qquad\qquad - 2\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)^2 \mu_c^+ \mu_c^- \\
&\qquad\qquad + \pi_{t_{k+1}}^+ (\pi_{t_{k+1}}^-)^2 (\alpha_{t_{k+1}} - \phi)\mu_c^- \left((\alpha_{t_{k+1}} - \phi)\mu_c^+ - 1\right) \\
&\qquad\qquad + \pi_{t_{k+1}}^2 (1,1)(\alpha_{t_{k+1}} - \phi)^2 \mu_c^+ \mu_c^- \\
&\qquad\qquad - \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \left((\alpha_{t_{k+1}} - \phi)\mu_c^+ - 1\right) \left((\alpha_{t_{k+1}} - \phi)\mu_c^- - 1\right) \Big] \\
&\triangleq \mu_c^+ \mu_c^- \ell(\mu_c^+, \mu_c^-).
\end{aligned} \tag{B.31}$$

To prove $D_k + N_k < 0$, it suffices to demonstrate that $\ell(\mu_c^+, \mu_c^-) < 0$. Note that $\ell(\mu_c^+, \mu_c^-)$ is a linear function in μ_c^+, and the coefficient of μ_c^+ is given by

$$\partial_{\mu_c^+} \ell(\mu_c^+, \mu_c^-) = m \left(\pi_{t_{k+1}}(1,1)\right) (\alpha_{t_{k+1}} - \phi)^2 \mu_c^- + (\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (1 - \pi_{t_{k+1}}^+),$$

where

$$m\left(\pi_{t_{k+1}}(1,1)\right) := (\pi^+_{t_{k+1}})^2\pi^-_{t_{k+1}} - 2\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\pi_{t_{k+1}}(1,1) + \pi^+_{t_{k+1}}(\pi^-_{t_{k+1}})^2 + \pi^2_{t_{k+1}}(1,1) - \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}.$$

We temporarily assume the validity of $m\left(\pi_{t_{k+1}}(1,1)\right) \leq 0$ for any $\pi_{t_{k+1}}(1,1)$ in Eq. (4.1), and we will provide the proof subsequently. Given $\mu^-_c \geq 0$, substituting 0 into μ^-_c yields $\partial_{\mu^+_c}\ell\left(\mu^+_c, \mu^-_c\right) \leq (\alpha_{t_{k+1}} - \phi)\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left(1 - \pi^+_{t_{k+1}}\right) < 0$. Consequently, $\ell\left(\mu^+_c, \mu^-_c\right)$ decreases with μ^+_c. Since $\mu^+_c \geq 0$, we have that

$$\begin{aligned}\ell\left(\mu^+_c, \mu^-_c\right) \leq \ell\left(0, \mu^-_c\right) &= -\pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left(1 - (\alpha_{t_{k+1}} - \phi)\mu^-_c\right) - \pi^+_{t_{k+1}}(\pi^-_{t_{k+1}})^2(\alpha_{t_{k+1}} - \phi)\mu^-_c \\ &= -\pi^+_{t_{k+1}}\pi^-_{t_{k+1}} + \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}(\alpha_{t_{k+1}} - \phi)\mu^-_c(1 - \pi^-_{t_{k+1}}) < 0.\end{aligned}$$

Therefore, $\ell\left(\mu^+_c, \mu^-_c\right) < 0$ for any $\mu^\pm_c \geq 0$. We then have $D_k + N_k < 0$ from Eq. (B.31), which implies that $N_k/D_k > -1$.

It remains to show that $m\left(\pi_{t_{k+1}}(1,1)\right) \leq 0$ holds for any $\pi_{t_{k+1}}(1,1)$ in Eq. (4.1). From Eq. (4.1), we know that $(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \vee 0 \leq \pi_{t_{k+1}}(1,1) \leq \pi^+_{t_{k+1}} \wedge \pi^-_{t_{k+1}}$. Recognizing that $m\left(\pi_{t_{k+1}}(1,1)\right)$ is a quadratic function of $\pi_{t_{k+1}}(1,1)$ opening upwards, we focus on confirming that the values of $m\left(\pi_{t_{k+1}}(1,1)\right)$ at the two endpoints $(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \vee 0$ and $\pi^+_{t_{k+1}} \wedge \pi^-_{t_{k+1}}$ are non-positive. Without loss of generality, we assume $\pi^-_{t_{k+1}} \leq \pi^+_{t_{k+1}}$. Initially, we examine $m\left(\pi^-_{t_{k+1}}\right) \leq 0$:

$$\begin{aligned}m(\pi^-_{t_{k+1}}) &= (\pi^-_{t_{k+1}})^2 - 2\pi^+_{t_{k+1}}(\pi^-_{t_{k+1}})^2 + \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \\ &= (\pi^+_{t_{k+1}} - \pi^-_{t_{k+1}})\pi^-_{t_{k+1}}(\pi^+_{t_{k+1}} - 1) \leq 0.\end{aligned}$$

Continuing our examination, we proceed to verify $m\left((\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \vee 0\right) \leq 0$. In the case where $\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1 < 0$, we immediately ascertain that $m(0) = \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}(\pi^+_{t_{k+1}} +$

$\pi^-_{t_{k+1}} - 1) \leq 0$. On the other hand, if $\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1 \geq 0$,

$$\begin{aligned} m(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) &= (\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1)^2 - \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \\ &= (1 - \pi^-_{t_{k+1}})(\pi^+_{t_{k+1}})^2 + (1 - \pi^-_{t_{k+1}})(\pi^-_{t_{k+1}} - 2)\pi^+_{t_{k+1}} + (\pi^-_{t_{k+1}} - 1)^2 \\ &\triangleq n(\pi^+_{t_{k+1}}). \end{aligned}$$

Considering $n(\pi^+_{t_{k+1}})$, a quadratic function with an upward opening, we note that, under the assumption $\pi^-_{t_{k+1}} \leq \pi^+_{t_{k+1}} \leq 1$ and $\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1 \geq 0$, the feasible range for $\pi^+_{t_{k+1}}$ is given by

$$\begin{cases} 1 - \pi^-_{t_{k+1}} \leq \pi^+_{t_{k+1}} \leq 1, & \text{when } 0 \leq \pi^-_{t_{k+1}} \leq 0.5, \\ \pi^-_{t_{k+1}} \leq \pi^+_{t_{k+1}} \leq 1, & \text{when } 0.5 \leq \pi^-_{t_{k+1}} \leq 1. \end{cases}$$

It suffices to examine the non-positivity of $n(\pi^+_{t_{k+1}})$ at the boundary:

$$n(1) = (1 - \pi^-_{t_{k+1}}) + (1 - \pi^-_{t_{k+1}})(\pi^-_{t_{k+1}} - 2) + (\pi^-_{t_{k+1}} - 1)^2 = 0.$$

When $0 \leq \pi^-_{t_{k+1}} \leq 0.5$,

$$n(1 - \pi^-_{t_{k+1}}) = (1 - \pi^-_{t_{k+1}})^3 + (1 - \pi^-_{t_{k+1}})^2(\pi^-_{t_{k+1}} - 2) + (\pi^-_{t_{k+1}} - 1)^2 = 0.$$

When $0.5 \leq \pi^-_{t_{k+1}} \leq 1$,

$$n(\pi^-_{t_{k+1}}) = (1 - \pi^-_{t_{k+1}})(\pi^-_{t_{k+1}})^2 + (1 - \pi^-_{t_{k+1}})(\pi^-_{t_{k+1}} - 2)\pi^-_{t_{k+1}} + (\pi^-_{t_{k+1}} - 1)^2 \leq 0.$$

Consequently, $m(\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1) \leq 0$ when $\pi^+_{t_{k+1}} + \pi^-_{t_{k+1}} - 1 \geq 0$. This concludes the verification of the assertion that $m\left(\pi_{t_{k+1}}(1,1)\right) \leq 0$ holds for any $\pi_{t_{k+1}}(1,1)$ in Eq. (4.1).

□

B.3 Proof of Proposition 4.2.5

Proof. Initially, we establish the result under the martingale condition (4.7). Utilizing Eq. (4.10), our objective is to demonstrate that

$$L_{t_k}^{+,*} + L_{t_k}^{-,*} = ({}^{(1)}A_{t_k}^{+} - {}^{(1)}A_{t_k}^{-})I_{t_k} + ({}^{(2)}A_{t_k}^{+} - {}^{(2)}A_{t_k}^{-}) + ({}^{(3)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{-}) > 0.$$

Given the conditions (4.21)-(4.22) in Proposition 4.2.5, it is evident that

$$\begin{aligned}\beta_{t_k}^{+} - \beta_{t_k}^{-} = {} & \pi_{t_{k+1}}^{+}\pi_{t_{k+1}}^{-}(\alpha_{t_{k+1}} - \phi)(\mu_c^{+}\mu_{c^2}^{-} - \mu_c^{-}\mu_{c^2}^{+}) \\ & - \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^{-}\mu_c^{+}(\pi_{t_{k+1}}^{-}\mu_c^{-} - \pi_{t_{k+1}}^{+}\mu_c^{+}) = 0\end{aligned}$$

This implies that ${}^{(1)}A_{t_k}^{+} - {}^{(1)}A_{t_k}^{-} = 0$ and ${}^{(2)}A_{t_k}^{+} - {}^{(2)}A_{t_k}^{-} = 0$. The subsequent demonstration focuses on establishing ${}^{(3)}A_{t_k}^{+} - {}^{(3)}A_{t_k}^{-} > 0$. To achieve this, it is essential to note, as indicated in Eq. (B.26), that γ_{t_k} (i.e., D_k), the denominator of ${}^{(3)}A_{t_k}^{+} - {}^{(3)}A_{t_k}^{-}$, is negative. Consequently, our task is now to confirm that the numerator of ${}^{(3)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{-}$ is also negative. According to condition (4.23) in Proposition 4.2.5, the numerator of ${}^{(3)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{-}$ can be expressed as

$$\begin{aligned}N\left({}^{(3)}A_{t_k}^{+} + {}^{(3)}A_{t_k}^{-}\right) = {} & \Big\{\pi_{t_{k+1}}^{+}\pi_{t_{k+1}}^{-}\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{-} - \mu_c^{-}\right)\left(\mu_c^{+} - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{+}\right) \\ & + 2\left[(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^{+}\mu_c^{-}\right]^2 \\ & - \pi_{t_{k+1}}^{+}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^{+})^2\mu_c^{-}\Big\}\mu_p^{+} \\ & + \Big\{\pi_{t_{k+1}}^{+}\pi_{t_{k+1}}^{-}\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{+} - \mu_c^{+}\right)\left(\mu_c^{-} - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}^{-}\right) \\ & + 2\left[(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^{+}\mu_c^{-}\right]^2 \\ & - \pi_{t_{k+1}}^{-}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^{-})^2\mu_c^{+}\Big\}\mu_p^{-}.\end{aligned}$$

Next, we can demonstrate that the coefficients of μ_p^+ is negative. Specifically, let's denote the coefficients of μ_p^+ as $r(\mu_{c^2}^+, \mu_{c^2}^-)$, constituting a linear function of $\mu_{c^2}^-$ with a coefficient $\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- (\alpha_{t_{k+1}} - \phi) \left(\mu_c^+ - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}^+\right) < 0$. Given that $\mu_{c^2}^- \geq (\mu_c^-)^2$, it follows that:

$$
\begin{aligned}
r(\mu_{c^2}^+, \mu_{c^2}^-) &\leq r\left(\mu_{c^2}^+, (\mu_c^-)^2\right) \\
&= \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \left((\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2 - \mu_c^-\right) \left(\mu_c^+ - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}^+\right) \\
&\quad + 2\left[(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^+\mu_c^-\right]^2 - \pi_{t_{k+1}}^+ \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2\mu_c^-.
\end{aligned}
$$

Similarly, $r\left(\mu_{c^2}^+, (\mu_c^-)^2\right)$ is a linear function of $\mu_{c^2}^+$ with a coefficient $-2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \left((\alpha_{t_{k+1}} - \phi)\left(\mu_c^-\right)^2 - \mu_c^-\right) < 0$. Therefore,

$$
\begin{aligned}
r\left(\mu_{c^2}^+, (\mu_c^-)^2\right) &\leq r\left((\mu_c^+)^2, (\mu_c^-)^2\right) \\
&= \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \left((\alpha_{t_{k+1}} - \phi)(\mu_c^-)^2 - \mu_c^-\right) \left(\mu_c^+ - 2(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2\right) \\
&\quad + 2\left[(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}(1,1)\mu_c^+\mu_c^-\right]^2 - \pi_{t_{k+1}}^+ \pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)(\mu_c^+)^2\mu_c^- \\
&= \mu_c^+\mu_c^- \Big\{ \left[2(\alpha_{t_{k+1}} - \phi)^2 \pi_{t_{k+1}}^2(1,1) - 2(\alpha_{t_{k+1}} - \phi)^2 \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^-\right]\mu_c^+\mu_c^- \\
&\qquad + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \mu_c^+ + \pi_{t_{k+1}}^+ (\alpha_{t_{k+1}} - \phi)\left[\pi_{t_{k+1}}^- \mu_c^- - \pi_{t_{k+1}}(1,1)\mu_c^+\right] \\
&\qquad - \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \Big\}.
\end{aligned}
$$

According to Eq. (4.1), $\pi_{t_{k+1}}(1,1) \leq \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^-$. Additionally, as indicated by Lemma 4.2.2, $\alpha_{t_{k+1}} < 0$. Consequently, the sum of the first two terms within the curly brackets, denoted as $\left[2(\alpha_{t_{k+1}} - \phi)^2 \pi_{t_{k+1}}^2(1,1) - 2(\alpha_{t_{k+1}} - \phi)^2 \pi_{t_{k+1}}^+ \pi_{t_{k+1}}^-\right]\mu_c^+\mu_c^- + 2(\alpha_{t_{k+1}} - \phi)\pi_{t_{k+1}}^+ \pi_{t_{k+1}}^- \mu_c^+$, is negative.

Under the Condition (4.21), the third term within the curly brackets, namely $\pi_{t_{k+1}}^+ (\alpha_{t_{k+1}} - \phi)\left[\pi_{t_{k+1}}^- \mu_c^- - \pi_{t_{k+1}}(1,1)\mu_c^+\right]$, can be expressed as $\pi_{t_{k+1}}^+ (\alpha_{t_{k+1}} - \phi)\left(\pi_{t_{k+1}}^- - \pi_{t_{k+1}}(1,1)\right)\mu_c < 0$. Consequently, the coefficient of μ_p^+ in $N({}^{(3)}A_{t_k}^+ + {}^{(3)}A_{t_k}^-)$, represented by $r(\mu_{c^2}^+, \mu_{c^2}^-)$, is

negative. Similarly, the coefficient of μ_p^- is also negative. Thus, $N({}^{(3)}A_{t_k}^+ + {}^{(3)}A_{t_k}^-) < 0$ and ${}^{(3)}A_{t_k}^+ + {}^{(3)}A_{t_k}^- > 0$.

The demonstration for the non-martingale case is straightforward. As highlighted earlier, conditions (4.21)-(4.22) ascertain $\beta_{t_k}^+ - \beta_{t_k}^- = 0$. Moreover, according to (4.17), we obtain

$$\widetilde{L}_{t_k}^{+,*} + \widetilde{L}_{t_k}^{-,*} = L_{t_k}^{+,*} + L_{t_k}^{-,*}.$$

□

B.4 Proof of Corollary 4.3.1

Proof. Under the Conditions (4.21)-(4.22), a straightforward verification reveals $\beta_{t_k}^+ = \beta_{t_k}^-$. Subsequently, based on (4.17), it becomes evident that the optimal spreads, hereinafter denoted as $Sprd_{t_k}$, remain consistent across both the martingale and non-martingale mid-price scenarios. Furthermore,

$$Sprd_{t_k} = \widetilde{L}_{t_k}^{+,*} + \widetilde{L}_{t_k}^{-,*} = L_{t_k}^{+,*} + L_{t_k}^{-,*} = {}^{(3)}A_{t_k}^+ + {}^{(3)}A_{t_k}^-, \tag{B.32}$$

which establishes the independence of optimal spreads from both inventory levels and local drifts, denoted as $\{\Delta_{t_k}\}_{k=0,\dots,N}$. With additional assumptions of Condition (4.23) and Condition (4.26), the optimal spread assumes the form

$$Sprd_{t_k} = \frac{\left[\pi\left(\mu_c - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right) + 2(\alpha_{t_{k+1}} - \phi)\pi(1,1)\mu_c^2\right]\left(\mu_p^+ + \mu_p^-\right)}{2\left[\pi(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^2 - \pi\big((\alpha_{t_{k+1}} - \phi)\mu_{c^2} - \mu_c\big)\right]}, \tag{B.33}$$

where $\pi = \pi^{\pm}$.

We check the difference between $Sprd_{t_k}$ and $Spd_{t_{k-1}}$ here.

$$Sprd_{t_k} - Sprd_{t_{k-1}} = \frac{\mu_p^+ + \mu_p^-}{2} \cdot \frac{(\alpha_{t_{k+1}} - \alpha_{t_k})\pi\mu_c\big(\pi(1,1)\mu_c^2 - \pi\mu_{c^2}\big)}{\prod_{\ell=k,k+1}\left[\pi(1,1)(\alpha_{t_\ell} - \phi)\mu_c^2 - \pi\big((\alpha_{t_\ell} - \phi)\mu_{c^2} - \mu_c\big)\right]}. \tag{B.34}$$

Initially, we establish the positivity of the denominator. Given the negativity of α_{t_k} and by definition $0 \leq \pi(1,1) \leq \pi$ and $0 < \mu_c^2 \leq \mu_{c^2}$, we have

$$\pi(1,1)\alpha_{t_\ell}\mu_c^2 - \pi\left(\alpha_{t_\ell}\mu_{c^2} - \mu_c\right) \geq \pi\alpha_{t_\ell}\mu_{c^2} - \pi\left(\alpha_{t_\ell}\mu_{c^2} - \mu_c\right) = \pi\mu_c > 0.$$

This demonstrates the positivity of the denominator.

Regarding the numerator, we observe that $\pi(1,1)\mu_c^2 \leq \pi\mu_{c^2}$. Therefore, the positivity of the numerator relies on the positivity of $\alpha_{t_{k+1}} - \alpha_{t_k}$, which is determined by the value of ϕ according to the paragraph right after the Lemma 4.2.2.

Particularly, if $\pi(1,1)\mu_c^2 = \pi\mu_{c^2}$, $Sprd_{t_k} - Sprd_{t_{k-1}} = 0$.

To demonstrate the optimal spreads decrease with $\pi(1,1)$ at a given time instance, we have

$$\begin{aligned}
\partial_{\pi(1,1)}Sprd_{t_k} &= \frac{(\alpha_{t_{k+1}} - \phi)\mu_c^2(\mu_p^+ + \mu_p^-)\left[\pi(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^2 - \pi\big((\alpha_{t_{k+1}} - \phi)\mu_{c^2} - \mu_c\big)\right]}{\left[\pi(1,1)(\alpha_{t_{k+1}} - \phi)\mu_c^2 - \pi\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2} - \mu_c\right)\right]^2} \\
&\quad - \frac{(\alpha_{t_{k+1}} - \phi)\mu_c^2\left[\pi\big(\mu_c - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}\big) + 2(\alpha_{t_{k+1}} - \phi)\pi(1,1)\mu_c^2\right](\mu_p^+ + \mu_p^-)}{2\left[\pi(1,1)\alpha_{t_{k+1}}\mu_c^2 - \pi\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2} - \mu_c\right)\right]^2} \\
&= \frac{(\alpha_{t_{k+1}} - \phi)\mu_c^3\left(\mu_p^+ + \mu_p^-\right)}{2\left[\pi(1,1)\alpha_{t_{k+1}}\mu_c^2 - \pi\left((\alpha_{t_{k+1}} - \phi)\mu_{c^2} - \mu_c\right)\right]^2} < 0.
\end{aligned}$$

The proof of corollary 4.3.1 is now concluded. □

B.5 Proof of Corollary 4.3.2

Proof. Recall

$$\widetilde{a}^*_{t_k} = S_{t_k} + {}^{(1)}A^+_{t_k} I_{t_k} + {}^{(2)}\widetilde{A}^+_{t_k} + {}^{(3)}\widetilde{A}^+_{t_k},$$
$$\widetilde{b}^*_{t_k} = S_{t_k} - {}^{(1)}A^-_{t_k} I_{t_k} - {}^{(2)}\widetilde{A}^-_{t_k} + {}^{(3)}\widetilde{A}^-_{t_k}.$$

To establish the strictly decreasing monotonicity of $\widetilde{a}^*_{t_k}$ and $\widetilde{b}^*_{t_k}$ with respect to I_{t_k}, we need to demonstrate that ${}^{(1)}A^{\pm}_{t_k} < 0$. According to Proposition 4.2.1, we have that ${}^{(1)}A^{\pm}_{t_k} = \frac{\beta^{\pm}_{t_k}(\alpha_{t_{k+1}} - \phi)}{\gamma_{t_k}}$ and

$$\gamma_{t_k} = \left[\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^+_c \mu^-_c\right]^2 - \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu^+_{c^2} - \mu^+_c\right)\left((\alpha_{t_{k+1}} - \phi)\mu^-_{c^2} - \mu^-_c\right),$$
$$\beta^{\pm}_{t_k} = \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\mu^{\pm}_c\left((\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2} - \mu^{\mp}_c\right) - \pi^{\mp}_{t_{k+1}}\pi_{t_{k+1}}(1,1)(\alpha_{t_{k+1}} - \phi)\mu^{\pm}_c\left(\mu^{\mp}_c\right)^2.$$

Given that $\pi_{t_{k+1}}(1,1) \le \pi^+_{t_{k+1}} \wedge \pi^-_{t_{k+1}}$, along with $\left(\mu^{\pm}_c\right)^2 \le \mu^{\pm}_{c^2}$ and the negativity of $\alpha_{t_{k+1}}$ as established in Lemma 4.2.2,

$$\begin{aligned}
\gamma_{t_k} &\le \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left\{(\alpha_{t_{k+1}} - \phi)^2\left[(\mu^+_c)^2(\mu^-_c)^2 - \mu^+_{c^2}\mu^-_{c^2}\right] + (\alpha_{t_{k+1}} - \phi)(\mu^+_{c^2}\mu^-_c + \mu^+_c\mu^-_{c^2}) - \mu^+_c\mu^-_c\right\} \\
&\le \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left[(\alpha_{t_{k+1}} - \phi)(\mu^+_{c^2}\mu^-_c + \mu^+_c\mu^-_{c^2}) - \mu^+_c\mu^-_c\right] < 0, \\
\beta^{\pm}_{t_k} &\le \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left[(\alpha_{t_{k+1}} - \phi)\mu^{\mp}_{c^2}\mu^{\pm}_c - \mu^{\pm}_c\mu^{\mp}_c - (\alpha_{t_{k+1}} - \phi)\mu^{\pm}_c(\mu^{\mp}_c)^2\right] \\
&= \pi^+_{t_{k+1}}\pi^-_{t_{k+1}}\left\{(\alpha_{t_{k+1}} - \phi)\mu^{\pm}_c[\mu^{\mp}_{c^2} - (\mu^{\mp}_c)^2] - \mu^{\pm}_c\mu^{\mp}_c\right\} < 0,
\end{aligned}$$

Therefore, it follows that ${}^{(1)}A^{\pm}_{t_k} < 0$ for any t_k. □

B.6 Proof of Corollary 4.3.3

Proof. Given the assumptions in Corollary 4.3.3, the expressions for the optimal strategies can be formulated as follows

$$a^*_{t_k} = S_{t_k} + \overbrace{\frac{(\alpha_{t_{k+1}} - \phi)\mu_c}{\mu_c - \alpha_{t_{k+1}}\mu_{c^2}} I_{t_k} + \frac{\mu_c - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}\mu_p + \frac{h_{t_{k+1}}\mu_c}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}}^{L^{+,*}_{t_k}}, \tag{B.35}$$

$$b^*_{t_k} = S_{t_k} + \overbrace{\frac{(\alpha_{t_{k+1}} - \phi)\mu_c}{\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}} I_{t_k} - \frac{\mu_c - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}\mu_p + \frac{h_{t_{k+1}}\mu_c}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}}^{-L^{-,*}_{t_k}}, \tag{B.36}$$

where

$$\alpha_{t_k} = (\alpha_{t_{k+1}} - \phi) + \frac{2\pi_{t_{k+1}}\left((\alpha_{t_{k+1}} - \phi)\mu_c\right)^2}{\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}}, \quad h_{t_k} = h_{t_{k+1}} + \frac{2\pi_{t_{k+1}}(\alpha_{t_{k+1}} - \phi)\mu_c^2 h_{t_{k+1}}}{\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}},$$

and $\pi_{t_{k+1}} := \pi^{\pm}_{t_{k+1}}$. Given that $h_T = 0$, it follows that h_{t_k} remains zero throughout the trading horizon. Upon closer examination, it becomes evident that regardless of the specific time t_k and the penalty levels leading to distinct values of α_{t_k}, the optimal ask price can be expressed as $a^*_{t_k} = S_{t_k} + \frac{\mu_p}{2}$ when the inventory level $I_{t_k} = \bar{I}^+ = \frac{\mu_{c^2}\mu_p}{2\mu_c}$. Similarly, the optimal bid price is given by $b^*_{t_k} = S_{t_k} - \frac{\mu_p}{2}$ when $I_{t_k} = \bar{I}^- = -\frac{\mu_{c^2}\mu_p}{2\mu_c}$.

First, let us examine the case where the inventory level is non-negative. Specifically, when $I_{t_k} = 0$, we can deduce from Eq. (B.36) that the optimal bid price is as follows:

$$S_{t_k} - \frac{\mu_c - 2(\alpha_{t_{k+1}} - \phi)\mu_{c^2}}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}\mu_p = S_{t_k} - \frac{\mu_p}{2} + \frac{(\alpha_{t_{k+1}} - \phi)\mu_{c^2}}{2\left[\mu_c - (\alpha_{t_{k+1}} - \phi)\mu_{c^2}\right]}\mu_p < S_{t_k} - \frac{\mu_p}{2},$$

since $\alpha_{t_{k+1}} < 0$. As indicated in Corollary 4.3.2, the optimal bid price exhibits a strict decreasing trend with respect to the inventory level. Consequently, when $I_{t_k} \geq 0$, it can be concluded that the optimal bid price is consistently lower than $S_{t_k} - \frac{\mu_p}{2}$.

As mentioned in Corollary 4.3.2, the optimal ask price exhibits a strict decreasing relationship with the inventory level. Building upon the previous discussion, it can be inferred that the optimal ask price has the following characteristics:

$$a^*_{t_k} = S_{t_k} + \frac{\mu_p}{2} \text{ when } I_{t_k} = \bar{I}^+ = \frac{\mu_{c^2}\mu_p}{2\mu_c}.$$

Therefore, when considering the range of inventory levels $I_{t_k} \in [0, \bar{I}^+)$, it is evident that the optimal ask price consistently surpasses the threshold of $S_{t_k} + \frac{\mu_p}{2}$. Conversely, for inventory levels exceeding $\bar{I}^+$, the optimal ask price consistently falls below $S_{t_k} + \frac{\mu_p}{2}$. The symmetric nature of the proof holds true in the case of non-positive inventory levels. Thus, the proof of Corollary 4.3.3 is substantiated.

□