Delft University of Technology
Faculty of EEMCS

A Static and Dynamic Approach to the Optimal Placement Problem

by
J.A.W. Claes

to obtain the degree of Master of Science in Applied Mathematics - Financial Engineering
at the Delft University of Technology,
to be defended publicly on Friday August 26th, 2025 at 15:00.

Student number:	4935802		
Project duration:	September 2, 2024 – August 26, 2025		
Thesis committee:	Dr. F. Yu,	TU Delft,	Daily supervisor
	Prof.dr. A. Papapantoleon,	TU Delft,	Chair
	Dr. N. Parolya,	TU Delft,	Committee member

TUDelft Delft University of Technology

Abstract

The digitalisation of financial markets has led to an increase in the share of trades being executed by algorithmic trading systems. A key application of these algorithms arises when a trader needs to buy or sell a large volume of shares within a fixed time window. The trader must decide between placing reliable but costly market orders or cheaper limit orders, at a risk of not filling her target amount within the desired time frame. This is known as the Optimal Placement Problem (OPP) and this thesis aims to provide a general solution to this problem.

Our research builds forth upon two models. The first, proposed by Cont and Kukanov [2017] [15], formulates the problem as a stochastic convex optimization problem. This model provides a static solution to the OPP, where the trader can only place orders at the initial time. We extend their model beyond the best quote price-level to consider any number of price-levels. This generalization still allows us to derive analytical expressions for the optimal placement strategy, including a closed-form solution under the assumption that market and limit order arrivals follow exponential distributions.

However, large deviations in the calibrated parameters or distributions can lead to substantially higher costs. A dynamic strategy that allows the trader to adjust her placement strategy at intermediate times offers increased flexibility and resilience against such unexpected order flow deviations. The second model, proposed by Cartea and Jaimungal [2015] [5], provides such a dynamic solution and frames the problem as a stochastic optimal stopping and control problem. We extend this model to include mid-price drift and a general running inventory penalty, and formulate a recursive expression for the general optimal strategy for an inventory of any size.

The static solution gives an accurate, efficient and robust indication of the optimal allocation for larger inventories and longer trading windows. The dynamic solution, on the other hand, is more effective at liquidating (or acquiring) smaller inventories in shorter time frames at an optimal price, by providing a precise, adaptive strategy in volatile market conditions.

Overall, this thesis improves our understanding of the Optimal Placement Problem by analytically deriving its solution under two distinct model setups: the static model of [15] and the dynamic model of [5]. In the static approach, we find the optimal allocation switches from limit orders at one price-level to the next at distinct critical points. This provides an intuitive and broadly applicable rule of thumb for practical implementation, leading to lower expected costs across the board. The optimal strategy, derived from the dynamic model, leads to higher average earnings per share than benchmark models like the Time Weighted Average Price, the solution only considering limit orders and the solution found in [5], by taking into account market orders and possible mid-price drift.

Future extensions, such as testing on different arrival distributions, depth discretisation, parameter calibration and using heuristic methods to improve computational speed, could further assess the practical viability and performance trade-offs of these strategies in real-world trading systems.

Keywords: Limit Order Books, Optimal Order Placement, Convex Optimization, Optimal Control Theory, Optimal Stopping Theory, Dynamic Programming Principle, Queueing Systems, Hamilton-Jacobi-Bellman Equation, Algorithmic Trading

Acknowledgements

The work before you is written to obtain the degree Master of Science in Applied Mathematics – Financial Engineering at TU Delft. This thesis was an incredible journey and taught me a lot about what it means to be a mathematician, do research and overcome obstacles. This thesis also marks the end of my time as a student in Delft, which I could not have enjoyed more.

I would like to express my gratitude toward dr. Fenghui Yu for her time and help during this project and for the useful discussions we have had regarding the project. Her guidance helped me find innovative ways to tackle problems and take the quality of my thesis to a higher academic level. I would like to thank the Chair of my committee Prof. dr. Antonis Papapantoleon and the final member of my committee Dr. Nestor Parolya for being a part of my thesis committee.

Writing a thesis can be a long and painstaking process, but what I will take away from this final part of my studies is mainly the joy I had during the project. I would like to thank my family, girlfriend and friends for always being there for me and helping me when I had trouble finding motivation or direction.

I wish you a pleasant read,

Joep Claes
Delft, August 2025

Contents

List of Figures

List of Tables

1 | Introduction

Technological advancements in financial markets have led to an increase in trades being performed by algorithms. In 2019, algorithmic trading accounted 92% of all traded equity volume in the United States [30]. Traders range from small retail investors to specialised market-makers, each trading not only for speculative investment purposes, but also for risk hedging strategies, arbitrage, and liquidity provision.

Electronic platforms replaced traditional floor trading, enabling traders to trade at ultra-high speeds. High-frequency trading (HFT) firms use algorithms to place and cancel orders within microseconds. Traditional financial institutions like banks and hedge funds have also embraced algorithms to execute large volumes of trades. This shift drives the need for fast and efficient algorithms and the ability to process large amounts of data in real-time.

A key challenge in trade execution is buying or selling a large amount of shares in a finite time horizon. To mitigate negative price impact, the trader must decide how to split up the large parent order into smaller child orders. This is known as the Optimal Execution Problem, and has been vastly studied in the literature. Once this order scheduling decision is made, a trader must decide how to allocate each child order between market orders (which execute immediately but incur higher costs) and limit orders (which offer better prices but may not be filled in time). This is called the Optimal Placement Problem and will be the focus of this thesis.

Solving the Optimal Placement Problem is of critical importance. Even small improvements in execution efficiency can yield substantial cost savings. For example, Almgren and Chriss [2001] [1] estimate that a 10 basis point improvement in execution price translates to millions of dollars saved annually for large institutions. Poorly placed orders may either miss execution entirely or lead to unfavourable price impact due to adverse selection. Moreover, the decision must be made under uncertainty, as future order flow, price movements, and cancellations are inherently stochastic.

In this thesis we aim to derive tractable analytical solutions to the Optimal Placement Problem, providing insight into how order book features, such as queue size, spread, depth, and cancellation rates, affect optimal order allocation. First, we build on the static model proposed by Cont and Kukanov [2017] [15], which poses the problem as a stochastic convex optimization problem. Next, we explore a dynamic model, as proposed by Cartea and Jaimungal [2015] [5], which poses the problem as a stochastic optimal stopping and control problem in which the trader optimally adapts her placement strategy over time in response to evolving market conditions. Together, these two approaches offer complementary perspectives: one focusing on immediate placement under static assumptions, and the other incorporating adaptive decision-making in a dynamic environment.

1.1 Limit Orders Books

In order to buy any asset, someone who owns that asset has to be willing to sell it. In financial markets, trading occurs through the matching of these buyers and sellers. For trading to take place, there must be a buyer willing to purchase at a certain price and a seller willing to sell at that same price. This is done by placing orders in a digital order book, where all outstanding buy and sell orders are recorded

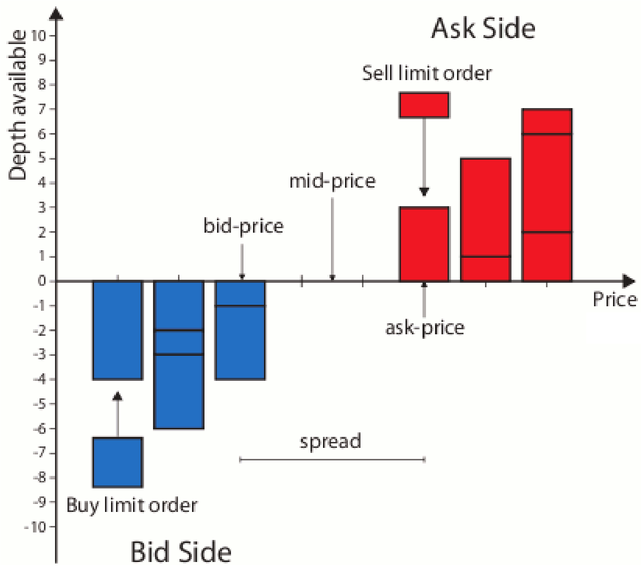

Figure 1.1: A schematic representation of the limit order book with a price-time priority [18]

and matched according to specific rules.

While there are many trading mechanisms (e.g. request-for-quote, dark pools and over the counter markets) trading is often done in limit order books, which are crucial to the functioning of modern financial markets. A limit order book is a record of all outstanding limit orders for a particular asset (share, bond, commodity etc.), organized by price-level. It displays the buy orders (bids) and sell orders (asks) that have been placed but not yet executed. These orders are placed as buy/sell *limit orders*, which indicate that a trader is willing to buy/sell at the price at which the limit order is placed. The order book consists of discrete price-levels with one "tick" difference between each level. The best bid is the highest price someone is willing to buy at, and the best ask is the lowest price someone is willing to sell at. The difference between the best bid and best ask is known as the (bid-ask) spread. The depth of a placed limit order refers to how far below (for a buy order) or above (for a sell order) the best bid/ask the order is placed. When a *market order* is placed, it is immediately executed at the best available price, which means it will match with the best bid (if it is a sell market order) or ask (if it is a buy market order) in the book, filling the first outstanding limit order at the best bid/ask price.

Most exchanges employ a price-time priority policy to organize limit orders. At a given price-level, orders are executed in the order they were received (first-in, first-out or FIFO). This creates queues at each price-level: a newly submitted limit order joins the back of the queue. When a matching order arrives, it is matched with the first order in the queue. If a market order exceeds the size of the best quote, it walks up the book, executing against multiple levels. This queueing mechanism ensures fairness and transparency in the trading process.

Mathematically, a limit order book (LOB) can be modelled as a multi-level queueing system, where each price-level represents a queue of buy or sell orders. The arrival processes include market orders (MOs), which are executed immediately, limit orders (LOs), which are added to the book at the specified price-level, and cancellations, which remove outstanding limit orders from the book. These three processes dictate the dynamics of the entire LOB and can be described stochastically, such as Poisson processes [15], while the queue dynamics at each level can be modelled as Markov jump processes or diffusion approximations [34][19]. These models capture the probabilistic nature of order arrivals, cancellations, and executions, providing a framework for analysing market microstructure and developing effective trading strategies. Figure 1.1 shows a typical schematic representation of the limit order book.

1.2 Optimal Placement Problem

Common tasks that banks, flash-traders, and other market participants often face are adjusting their position, risk, or exposure. This is done by liquidating or acquiring a (large) number of shares within a certain period of time. After the decision has been made to execute a child order at certain moment, within a specific time frame and the problem comes down to where in the LOB and when in the time frame the trader must place her orders. This problem is not studied as thoroughly in the literature, and we choose to call this the **Optimal Placement Problem** (OPP) (as in [23]), which will be the focus of this thesis.

In principle, a trader favours acquiring/liquidating their target quantity using limit-orders, because they avoid crossing the bid-ask spread and often earn a rebate for providing liquidity to the market. The structure of this rebate depends on the trading venue: it could be a fixed amount per share or a percentage of the transaction value. However, placing only limit orders, or limit orders too deep into the book, carries the risk that they may not be filled in time. This could force the trader to execute market orders at the end of the time period to fill the remaining inventory, potentially at unfavourable prices. On the other hand, if the trader chooses to execute only market orders, she risks "walking up the order book", meaning she will have to pay higher prices as they exhaust the liquidity at each price-level. Due to the fact that these market orders are in essence "liquidity-taking", they are often penalised with a fee. Another factor to consider is the price impact of placing orders on a certain side of the order book, as placing buy market or limit orders generally shifts the price of the assets upward [12]. Almgren and Chriss [2001] [1] estimate that a 10 basis point improvement in execution price translates to millions of dollars saved annually for large institutions.

Mathematically, the problem boils down to determining how many market orders to place, how many limit orders and at what depth to place these limit orders. The trader aims to minimise the expected total cost, which includes the cost of execution (the price paid or received for filled orders) and the cost of non-execution (the penalty for not filling orders by in the set time frame). This can be modelled in many different ways, as we will see in the Literature section. The solution to this problem often involves trade-offs between immediacy (using market orders) and price improvement (using limit orders), while considering the risk of under-filling the target amount of shares.

In practice, many traders, especially retail investors, might only use market orders for simplicity, accepting the potential for higher execution costs. One common strategy is the Time-Weighted Average Price (TWAP), where orders are split evenly over time to reduce market impact. Another is the Volume-Weighted Average Price (VWAP), which aims to match the volume profile of the market over the trading period. More advanced strategies involve optimising over targeting schedules, where the trader specifies a desired execution profile and adjusts the order placement accordingly [1][8].

In this thesis, the main question we try to answer is:

> *What is the optimal allocation of orders between market orders and limit orders at various price levels in a limit order book, to minimize trading costs while ensuring timely execution?*

We will build on existing work by Cont and Kukanov [2017] [15], first formulating the static version of the problem, where the trader has to place all her orders at time $t = 0$ and wait until the terminal time to place any market orders to fill the under-filled target amount. We extend on the paper by [15] generalising it to consider multiple price levels deeper in the LOB. Next we formulate the OPP as a dynamic optimal stopping and control problem, as in [5], where a trader optimises the depth at which to place her limit orders at every time between the initial and terminal time (control) and when to place market orders instead (stopping). We extend on this paper by considering possible drifts in the mid-price, introducing fees and rebates for filled market and limit orders respectively, and a general remaining inventory and running inventory penalty function.

These two models complement each other nicely. The static model creates an intuitive sense of the dynamics of the whole queueing system underlying the LOB and how an optimal order allocation is influenced by different LOB events, even for trading large inventories across longer time windows. The dynamic model then focusses our perspective on finding a strategy that is precisely optimal for smaller inventories, re-assessing our strategy at every time step and adapting to changing LOB conditions.

1.3 Literature

In the literature, many different ways to model the LOB have been proposed, each with their own (dis)advantages. These models allow us to describe the dynamics of the LOB, providing solutions to many different problems. Some important problems are finding the filling-probability of limit orders, predicting movements in the mid-price and the OPP, which we will be looking at. The mathematical model for the LOB that is chosen is crucial in determining the approach to solving these problems.

1.3.1 Models for Limit Order Book Dynamics

Accurately modelling and simulating limit order books is very important for both researches and traders. Having an accurate simulation of an LOB not only allows for prediction of future LOB events, but also for the back-testing of algorithmic trading strategies and a richer set of training data to run the strategy on.

When modelling a limit order book several points of focus can be taken. Taking the mid-price as a central focus point, modelled as a (correlated) random walk [19][23] or a mean reversion process [24], is known as the centred order book representation [28][14]. Changes to the spread and best price queues occur when incoming market orders deplete these queues. These market orders have been modelled as simple Poisson point or jump processes [32], simple Hawkes processes [33], but in later work also as Compound Poisson [2] and Compound Hawkes processes [28]. The arrival of market orders, the arrival of limit orders and the cancellation of limit orders completely determine the evolution of the order book in this model. This has the advantage that only these 3 processes need to be modelled accurately to simulate the dynamics of the LOB. Since the 3 processes describe actual real-world events, these types of models closely represent real LOBs and offer intuitive insights into their mechanics. For these reasons we chose to adopt the model introduced by Cont and Kukanov [2017] [15], where the queues in the LOB are depleted by incoming market orders and cancelled limit orders and the queues are formed by incoming limit orders. These events are modelled as random variables with no specified distribution, requiring minimal assumptions.

Another modelling approach focusses on the shape of the limit order book, by representing the state of the order book as (a pair of) densities, described by a system of stochastic partial differential equations (SPDEs) [16][26]. The shape of the order book could help indicate the direction of the movement of the mid-price, based on the imbalance in the bid and ask volumes [6] or the evolution of the incoming market order flow, also referred to as the Order Flow Imbalance [7][4][31]. These models capture volume distribution and features like convexity, depth resilience, and dynamic liquidity, and are useful when solving problems where mid-price movement play an essential role. The drawbacks of these models is that they can are computationally complex and they require some strong assumptions and simplifications, making them less representative of the real mechanics behind the LOB.

Lastly placed orders could be the focus of the model, treating the filling of an order as a birth-death process. Survival analysis is particularly useful in this case, as it is able to handle variables that are partially censored or truncated, like cancelled limit orders. The hazard function, which describes the probability that an order remains unfilled until a certain time, can then be derived analytically using underlying assumptions (as in [32]), but is more commonly derived using deep-learning methods [37][38][10]. Although these types of models often require strong assumptions, they allow for analytical solutions for problems where the fill probability of a limit order plays a central role. For this reason we choose to adopt the model introduced by Cartea and Jaimungal [2015] [5] as our second model, to find a dynamic trading strategy. It assumes limit orders can be placed at a continuous depth in the LOB and have an exponential probability distribution, that describes the probability that a limit order is filled when a market order arrives.

Recently, machine-learning methods have become very popular, not just in the case of survival analysis. In [25], they use machine-learning to forecast mid-price movements and [31] use neural networks to model the order book shape, but these methods require large amounts of training data and provide relatively little tractable insights in the inner workings of the limit order book.

1.3.2 Models for the Optimal Placement Problem

The Optimal Placement Problem (OPP) addresses the question of how a trader can optimally allocate an order across different types (limit or market) and price levels in a limit order book (LOB) to minimize expected execution cost or maximize execution probability. In this subsection we will go through the different models for the OPP used in the literature and the different scopes considered, before motivating our choice of model.

Firstly, we look at static models. In Markov chain models (as in [11]), the dynamics of the LOB are approximated as a discrete-time or continuous-time Markov process, where states may represent queue sizes, price levels, or order book imbalances. The OPP becomes one of determining the probability that a limit order is executed before the mid-price moves unfavourably, along with the expected execution price. Guo et al. [2013] [24] find the static optimal placement strategy by minimising a cost function, where the fill probability of limit orders at a certain depth is determined by the movement of the mid-price, modelled as a random walk. Cont and Kukanov [2017] [15] also formulate the OPP as a convex optimization problem, modelling the expected cost of execution as a function of the allocation of an order across MOs and limit orders at different price levels. Their static framework assumes a known distribution of LO and MO arrival and LO cancellation rates and fill probabilities are determined by these three events. Both papers find a static solution, where the trader may only place her orders at the initial time and execute any unfilled orders at the terminal time as MOs. The main strength of the optimal strategy found in [15] lies in its ability to capture the true workings of a limit order book and provide tractable solutions. However, it lacks the flexibility to adapt in real-time to market changes, as the optimal strategy is computed once and not updated dynamically.

Other researches model the OPP in a dynamic setting resulting in optimal strategies that can be adapted at intermediate times. Cartea, Jaimungal, and Ricci [2014] [9] formulate an optimal order placement strategy by modelling the LOB using a system of SDEs and use dynamic programming to derive optimal policies. Aydogan and Ugur [2023] [3] extend this to multi-period settings and include stochastic price impact and inventory risk. Although these methods are accurate at capturing intricate details of the LOB, their complexity makes analytical solutions unattainable and their dynamic programming equation often has to be solved numerically. Cartea and Jaimungal [2015] [5] formulate a dynamic optimal stopping and control problem where the trader optimally decides, at each time step, whether to place a market or a limit order (and at what depth). Assuming an exponential distribution for the fill probability of limit orders allows for the derivation of a closed form analytical solution. This offers an intuitive explicit adaptive trading strategy that allows the trader to re-optimise at every time point, leading to precisely optimal strategies.

Now, some models restrict their scope to only limit orders, typically at the best bid or ask level. For instance, Gueant et al. [2012] [21] and Cartea and Jaimungal [2015] [8] consider only passive limit order execution and focus on the fill probability of these limit orders. [20] extends the model from [8] to also consider mid-price drift, but considering only limit orders can result in trivial optimal strategies, especially for a trader looking to acquire/liquidate a large inventory or when mid-price drift is unfavourable.

Other researches, on the other hand, allow for both market orders and limit orders, but the focus is limited to choosing between a market order (MO) or a limit order (LO) placed at the best bid or ask price, often without considering the possibility of placing orders deeper into the LOB (e.g., Guo et al. [2023] [25] and Guilbaud and Pham [2011] [22]). Figueroa-Lopez et al. [2017] [19] allow for market and limit orders at any depth, but only consider the optimal placement of 1 order, rather an inventory of multiple orders. Considering market orders and limit orders at all depths allows for a truely optimal strategy even in a broader spectrum of cases (like high order-flow, significant mid-price drift or short time frames).

Recently, reinforcement learning (RL) has gained popularity for solving the OPP without requiring full knowledge of the underlying LOB dynamics. Karpe et al. [2020] [29], Fang et al. [2022] [17], and Schnaubelt [2021] [35] train deep RL agents to learn optimal order placement policies from historical LOB data. These models can uncover non-intuitive strategies and perform well in complex, high-dimensional environments. However, they require substantial amounts of training data and may lack interpretability. Additionally, RL policies are difficult to analyse formally, which can be a drawback for risk-sensitive trading environments.

The most general formulation of the OPP considers the full LOB, allowing placement of MOs and LOs across all levels and inventory sizes. We first choose to adopt the mode by Cont and Kukanov [2017] [15], as it very closely represents the mechanics that govern a real LOB, and allows for an analytical solution to the OPP for any inventory size considering market orders and limit orders (at all price-levels). Since deviations in the expected order flow can significantly throw off the static optimal allocation, we then look for a dynamic strategy that allows the trader to re-evaluate and adjust her strategy, while still considering the most general OPP formulation. Cartea and Jaimungal [2015] [5] propose frameworks that allow for (extensions to) such generality. Furthermore, these models allow for analytical solutions to this general formulation of the OPP, giving tractable insights in the workings of optimal placement strategies.

1.4 Main Contributions

This thesis builds on two models that provide analytical solutions to the OPP. Firstly, we will build on existing work by Cont and Kukanov [2017] [15], extending the paper by considering multiple price-levels deeper in the LOB, rather than just the decision between market orders and limit orders at the best price level. We derive a general expression for the solution to the n price-level case, without any assumptions on the probability distribution of market order arrivals, limit order arrivals and limit order cancellations. Furthermore, we assume these three events are exponentially distributed. This allows us to derive the analytical expression for cumulative distribution function (CDF) for the convolution of the order outflows at the different price-levels. The expression of this CDF allows us to represent the solution of the OPP explicitly, allowing us to exactly determine the optimal number of market orders and limit orders and at what price-level to place the latter.

Secondly, we formulate the OPP as a dynamic optimal stopping and control problem, as in [5]. In this model a trader optimises the depth at which to place her limit orders at every time between the initial and terminal time (control) and when to place market orders instead (stopping). We extend on this paper by formulating the optimal solution for general inventory size, rather than only providing analytical solutions for inventories of size 1 or 2. We derive a recursive formula for the optimal depth at which to place limit orders and the optimal stopping time at which to place market orders. Furthermore we extend the model by Cartea and Jaimungal [2015] [5] by considering possible drifts in the mid-price, introducing fees and rebates for filled market and limit orders respectively, and a general remaining inventory and running inventory penalty function.

1.5 Thesis Structure

Chapter 2 introduces the two models of the limit order book that are used in this thesis. Chapter 3 provides the mathematical preliminaries required to understand the derivations of the solutions. Next, in chapter 4, we will start with the static model and derive its general solution for n price-levels. In chapter 5 we move on to the dynamic model and derive its general solution for all inventory sizes. Chapter 6 shows the numerical results, where we analyse the solutions found and compare them to benchmark strategies. Chapter 7 concludes with the most important conclusions and discussion, followed by the proofs to our propositions in appendix A. Appendix B contains long derivations and other work less important to the main body of this thesis and my code for making the plots can be found in appendix C.

2 | Model of the Limit Order Book

In section 1.1 of the Introduction we explain how the limit order book works and section 1.3 shows different ways the LOB has been modelled in the literature and our motivation for the chosen models. In this chapter we will explain the chosen models in further detail, before moving on to solving the Optimal Placement Problem in both models in chapter 4 and 5.

This thesis uses 2 different models for the limit order book. Section 2.1 describes the static model from the paper by Cont and Kukanov [2017] [15] and section 2.2 describes the dynamic model from the paper by Cartea and Jaimungal [2015] [5]. In the first model in chapter 4 the dynamics of the limit order books are determined by the number of arriving market orders, arriving limit orders and cancelled limit orders. Each is modelled as a stochastic random variable, which allows us to calculate the expected cost of any order allocation analytically. The OPP then becomes a "static" convex optimization problem, where we minimise the expected cost function. In the second model in chapter 5 we directly model the filling-probability of a limit order placed at a certain depth, conditional on the arrival of a market order. The arrival of market orders is subsequently modelled as a Poisson process, to be able to calculate the expected number of filled limit orders a different depths. Using this model the optimal allocation problem considering both limit and market orders becomes a stochastic optimal stopping and control problem. The value function of the continuation and stopping scenario is optimised and compared at every time, to find the "dynamic" optimal allocation strategy.

2.1 Static Model

The first, static model was introduced by Cont and Kukanov [2017] [15] and is centred around the mid-price. Half of the spread, h, below the mid-price is the best buy quote, and half of the spread above the mid-price is the best sell quote. We look at a case of the Optimal Placement Problem, where a trader wishes to buy S shares, but if the trader was selling the problem would be completely analogous. One "tick size", δ, below the best buy quote (the first price-level) is the second best buy quote (the second price-level) and another tick size below that the third etc.. At every-price level there is a queue of outstanding limit orders, $Q_1, Q_2, Q_3, ...$. The traders' order placement problem consist of deciding how many market orders M to place and how many limit orders L_i to place at each price level $i = 1, ..., n$. This order placement decision can be summarized by a vector $\boldsymbol{X} = (M, L_1, ..., L_n) \in \mathbb{R}_+^{n+1}$ of order sizes.

The evolution of the limit order book is then determined by three processes: the arrival of market orders, the arrival of limit orders and the cancellation of limit orders. The number of arriving market orders in the time frame $t \in (0, T)$ is described by the stochastic process $\mu \in (0, \infty)$. The cancelled limit orders from each queue $Q_1, Q_2, Q_3, ...$ is described by the stochastic process $\gamma_1, \gamma_2, \gamma_3, ...$, where $\gamma_i \in (0, Q_i)$. The arriving limit orders at each price-level after time $t = 0$ is also a stochastic process denoted by $\eta_1, \eta_2, \eta_3, ...$, with $\eta_i \in (0, \infty)$. The cancelled limit orders only apply to the orders in the existing queue at $t = 0$. For the limit orders after that, η_i describes the net new number of limit orders that have arrived at time $t = T$, so not counting the ones that have been placed and cancelled inside the time frame $t \in (0, T)$.

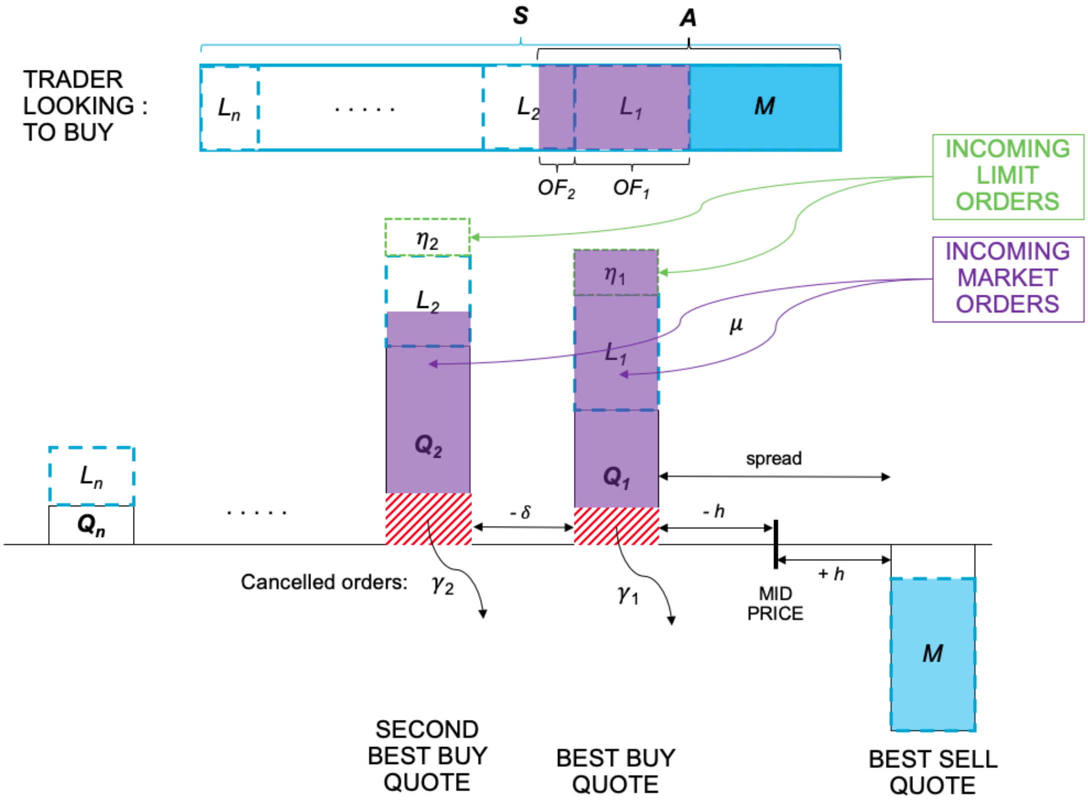

Figure 2.1: Schematic representation of the static model

To allow for analytical solutions and for ease of notation we describe the evolution of the LOB in terms of the outflows from each queue, rather than the three separate stochastic processes mentioned in the previous paragraph. The outflow at the first queue, ξ, is made up of the arriving market orders and the cancelled orders from the first queue,

$$\xi = \mu + \gamma_1 \tag{2.1}$$

with $\xi \in (0, \infty)$.

From the second queue, the outflow is obviously also determined by the number of incoming market orders, but since this is already captured in ξ, we define the (extra) outflow from the second queue, ψ_1, as the number of cancelled limit orders from that queue minus the number of limit orders that arrived before this queue (at the first price-level after $t = 0$). In general, the outflow from price-level $i + 1$ is given by

$$\psi_i = \gamma_{i+1} - \eta_i \tag{2.2}$$

with $\psi_i \in (-\infty, Q_{i+1})$. The reason we subtract η_i is because arriving limit orders are actually inflow, rather than outflow, since market orders have to deplete those first before they start to fill limit orders at price-level $i + 1$.

The Optimal Placement Problem in this model comes down to a trader, who is looking to buy S shares within the time frame $t \in (0, T)$, deciding how many market orders and limit orders to place at what price-level. In Figure 2.1 we can see how the order outflows influence the number of trader's orders that are filled at each price-level i, OF_i. To start filling L_2 orders, the incoming market order must first deplete $(Q_1 - \gamma_1), L_1, \eta_1$, and $(Q_2 - \gamma_2)$.

2.2 Dynamic Model

Where in the static model the trader had to make her placement decision at time $t = 0$ and wait until the terminal time $t = T$ to fill any remaining orders as market orders, the dynamic model allows the trader to make decisions at intermediary times, adapting to the current market conditions.

In this dynamic model of the OPP, introduced by Cartea and Jaimungal [2015] [5], the trader is looking to sell the shares in her inventory. This model combines an optimal stopping problem and an optimal control problem. At each point in time the trader faces a choice between selling a market order (stopping) or placing a limit order at a certain depth δ (controlling). This decision is based on the value function associated with either selling a market order or placing a limit order at the optimal depth. If the value of continuing to place a limit order at the optimal depth is higher than the value of stopping and placing a market order, she will place a limit order, controlling the depth at which it is placed at every time step. Once the value of stopping is equal to the continuation value, she will place a market order, as this value is guaranteed rather than uncertain like the value of limit orders. The value function depends on the trader's wealth, the size of her inventory I, the time she has left to liquidate the inventory $T - t$ and market conditions like the mid-price p and incoming market orders ν.

Stopping guarantees the trader will sell a market order. However, when a limit order is placed the probability that it is filled depends on the depth it is placed at and the arrival rate of incoming buy market orders, λ_ν. The probability that a limit order, placed at a depth δ, is filled when a buy market order arrives is assumed to be

$$P(\delta) = e^{-\kappa\delta} \tag{2.3}$$

where $\kappa > 0$. At every point in time the trader optimizes the depth at which she places her limit orders, making sure they are deep enough to sell at an attractive price, but not too deep ensuring they are filled within the desired time frame. When the trader places a market order or her limit order is filled, her inventory size decreases by 1. She then faces the same decision again, this time governed by different value functions, due to her smaller inventory size.

The trader wishes to liquidate her entire portfolio before the terminal time T, so whatever inventory she is left with at an instant before this terminal time, she will liquidate in the form of market orders.

We see that for the static model the focus was on the queue at each price-level, and the LOB dynamics were described by the three main LOB events. The dynamic model, on the other hand, focusses on the mid-price, and the dynamics are described in terms of the filling probability of a limit order at a certain depth.

3 | Mathematical Preliminaries

In this chapter we introduce preliminary mathematical concepts and theory which are used throughout this thesis.

For the static model the underlying mathematics is simpler: we minimize a cost function, use convoluted probability distributions to express our solution and we require numerical solvers to minimize (or find the root of) mathematical equations which cannot be solved algebraically in a general sense, like transcendental equations.

For the dynamic model we will require more advanced theoretical frameworks. Firstly we express the Optimal Placement Problem as a dynamic programming problem that combines stochastic control theory and an optimal stopping problem. To solve it we must apply Ito's Lemma to arrive at a Hamilton-Jacobi-Bellman equation which can then be solved iteratively as an Ordinary Differential Equation.

3.1 Numerical Methods for Root Finding

Throughout this thesis we encounter moments where we want to find the root of certain expressions, for the static model to find the optimal order allocation in equation (4.47) and for the dynamic model to find the optimal stopping time in equation (5.52) for example. In this thesis our aim is to derive an explicit formula for the optimal allocation of orders, but some expression are too complex (or even impossible) to solve algebraically. In that case, we need numerical minimizers or root finders to find our solution.

In this work, we encounter transcendental equations of the following form several times:

$$e^{Ax} + e^{Bx} + c = 0$$

where A, B, and c are constants derived from the problem's constraints, and x represents the variable to be solved, like the position, time, or parameter optimizing the placement. These equations cannot be solved in closed-form, in terms of elementary functions, due to the different exponential terms. This section explores both analytical and numerical approaches to solve such equations, evaluates their applicability, and justifies the adoption of the Newton-Raphson method as the most suitable technique for our purposes.

3.1.1 Analytical Methods

Analytical solutions, which yield exact expressions for x, are desirable but challenging for transcendental equations. We consider a few special cases where such solutions are feasible:

- **Case 1:** $A = B$
  If $A = B$, the equation simplifies to $2e^{Ax} + c = 0$, yielding:

$$e^{Ax} = -\frac{c}{2} \quad \Rightarrow \quad x = \frac{1}{A}\ln\left(-\frac{c}{2}\right)$$

This solution requires $c < 0$ to ensure $-\frac{c}{2} > 0$, making the logarithm real-valued. While elegant, this case is unlikely to hold universally in the optimal placement context.

- **Case 2:** $B = kA$
  If $B = 2A$, let $u = e^{Ax}$, transforming the equation into $u + u^2 + c = 0$. Solving this quadratic equation for u gives:
  $$u = \frac{-1 \pm \sqrt{1-4c}}{2}$$
  Taking the positive root (since $u > 0$) and back-substituting, $x = \frac{1}{A}\ln(u)$. This approach depends on a specific relationship between A and B, which may not apply generally.

In the absence of such relationships, an analytical solution is not possible. Techniques like the Lambert W function, effective for single-exponential equations, do not readily extend to sums of exponentials without complex transformations. Thus, to solve equations of the general form $e^{Ax} + e^{Bx} + c = 0$, we turn to numerical methods.

3.1.2 Numerical Methods

Numerical methods approximate the root of a function $f(x)$ through iterative refinement. We evaluate three common techniques:

1. **Bisection Method**
   This method tries to find a root within an interval $[a, b]$ where $f(a)f(b) < 0$, iteratively halving the interval at mid-point $c = \frac{a+b}{2}$. At each time step it checks in which half the root lies, by checking if $f(a)f(c) < 0$ or $f(b)f(c) < 0$. It then choose the half containing the root and halves this interval again, continuing until $|f(c)|$ is sufficiently close to the root.

   This method is robust but converges slowly (linear convergence), making it less efficient for high-precision needs. Furthermore, it requires that on its defined interval $[a, b]$, the condition $f(a)f(b) < 0$ is satisfied. However, this is not necessarily the case in the situations where we wish to apply the root-finder, making this method unsuitable for our application.

2. **Newton-Raphson Method**
   The Newton-Raphson method starts with an initial guess x_0 and iteratively refines it using the formula:
   $$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$
   where $f'(x)$ is the derivative of the function. This method uses the tangent line at x_n to approximate the root, leading to quadratic convergence when the initial guess is close to the true root. For our transcendental equation $f(x) = e^{Ax} + e^{Bx} + c$, the derivative is $f'(x) = Ae^{Ax} + Be^{Bx}$, which is straightforward to compute.
   The method's efficiency depends on a good initial guess and a non-zero derivative at each iteration. Poor initial guesses or points where $f'(x_n) \approx 0$ can lead to divergence or slow convergence. Despite these challenges, its speed makes it highly suitable for problems requiring repeated root-finding in our case.

3. **Secant Method**
   The Secant method approximates the root using two initial guesses, x_0 and x_1, and iterates via:
   $$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

This approach approximates the derivative using the slope of the secant line between two points, eliminating the need for explicit derivative computation. While this makes it simpler to implement for functions where the derivative is complex, its convergence is super-linear, slower than Newton-Raphson's quadratic rate. Additionally, it requires two initial guesses, and its performance can degrade if the guesses are poorly chosen or if the function behaves non-smoothly near the root.

We select the Newton-Raphson method for solving equations of the form $e^{Ax} + e^{Bx} + c = 0$ due to its quadratic convergence, ensuring fast computation. The method's reliance on the derivative $f'(x) = Ae^{Ax} + Be^{Bx}$ is not a limitation, as it is easily computed. Additionally, the real-world context of our variables allows informed initial guesses, enhancing convergence reliability. We also implement a tolerance of $\epsilon = 10^{-6}$ (stopping when $|f(x_n)| < \epsilon$) and may employ the Bisection method as a fallback for poor initial guesses.

3.2 Dynamic Programming

3.2.1 Dynamic Programming Principle for Optimal Stopping and Control Problems

In a combined optimal stopping and control problem, we are solving a problem where we try to optimise our value function using the control variables and at the same time decide when it is optimal to decide to "stop". First, let $\boldsymbol{X}^u \in \mathbb{R}^m$ denote the vector containing the controlled processes that determine the trader's performance criteria. $\boldsymbol{X}^u$ is controlled using the control process $u = (u_t)_{0\le t\le T}$, and satisfies the SDE

$$d\boldsymbol{X}_t^u = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t + \boldsymbol{\gamma}_t^u d\boldsymbol{N}_t^u \tag{3.1}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_t)_{0\le t\le T} \in \mathbb{R}^m$ is the drift, $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_t)_{0\le t\le T} \in \mathbb{R}_+^m$ is the volatility and $\boldsymbol{W} = (\boldsymbol{W}_t)_{0\le t\le T} \in \mathbb{R}^m$ a collection of independent Brownian motions. Furthermore, $\boldsymbol{N}^u = (\boldsymbol{N}_t^u)_{0\le t\le T} \in \mathbb{N}^n$ denotes a collection of counting processes with controlled intensities $\boldsymbol{\lambda}^u = (\boldsymbol{\lambda}_t^u)_{0\le t\le T} \in \mathbb{R}^n$, and $\boldsymbol{\gamma}_t^u := \boldsymbol{\gamma}(t, u_t) \in \mathbb{R}^{n\times m}$ denotes the controlled jump size.

The trader's reward is a function of these controlled processes $\boldsymbol{X}^u$, which are stochastic process, which is why we need to take the expectation. The trader has a performance criterion, for each admissible control process $u \in \mathcal{A}_{[t,T]}$ and admissible stopping time $\tau \in \mathcal{T}_{[t,T]}$, given by

$$H^{\tau,u}(t, \boldsymbol{x}) = \mathbb{E}_{t,\boldsymbol{x}}[G(\boldsymbol{X}_\tau^u) + \int_t^\tau F(s, \boldsymbol{X}_s^u)ds] \tag{3.2}$$

where $G(\boldsymbol{X}_\tau^u)$ is the reward upon exercise, $F(s, \boldsymbol{X}_s^u)$ is the running penalty, and she seeks to find the value function

$$H(t, \boldsymbol{x}) = \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} H^{\tau,u}(t, \boldsymbol{x}) = \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}[G(\boldsymbol{X}_\tau^u) + \int_t^\tau F(s, \boldsymbol{X}_s^u)ds] \tag{3.3}$$

and control strategy u and stopping time τ which attain the supremum of the performance criterion (if it exists).

Say we are at a point $(t = t_0, \boldsymbol{X} = \boldsymbol{x})$ and $G(\boldsymbol{X}_\tau^u)$ is the reward paid out at the chosen stopping time (and optimal strategy u). Our value function is described by the optimal stopping time $\tau^\star \in [t_0, T]$ that maximizes the expected reward, given the point $(t_0, \boldsymbol{x})$ which we are currently at. Now that we know which stopping time maximizes our expected reward, the value function $H(t, x)$ with $t_0 \le t < \tau^\star$ describes the current value of our position.

The dynamic programming principle (DPP) for optimal stopping and control problems, as described above and in [8], is given by the following theorem.

Theorem 3.1. ***Dynamic Programming Principle for Optimal Stopping and Control.*** *The value function $H(t, \boldsymbol{x})$ satisfies the DPP*

$$H(t, \boldsymbol{x}) = \sup_{\tau \in \mathcal{T}_{[t,T]}} \sup_{u \in \mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}\Big[\Big(H(\theta, \boldsymbol{X}_\theta^u) + \int_t^\theta F(s, \boldsymbol{X}_s^u) ds\Big) 1_{\tau \geq \theta} + \Big(G(\boldsymbol{X}_\tau^u) + \int_t^\tau F(s, \boldsymbol{X}_s^u) ds\Big) 1_{\tau < \theta}\Big] \tag{3.4}$$

for all $(t, \boldsymbol{x}) \in [0, T] \times \mathbb{R}^m$ and all stopping times $\theta \leq T$, where $G(X_\tau^u)$ is the reward function and $F(s, \boldsymbol{X}_s^u)$ is the running penalty.

The proof to this theorem can be found in section A.8 of the Proofs, and the intuition behind this DPP is as follows. If stopping has not yet occurred ($\tau \geq \theta$), then the trader receives the value function at that time. If stopping has already occurred ($\tau < \theta$), then the trader has received the reward at this stopping time.

The Dynamic Programming Equation (DPE) satisfied by the value function can be obtained by looking at the DPP over infinitesimally small amounts of time. The DPE that arises in this problem, as in [8], is stated by the following theorem.

Theorem 3.2. ***Dynamic Programming Equation for Stopping and Control Problems.*** *Assume that the value function $H(t, \boldsymbol{x})$ is once differentiable in t and all seconder-order derivatives in $\boldsymbol{x}$ exist, i.e. $H \in C^{1,2}([0, T], \mathbb{R}^m)$ and that $G : \mathbb{R}^m \to \mathbb{R}$ and $F : \mathbb{R}^m \to \mathbb{R}$ are continuous. Then H solves the* ***quasi-variational inequality (QVI)****,*

$$\max\Big(\partial_t H + \sup_{u \in \mathcal{A}_t} \big(\mathcal{L}_t^u H + F\big) \ ; \ G - H\Big) = 0, \quad on \ \mathcal{D}, \tag{3.5}$$

where $\mathcal{D} = [0, T] \times \mathbb{R}^m$ and the generator of the process $\mathcal{L}_t$ acts on twice differentiable functions as follows:

$$\mathcal{L}_t h(t, x) = \boldsymbol{\mu}(t, \boldsymbol{x}) \cdot \boldsymbol{D}_x h(t, x) + \frac{1}{2} \boldsymbol{Tr}(\boldsymbol{\sigma}(t, \boldsymbol{x})\boldsymbol{\sigma}(t, \boldsymbol{x})' \boldsymbol{D}_x^2 h(t, \boldsymbol{x})) + \sum_{j=1}^p \lambda_j(t, \boldsymbol{x})[h(t, \boldsymbol{x} + \boldsymbol{\gamma}_{\cdot j}(t, \boldsymbol{x})) - h(t, \boldsymbol{x})] \tag{3.6}$$

where $\boldsymbol{\gamma}_{\cdot j}$ represents the j-th column of $\boldsymbol{\gamma}$, $\boldsymbol{D}_x h$ represents the vector of partial derivatives w.r.t. x, and $\boldsymbol{D}_x^2 h$ represents the matrix of second-order partial derivatives w.r.t. x.

The proof to this theorem can be found in section A.9 of the Proofs. The intuition behind this QVI can best be explained when handling the two possibilities separately, we either have

$$\text{(i)} \quad \partial_t H + \mathcal{L}_t H + F = 0 \quad \text{and} \quad G < H, \tag{3.7}$$

or we have

$$\text{(ii)} \quad H = G \quad \text{and} \quad \partial_t H + \mathcal{L}_t H + F < 0. \tag{3.8}$$

In the first case, the reward G is lower than the value function H, which means that we are in the continuation region (i.e. it is optimal for the trader to continue rather than choosing the reward) and the value function satisfies a linear PDE. In the second case, the value function H has the same value as the reward G and the PDE tells us the value function will decrease in time. This means that it is optimal for the trader to choose the instant, guaranteed reward rather than continuing.

4 | The Static Optimal Placement Problem

The first model to tackle the Optimal Placement Problem in this thesis is based on the model introduced by Cont and Kukanov [2017] [15]. They choose to analyse strategies where a trader can place her limit orders at the best buy quote on multiple different exchanges. In this thesis, however, we choose to only consider one exchange, but explore the situation where a trader can place her limit orders at different price levels. We start by introducing the simplest problem, where the trader can only post at the best buy quote and then consider two, three and four price-levels, to eventually arrive at a general case where the trader can choose any price-level to post her limit orders. We consider the case where the trader wishes to acquire shares within a certain time frame, but the case where she wishes to liquidate an inventory by selling shares is completely analogous.

Below is a brief overview of all symbols and notations used in this chapter.

- S is the target number of shares the trader wishes to acquire
- T is the terminal time at which the trader must have acquired her target number of shares S
- M is the number of market orders she places at $t = 0$
- L_i is the number of limit orders she places at price-level i at $t = 0$
- $\boldsymbol{X}$ is the placement vector containing the number of market orders and limit orders (at each price-level) the trader places
- Q_i is the queue size at price-level i at $t = 0$
- γ_i is the number of cancelled limit orders from the queue Q_i within the time frame $t \in (0, T)$
- μ is the number of incoming number of market sell orders within the time frame $t \in (0, T)$
- η_i is the number of incoming limit buy orders at price-level i within the time frame $t \in (0, T)$
- ξ is the order outflow at the first price-level, and consists of incoming market sell orders μ and cancelled limit orders from the first price-level γ_1
- ψ_i is the order outflow at price-level $i + 1$, and consists of incoming limit buy order at price-level i, η_i, and cancelled limit orders from price-level $i + 1$, γ_{i+1}
- OF_i is the number the trader's limit orders placed at price-level i that are filled within the time frame $t \in (0, T)$
- A is total number of trader's limit orders that are filled within the time frame $t \in (0, T)$
- δ is the tick size between the price-levels

- h is half of the bid-ask spread
- f is a fee the trader must pay for placing market orders
- r is a rebate the trader receives for filling her limit orders
- ρ_u is the penalty coefficient when the trader under-fills her target amount of shares S
- ρ_o is the penalty coefficient when the trader over-fills her target amount of shares S
- θ is the price impact coefficient for any order the trader places at $t = 0$ or at terminal time $t = T$

At time $t = 0$, the trader, whose objective is to buy S shares before time $t = T$, may submit limit orders (LOs) of total size $L_1 \geq 0$ at the best buy quote (also called the first price-level) and market orders (MOs) for $M \geq 0$ shares. The order placement decision is denoted by the vector $\boldsymbol{X} = (M, L_1) \in \mathbb{R}^2_+$ of order sizes. The LOs L_1 join the best buy quote queue of size $Q_1 \geq 0$. The number of the trader's limit orders which are filled between time $t = 0$ and $t = T$ is a function of the arriving market orders, denoted by μ, and the cancelled limit orders, denoted by γ_1. For the cancelled limit orders we only consider limit orders in the best buy quote queue which were placed before the initial time $t = 0$. The number of orders at the best bid price that gets filled by time T, denoted as OF_1, is given by:

$$\mathrm{OF}_1 = \min(\max(\mu + \gamma_1 - Q_1, 0), L_1) = (\mu + \gamma_1 - Q_1)_+ - (\mu + \gamma_1 - Q_1 - L_1)_+ \tag{4.1}$$

where we use the notation $(x)_+ = max(x, 0)$. The total amount of shares bought by the trader during $[0, T)$, denoted by $A(X, \mu, \gamma_1)$, is given by

$$A(\boldsymbol{X}, \mu, \gamma_1) = M + \mathrm{OF}_1. \tag{4.2}$$

and consists of the amount of market orders placed plus the number of limit orders filled at terminal time T. We can generalise this model to consider all the possible price levels at which the trader can post limit orders. The price-levels start at the best buy quote (the first price-level) and then go down in tick sizes δ. The trader now has to decide how many MOs to place and how many LOs to place at every price-level. We choose to consider n price-levels and denote the order placement decision by the vector $\boldsymbol{X} = (M, L_1, L_2, ..., L_n)$. A limit buy order placed at the i-th price-level is placed at a price $h + (i-1)\delta$ below the mid-price, where h is half of the bid-ask spread. Every price-level i has its own queue Q_i in which the trader places her limit orders. We introduce a new stochastic variable ψ_{i-1}, which describes the inflow and outflow of orders that have priority after the trader's LOs placed at price-level $i-1$ and before LOs placed at price-level i. This order flow consists of LOs arriving at price-level $i-1$, in the time window $t \in (0, T)$, and LOs being cancelled from the queue Q_i. For the arriving LOs at price-level $i-1$, denoted by η_{i-1}, we don't consider orders that are placed and cancelled in the time window $t \in (0, T)$, and we denote the order cancelled from Q_i as γ_i. Because η_{i-1} represents incoming orders (inflow) and γ_i cancelled orders (outflow), we define the stochastic variable for the outflow at price-level i as

$$\psi_{i-1} = \gamma_i - \eta_{i-1}. \tag{4.3}$$

To ease notation and derivation, we also choose to define the order outflow at the first price-level as

$$\xi = \mu + \gamma_1, \tag{4.4}$$

consisting of the arriving MOs μ and the cancelled LOs from the first queue γ_1. We see that ψ_{i-1} is very distinct from the outflow at the first price-level ξ. These definitions mean that, when considering n price-levels, we have that $\xi \in [0, \infty)$ and $\psi_{i-1} \in (-\infty, Q_i]$, as obviously $\gamma_i \in [0, Q_1]$ and $\eta_i \in [0, \infty)$. A schematic overview of the static order placement model can be found in Figure 2.1 in Chapter 2.

The number of the trader's orders filled at price-level i (which is $i-1$ ticks below the best price-level) is given by the formula

$$\begin{aligned} \mathrm{OF}_i =& (\xi + \sum_{k=1}^{i-1} \psi_k - \sum_{k=1}^{i-1}(Q_k + L_k) - Q_i) \cdot 1_{\{\sum_{k=1}^{i-1}(Q_k+L_k)+Q_i<\xi+\sum_{k=1}^{i}\psi_k<\sum_{k=1}^{i}(Q_k+L_k)\}} \\ &+ L_i \cdot 1_{\{\xi+\sum_{k=1}^{i}\psi_k>\sum_{k=1}^{i}(Q_k+L_k)\}} \end{aligned} \tag{4.5}$$

where L_i is the size of the trader's order placed at the i-th price-level below the best quote. So, the number of the trader's orders filled at price-level i depends on the convolution of the probability functions of the order outflow at the first price-level, ξ, and the outflow at the other price levels before it, $\psi_1, ..., \psi_{i-1}$. When considering n price levels, the total number the trader's orders that are filled before terminal time $t = T$, denoted as A, is the sum of the number of her LOs that are filled at each price-level and the number of MOs placed at $t = 0$. So the total number of orders the trader fills before time T is given by

$$A(\boldsymbol{X}, \xi, \boldsymbol{\psi}) = M + \sum_{i=1}^{n} \mathrm{OF}_i \tag{4.6}$$

where $\boldsymbol{X} = (M, L_1, L_2, ..., L_n)$ is the order allocation for n price-levels, ξ is the order outflow at the first price-level and $\boldsymbol{\psi} = (\psi_1, ..., \psi_{n-1})$ are the outflows at the $n-1$ price levels below the first. For simplicity we use the notation $A = A(\boldsymbol{X}, \xi, \boldsymbol{\psi})$.

Most exchanges offer sellers of limit orders rebates, say r^e, but also set penalties for adverse selection, f_{AS}. This results in an effective rebate of $r = r^e + f_{\mathrm{AS}}$, that the trader, whose limit order is executed, receives as a reward for providing liquidity to the market. Using market orders to buy shares, which can be seen as liquidity-taking, is often penalized with an additional fee f. We can now calculate the execution cost relative to the mid-price for an order allocation $\boldsymbol{X} = (M, L_1, ..., L_n)$:

$$C(\boldsymbol{X}, \xi, \boldsymbol{\psi}) = (h + f)M - \sum_{i=1}^{n}(h + (i-1)\delta + r)\mathrm{OF}_i \tag{4.7}$$

where h is half of the spread.

A trader may also experience a shortfall due to unfilled orders, $A < S$, in which case the trader has to purchase the remaining $S - A$ shares at time T. In contrast, the trader might have bought too many shares $A > S$, and must sell the excess $A - S$ at time T. Both events introduce additional costs. Adverse selection, due to the placement of the buy limit orders at time 0, implies that conditionally on this event prices have likely increased and the cost of market orders at time T is higher than their cost at time 0, i.e. larger than $h + f$. This execution risk is to be included in the objective function and given by

$$R(\boldsymbol{X}, \xi, \boldsymbol{\psi}) = \rho_u(S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))_+ + \rho_o(A(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - S)_+ \tag{4.8}$$

where $\rho_u, \rho_o \geq 0$ are the marginal penalties (in dollars) of under- and over-filling the target number of shares S. We can see ρ_u as a trader's urgency to fill her orders.

Lastly, empirical studies have shown that both market and limit orders affect prices, and the average impact of small orders can be well approximated by a linear function [13][27]. Here buy orders (limit or market) drive the mid-price up and sell orders generally drive the price down. We assume the impact cost is paid on all orders placed at times 0 and T, irrespective of whether they are filled, with the total impact of all orders given by

$$I(\boldsymbol{X}, \mu, \gamma_1) = \theta\,(M + L_1 + (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))_+) \tag{4.9}$$

where θ is the impact coefficient.

Adding all these terms allows us to define the cost function as follows.

Definition 4.1. *[15](**Cost function**). Let* $\boldsymbol{X} = (M, L_1, ..., L_n) \in \mathbb{R}^{n+1}$ *be a placement vector,* ξ *a random variable and* $\boldsymbol{\psi}$ *a vector of* $n-1$ *random variables, then the cost function is defined as*

$$
\begin{aligned}
v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) :=& (h+f)M - \sum_{i=1}^{n} (h + r + (i-1)\delta) OF_i \\
&+ \rho_u (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))_+ + \rho_o (A(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - S)_+ \\
&+ \theta \left(M + \sum_{i=1}^{n} L_i + (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))_+ \right)
\end{aligned}
\tag{4.10}
$$

where $h, f, r, \delta, \rho_u, \rho_o, \theta, S \geq 0$ *are constants,* OF_i *is given by equation (4.5) and* $A(\boldsymbol{X}, \xi, \boldsymbol{\psi})$ *is given by equation (4.6).*

We can now formulate the *Optimal Placement Problem* considering n price-levels.

Problem 4.1. (Optimal Placement Problem) *An optimal order placement is a vector* $X^* = (M^\star, L_1^\star, ..., L_n^\star) \in \mathbb{R}_+^{n+1}$ *that minimises the expected cost function* $v(\boldsymbol{X}, \xi, \boldsymbol{\psi})$ *and therefore is the solution of*

$$
\underset{X \in \mathbb{R}_+^{n+1}}{min} \mathbb{E}[v(\boldsymbol{X}, \xi, \boldsymbol{\psi})] \tag{4.11}
$$

where $v(\boldsymbol{X}, \xi, \boldsymbol{\psi})$ *is given by equation (4.10).*

To avoid certain trivial solutions to the Optimal Placement Problem we begin by making some reasonable assumptions on the parameter values.

Suppose our trader has filled more LOs than here target quantity S, i.e. $A > S$. In that case the trader must place market sell orders at the terminal time T. Let us denote the price-level k as the lowest level in which at least 1 of the trader's placed limit orders is filled before time T, $k = \min_{k \geq 0}\{k : \text{OF}_k > 0\}$. When the trader has overfilled her target amount, it means that she has first bought shares (in the form of LOs) at a price greater or equal to $p - (h + (k-1)\delta)$ and must then sell the overfilled amount as market orders at a price $p - (h + (k-1)\delta)$ (or lower if the overfilled amount is large enough to walk the book), where p is the mid-price. This is because all the queues at the higher prices have all been depleted. For shares sold at a lower price, i.e. price-level $i > k$, this means a loss (or extra cost) of $(i-k)\delta$ per share. Therefore, we must argue that in reality it is never optimal to overfill the target amount. We use this argument and other reasonable economical restrictions to make the following assumptions.

Assumptions

A1: $\rho_u > h + f$: market orders sent at time 0 are less expensive compared to market orders that are sent at time T (avoids the trivial solution where only limit orders are placed at $t = 0$)
A2: $\rho_o + \theta > h + (k-1)\delta + r$ for $k = \min_{k>0}\{k : \text{OF}_k > 0\}$ and $\rho_o > -(h + f)$: it is suboptimal to over-fill the target size S regardless of fees and rebates
A3: $r + h > 0$: possibly negative rebates do not eliminate price improvement from limit order execution (limit orders are still cheaper than market orders)

These assumptions lead to the following proposition:

Proposition 4.1. *Under assumption A1-A3 all optimal order allocations for the Optimal Placement Problem 4.1 belong to the set*

$$
\mathcal{C} = \{(M^\star, L_1^\star, L_2^\star, ...) : \ 0 \leq M^\star \leq S, \ \ 0 \leq L_i^\star \leq S, \ \ M^\star + \sum L_i^\star = S\} \tag{4.12}
$$

The proof to this proposition can be found in section A.1. Under the constraint $S = M + \sum_{i=1}^{n} L_i$, which means

$$(S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))_+ = (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))1_{\{\xi + \sum_{i=1}^{n} \psi_i < \sum_{i=1}^{n}(Q_i + L_i)\}} \tag{4.13}$$

equation (4.10) becomes

$$\begin{aligned} v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) :=& (h+f)M - (h+r)\mathrm{OF}_1 - \sum_{i=1}^{n-1}(h + r + i \cdot \delta)\mathrm{OF}_{i+1} \\ &+ \rho_u (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))1_{\{\xi + \sum_{i=1}^{n-1} \psi_i < \sum_{i=1}^{n}(Q_i + L_i)\}} \\ &+ \theta \left(M + \sum_{i=1}^{n} L_i + (S - A(\boldsymbol{X}, \xi, \boldsymbol{\psi}))1_{\{\xi + \sum_{i=1}^{n-1} \psi_i < \sum_{i=1}^{n}(Q_i + L_i)\}} \right) \end{aligned} \tag{4.14}$$

4.1 Solution Considering One Price-Level

In this section we review the case where the trader is looking to buy S shares before terminal time T and she can only post MOs and LOs at the best price-level, as in [15]. In this one price-level case, where $n = 1$, the cost function in equation 4.10 becomes

$$\begin{aligned} v(\boldsymbol{X}, \xi) =& (h+f)M - (h+r)\mathrm{OF}_1 \\ &+ \rho_u (S - A(\boldsymbol{X}, \xi))_+ + \rho_o (A(\boldsymbol{X}, \xi) - S)_+ \\ &+ \theta \left(M + L_1 + (S - A(\boldsymbol{X}, \xi))_+ \right) \end{aligned} \tag{4.15}$$

where $\boldsymbol{X} = (M, L_1)$ is the trader's order allocation consisting of M MOs and L_1 LOs at the best price-level, ξ is the order outflow at the first price-level given by 4.4, OF_1 is the number of trader's L_1 orders that are filled and $A(\boldsymbol{X}, \xi)$ is given by 4.2. The trader must find an optimal order allocation $\boldsymbol{X}^\star = (M^\star, L_1^\star)$ to minimise the expected cost function, which we denote as $V(X) = \mathbb{E}[v(X, \xi)]$.

When considering one price-level for limit orders, adhering to the Assumptions at the beginning of this chapter, means an optimal allocation to the Optimal Placement Problem must satisfy $M^* + L^* = S$, i.e. the trader does not place more orders than her target quantity. Solving the OPP it becomes clear that ρ_u plays a crucial role, as it increases so does the amount of market orders M^*, which is logical. For the $n = 1$ case of Problem 4.1 the optimal split between limit and market orders is given by the following proposition.

Proposition 4.2. *Assume that ξ has a continuous distribution and Assumptions (A1-A3) hold. Then the following allocations are optimal for the $n = 1$ case of the Order Placement Problem 4.1*

(i) If $\rho_u \leq \underline{\rho_u} = \frac{2h+f+r}{F(Q_1+S)} - (h + r + \theta)$, then $(M^*, L_1^*) = (0, S)$ is an optimal allocation.

(ii) If $\rho_u \geq \overline{\rho_u} = \frac{2h+f+r}{F(Q_1)} - (h + r + \theta)$, then $(M^*, L_1^*) = (S, 0)$ is an optimal allocation.

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, an optimal allocation is a mix of limit and market orders, given by

$$M^* = S - F^{-1}\left(\frac{2h + f + r}{\rho_u + h + r + \theta} \right) + Q_1, \tag{4.16}$$

$$L_1^* = F^{-1}\left(\frac{2h + f + r}{\rho_u + h + r + \theta} \right) - Q_1, \tag{4.17}$$

where $F(x) = \mathbb{P}(\xi \leq x)$ is the distribution of the outflow at the first price-level $\xi = \mu + \gamma_1$ and F^{-1} is its left-inverse.

The proof to this proposition can be found in section A.2.

4.1.1 Example: Exponentially Distributed Order Flow at the First Price-Level

We will now show an example for an analytical solution, when assuming that ξ exponentially distributed. For the exponential distribution, the CDF is given by $F_\xi(x) = 1 - e^{-\lambda x}$. Then (A.4) becomes

$$M^\star = \frac{1}{\lambda}\ln\left(\frac{-h - f + \rho_u + \theta}{h + r + \rho_u + \theta}\right) + Q + S, \tag{4.18}$$

and equation (A.5) becomes

$$L^\star = \frac{1}{\lambda}\ln\left(\frac{h + r + \rho_u + \theta}{-h - f + \rho_u + \theta}\right) - Q. \tag{4.19}$$

4.2 Solution Considering Two Price-Levels

In this section we will review the case where the trader can post MOs and LOs at the best and second-best price-levels. The second-best price-level is one *tick size* δ below the best price, with a queue of size Q_2 at time 0. In the Optimal Placement Problem 4.1 considering two price-levels (i.e. $n = 2$) the trader's order allocation vector becomes $\boldsymbol{X} = (M, L_1, L_2)$, where L_1 and L_2 represent the number of LOs placed at the best and second-best price-level, where $L_1, L_2 \geq 0$. We introduce the random variable ψ_1 defined in 4.3, which describes the order outflow at the second price-level. It consists of LOs arriving at the first price-level during $t \in (0, T)$, denoted by η_1, and LOs cancelled from Q_2, denoted by γ_2. For a schematic overview of the model please refer back to Figure 2.1. The number the trader's LOs in the second-best price that are filled level is given by

$$\begin{aligned}
\mathrm{OF}_2 &= \min(\max(\xi - Q_1 - L_1 + \psi - Q_2, 0), L_2) \\
&= (\xi - Q_1 - L_1 + \psi - Q_2)_+ - (\xi - Q_1 - L_1 + \psi - Q_2 - L_2)_+ \\
&= (\xi + \psi - Q_1 - L_1 - Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+L_1+Q_2+L_2\}} + L_2 \cdot 1_{\{\xi+\psi>Q_1+L_1+Q_2+L_2\}}
\end{aligned} \tag{4.20}$$

where $\psi = \psi_1$ is the order outflow at the second price-level. Equation 4.10 in this case is given by

$$\begin{aligned}
v(\boldsymbol{X}, \xi, \psi) := {}& (h + f)M - (h + r)\mathrm{OF}_1 - (h + \delta + r)\mathrm{OF}_2 \\
&+ \rho_u(S - A(\boldsymbol{X}, \xi, \psi))_+ + \rho_o(A(\boldsymbol{X}, \xi, \psi) - S)_+ \\
&+ \theta\left(M + L_1 + L_2 + (S - A(\boldsymbol{X}, \xi, \psi))_+\right)
\end{aligned} \tag{4.21}$$

where $A(\boldsymbol{X}, \xi, \psi)$ is given by equation (4.2). Proposition 4.1 tells us that $S = M + L_1 + L_2$, so we have that $A - S = 0$ and $(S - A)_+ = (S - A)1_{\{\xi+\psi<Q_1+L_1+Q_2+L_2\}}$, which allows us to rewrite the expected cost function as

$$\begin{aligned}
\mathbb{E}[v(\boldsymbol{X}, \xi, \psi)] :&= \mathbb{E}[(h + f)M - (h + r)\mathrm{OF}_1 - (h + \delta + r)\mathrm{OF}_2 \\
&\quad + (\rho_u + \theta)(S - A(\boldsymbol{X}, \xi, \psi))1_{\{\xi+\psi<Q_1+L_1+Q_2+L_2\}} + \theta S] \\
&= \mathbb{E}[(h + f)M - (h + r)\mathrm{OF}_1 - (h + \delta + r)\mathrm{OF}_2 \\
&\quad + (\rho_u + \theta)(S - (M + \mathrm{OF}_1 + \mathrm{OF}_2))1_{\{\xi+\psi<Q_1+L_1+Q_2+L_2\}} + \theta S].
\end{aligned} \tag{4.22}$$

In the case where the trader can choose to place limit orders at the best bid or the second-best bid, we can express the solution to the Optimal Placement Problem 4.1 in the following proposition.

Proposition 4.3. *Assume that ξ and ψ have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocation for the $n = 2$ case of the Order Placement Problem 4.1 satisfies the following equations*

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+\delta+f+r}{F_{\xi+\psi}(Q_1+Q_2)} - (h + r + \delta + \theta)$, then $(M^*, L_1^*, L_2^*) = (S, 0, 0)$ is an optimal allocation.

(ii) *If* $\rho_u \leq \underline{\rho_u} = \frac{2h+\delta+f+r}{F_{\xi+\psi}(Q_1+Q_2+S)} - (h+r+\delta+\theta)$, *then* $M^* = 0$ *and* L_1^* *satisfies,*

$$\delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1^\star) - (h+\delta+r+\rho_u+\theta)F_{\xi+\psi}(Q_1+Q_2+L_1^\star) = 0, \tag{4.23}$$

and L_2^* *satisfies*

$$L_2^\star = S - M^\star - L_1^\star. \tag{4.24}$$

(iii) *If* $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, *then the optimal allocation is a mix of limit* (L_1^*, L_2^*) *and market* $(M^\star)$ *orders, where* $L_1^\star$ *satisfies (4.23),* $L_2^\star$ *satisfies (4.24), and* $M^\star$ *satisfies*

$$M^\star = S - F_{\xi+\psi}^{-1}\left(\frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta}\right) + Q_1 + Q_2, \tag{4.25}$$

where $F_\xi(x) = \mathbb{P}(\xi \leq x)$ *is the distribution of the outflow at the first price-level* $\xi = \mu + \gamma_1$, $F_{\xi+\psi}(x) = \mathbb{P}(\xi + \psi \leq x)$ *is the distribution of the outflow at the first and second price-level* ξ, $\psi = -\eta_1 + \gamma_2$ *and* $F_\xi^{-1}, F_{\xi+\psi}^{-1}$ *are their left-inverses.*

The derivation can be found in section A.3 and methods for finding the root of (4.23) are discussed in section 3.1.

4.2.1 Special Case: Empty Queue and No New LOs

We see that in a special case where the second price-level queue is empty and no new LOs arrive at the first price-level (i.e. $Q_2 = 0$ and $\psi = 0$), we are left with the one price-level case, where the order allocation only considers M and L_2, since equation (4.23) becomes

$$\begin{aligned} \frac{\partial V}{\partial L_1} = \delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1)) - (h+\delta+r+\rho_u+\theta)F_\xi(Q_1+L_1) = 0 \\ \delta - \delta F_\xi(Q_1+L_1) = 0 \end{aligned} \tag{4.26}$$

where $F_\xi(Q_1+L_1) \in (0,1)$ for $L_1 \in [0,S]$. We see that if Q_2 and ψ_2 are set to zero, the derivative of V w.r.t L_1 is always increasing, which is logical, since in that case L_1 orders are just as likely to be filled as L_2 orders, so it is never optimal to place orders at the price-level of L_1. This means we set $L_1 = 0$, which means

$$L_2^\star = S - M^\star - L_1^\star = S - M^\star = F_{\xi+\psi}^{-1}\left(\frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta}\right) - Q_1 \tag{4.27}$$

as is the case for the limit orders in the one price-level case (but this time the traders earns $(h+\delta+r)$ (instead of $(h+r)$) for a limit order).

4.2.2 Special Case: Many New LOs at the First Price-Level

When we set $\psi \to -\infty$, this means there are many limit orders arriving at the first price-level, behind our L_1 orders, making it impossible for our L_2 limit orders to be filled. We see that when $\psi \to -\infty$, we have $F_{\xi+\psi}(Q_1+Q_2+L_1^\star) \to 1$. In this case equation (4.23) becomes

$$\begin{aligned} \frac{\partial V}{\partial M} &= 2h+\delta+f+r - (h+\delta+r+\rho_u+\theta) \\ &= h + f - (\rho_u+\theta) \leq 0 \end{aligned} \tag{4.28}$$

where the inequality is due to Assumption **A2**, and (4.25) becomes

$$\begin{aligned} \frac{\partial V}{\partial L_1} &= \delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1^\star) - (h+\delta+r+\rho_u+\theta) \\ &= (h+r+\rho_u+\theta)(F_\xi(Q_1+L_1^\star) - 1) \leq 0. \end{aligned} \tag{4.29}$$

We see that in this case the expected cost function is decreasing in both M and L_1. To know what values of M and L_1 actually minimise V, we need to know which is smaller: $\frac{\partial V}{\partial M}$ or $\frac{\partial V}{\partial L_1}$. We find that we have $\frac{\partial V}{\partial M} < \frac{\partial V}{\partial L_1}$ when

$$\begin{aligned} h + f - (\rho_u + \theta) &< (h + r + \rho_u + \theta)(F_\xi(Q_1 + L_1^\star) - 1), \\ F_\xi(Q_1 + L_1^\star) - 1 &> \frac{h + f - (\rho_u + \theta)}{h + r + \rho_u + \theta}, \\ F_\xi(Q_1 + L_1^\star) &> \frac{2h + \delta + f + r}{h + r + \rho_u + \theta}, \\ L_1^\star &> F_\xi^{-1}\left(\frac{2h + \delta + f + r}{h + r + \rho_u + \theta}\right) - Q_1. \end{aligned} \tag{4.30}$$

Let us denote the solution to one price-level OPP (as described in Proposition 4.2) as $\boldsymbol{X}^{\star,\text{ 1 price-level}} = (M^{\star,\text{ 1 price-level}}, L_1^{\star,\text{ 1 price-level}})$ and the solution to the one price-level OPP as simply $\boldsymbol{X}^\star = (M^\star, L_1^\star, L_2^\star)$. We see that we find the expected cost to be decreasing w.r.t. M or L_1 when $L_1^\star > L_1^{\star,\text{ 1 price-level}}$ or $L_1^\star < L_1^{\star,\text{ 1 price-level}}$ respectively. Considering our constraint $S = M + L_1 + L_2$, this means that our truely optimal solution is $L_1^\star = L_1^{\star,\text{ 1 price-level}}$. We indeed find the two price-level solution in the case where $\psi \to -\infty$ is the same as in the one price-level case, as expected.

4.2.3 Example: Exponentially Distributed Order Flow at the First Two Price-Levels

To express the solution in Proposition 4.4 analytically, we again assume $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, and $\psi \sim Q_2 - Y$, where Q_2 is the length of the second price-level queue and Y is independent and exponentially distributed with rate λ_ψ. We must then know the cumulative distribution function of the convolution of the random variables ξ and ψ. This is given in the following lemma.

Lemma 4.1. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, and $\psi \sim Q_2 - Y$, where Q_2 is the length of the second price-level queue and Y is independent and exponentially distributed with rate λ_ψ, the cumulative distribution function of the convolution of the random variables ξ and ψ is given by*

$$F_{\xi+\psi}(A) = \begin{cases} \frac{\lambda_\xi}{\lambda_\xi + \lambda_\psi} e^{-\lambda_\psi (Q_2 - A)} & , \ A \le Q_2 \\ 1 - \frac{\lambda_\psi}{\lambda_\xi + \lambda_\psi} e^{-\lambda_\xi (A - Q_2)} & , \ A > Q_2. \end{cases} \tag{4.31}$$

The proof to this lemma can be found in section A.4 of the Proofs and can be used to derive an analytical expression for the two price-level case of the static Optimal Placement Problem 4.1. The following corollary describes this analytical solution.

Corollary 4.1. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, and $\psi \sim Q_2 - Y$, where Q_2 is the length of the second price-level queue and Y is independent and exponentially distributed with rate λ_ψ, the solution as in Proposition 4.3 is given by*

$$M^\star = S - \frac{1}{\lambda_\xi} \log\left(\frac{\lambda_\psi}{(\lambda_\xi + \lambda_\psi)} \frac{h + \delta + r + \rho_u + \theta}{-h - f + \rho_u + \theta}\right) + Q_1, \tag{4.32}$$

where $L_i^\star = S - M^\star \in [0, S]$ and $L_j^\star = 0$ for $j \ne i$. The optimal depth for limit orders is price-level i, where $i = 1$ if λ_ξ satisfies

$$\lambda_\xi < \frac{\lambda_\psi \delta}{h + r + \rho_u + \theta}, \tag{4.33}$$

and $i = 2$ if λ_ξ satisfies

$$\lambda_\xi \ge \frac{\lambda_\psi \delta}{h + r + \rho_u + \theta}. \tag{4.34}$$

4.3 Solution Considering Three Price-Levels

In this section we will review the case where the trader can post MOs and LOs at the first, second and third price-level. We again start by expressing the number of our orders filled at the third price-level, OF_3, as a function of the order outflows at the first, second and third price-level, ξ, ψ_1 and ψ_2 respectively.

$$\begin{aligned}\mathrm{OF}_3 =& \min(\max(\xi - Q_1 - L_1 + \psi_1 - Q_2 - L_2 + \psi_2 - Q_3, 0), L_3) \\ =& (\xi - Q_1 - L_1 + \psi_1 - Q_2 - L_2 + \psi_2 - Q_3)_+ - (\xi - Q_1 - L_1 + \psi_1 - Q_2 - L_2 + \psi_2 - Q_3 - L_3)_+ \\ =& (\xi + \psi_1 + \psi_2 - Q_1 - L_1 - Q_2 - L_2 - Q_3) \cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\ & + L_3 \cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+L_3\}},\end{aligned} \tag{4.35}$$

and the cost function in this case is described by

$$\begin{aligned} v(\boldsymbol{X}, \xi, \psi_1, \psi_2) :=& (h+f)M - (h+r)\mathrm{OF}_1 \\ & - (h+\delta+r)\mathrm{OF}_2 - (h+2\delta+r)\mathrm{OF}_3 \\ & + \rho_u(S - A(\boldsymbol{X}, \xi, \psi_1, \psi_2))_+ + \rho_o(A(\boldsymbol{X}, \xi, \psi_1, \psi_2) - S)_+ \\ & + \theta\left(M + L_1 + L_2 + L_3 + (S - A(\boldsymbol{X}, \xi, \psi_1, \psi_2))_+\right)\end{aligned} \tag{4.36}$$

where $A(\boldsymbol{X}, \xi, \psi_1, \psi_2)$ is simply given by (4.2). Again, under the condition that $S = M + L_1 + L_2 + L_3$, we have that $A - S = 0$ and $(S - A)_+ = (S - A)1_{\{\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}}$, which allows us to rewrite the expected cost function as

$$\begin{aligned}\mathbb{E}[v(\boldsymbol{X}, \xi, \psi_1, \psi_2)] :=& \mathbb{E}[(h+f)M - (h+r)\mathrm{OF}_1 - (h+\delta+r)\mathrm{OF}_2 - (h+2\delta+r)\mathrm{OF}_3 \\ & + (\rho_u + \theta)(S - A(\boldsymbol{X}, \xi, \psi_1, \psi_2))1_{\{\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} + \theta S] \\ =& \mathbb{E}[(h+f)M - (h+r)\mathrm{OF}_1 - (h+\delta+r)\mathrm{OF}_2 - (h+2\delta+r)\mathrm{OF}_3 \\ & + (\rho_u + \theta)(S - (M + \mathrm{OF}_1 + \mathrm{OF}_2 + \mathrm{OF}_3))1_{\{\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} + \theta S].\end{aligned} \tag{4.37}$$

We can formulate our the solution to this three price-level case of the Optimal Placement Problem 4.1 in the form of the following proposition.

Proposition 4.4. *Assume that ξ, ψ_1, and ψ_2 have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocation for the $n = 3$ case of the Order Placement Problem 4.1 satisfies the following, based on the valuw of ρ_u.*

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+2\delta+f+r}{F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3)} - (h + 2\delta + r + \theta)$, then $(M^*, L_1^*, L_2^*, L_3^*) = (S, 0, 0, 0)$ is an optimal allocation.

(ii) If $\rho_u \leq \underline{\rho_u} = \frac{2h+2\delta+f+r}{F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S)} - (h + 2\delta + r + \theta)$, then $M^* = 0$ and L_1^* satisfies

$$\delta + (h + r + \rho_u + \theta)F_\xi(Q_1 + L_1^\star) - (h + \delta + r + \rho_u + \theta)F_{\xi+\psi_1}(Q_1 + L_1^\star + Q_2) = 0, \tag{4.38}$$

L_2^* satisfies

$$\begin{aligned}\delta + (h + \delta + r + \rho_u + \theta)F_{\xi+\psi_1}(Q_1 + L_1^\star + Q_2 + L_2^\star) \\ - (h + 2\delta + r + \rho_u + \theta)F_{\xi+\psi_1+\psi_2}(Q_1 + L_1^\star + Q_2 + L_2^\star + Q_3) = 0,\end{aligned} \tag{4.39}$$

and L_3^* satisfies

$$L_3^\star = S - M - L_1^\star - L_2^\star. \tag{4.40}$$

(iii) *If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, then the optimal allocation is a mix of limit (L_1^*, L_2^*, L_3^*) and market $(M^\star)$ orders, where $L_1^\star$ satisfies (4.38), $L_2^\star$ satisfies (4.39), $L_3^\star$ satisfies (4.40) and $M^\star$ satisfies*

$$M^\star = S - F^{-1}_{\xi+\psi_1+\psi_2}\left(\frac{2h + 2\delta + f + r}{h + 2\delta + r + \rho_u + \theta}\right) + Q_1 + Q_2 + Q_3, \tag{4.41}$$

where $F_\xi(x) = \mathbb{P}(\xi \le x)$ is the distribution of the outflow at the first price-level ξ, $F_{\xi+\psi_1}(x) = \mathbb{P}(\xi+\psi_1 \le x)$ is the distribution of the outflow at the first and second price-level ξ, ψ_1, $F_{\xi+\psi_1+\psi_2}(x) = \mathbb{P}(\xi + \psi_1 + \psi_2 \le x)$ is the distribution of the outflow at the first, second and third price-level ξ, ψ_1,ψ_2 and $F^{-1}_\xi, F^{-1}_{\xi+\psi_1}, F^{-1}_{\xi+\psi_1+\psi_2}$ their left-inverses.

The proof can be found in section A.5 and methods for finding the root of (4.38) and (4.39) are discussed in section 3.1.

4.3.1 Example: Exponentially Distributed Order Flow at the First Three Price-Levels

To express the solution in Proposition 4.4 analytically, we again assume $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$, where Q_2 and Q_3 is the length of the second and third price-level queues and Y_1 and Y_2 are independent and exponentially distributed with rates λ_{ψ_1} and λ_{ψ_2}. We must then know the cumulative distribution function of the convolution of the random variables ξ, ψ_1 and ψ_2. This is given in the following lemma.

Lemma 4.2. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$, where Q_2 and Q_3 is the length of the second and third price-level queues and Y_1 and Y_2 are independent and exponentially distributed with rates λ_{ψ_1} and λ_{ψ_2}, the cumulative distribution function of the convolution of the random variables ξ, ψ_1 and ψ_2 is given by*

$$F_{\xi+\psi_1+\psi_2}(A) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} & , \ A \le Q_2 + Q_3 \\ 1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} e^{-\lambda_\xi(A-(Q_2+Q_3))} & , \ A > Q_2 + Q_3 \end{cases} \tag{4.42}$$

The proof to this lemma can be found in section A.6 of the Proofs and can be used to derive an analytical expression for the three price-level case of the static Optimal Placement Problem 4.1. The following corollary describes this analytical solution.

Corollary 4.2. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$, where Q_2 and Q_3 is the length of the second and third price-level queues and Y_1 and Y_2 are independent and exponentially distributed with rates λ_{ψ_1} and λ_{ψ_2}, the solution as in Proposition 4.4 is given by*

$$M^\star = S - \frac{1}{\lambda_\xi} \log\left(\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_\xi + \lambda_{\psi_2})} \frac{h + 2\delta + r + \rho_u + \theta}{-h - f + \rho_u + \theta}\right) + Q_1, \tag{4.43}$$

where $L_i^\star = S - M^\star \in [0, S]$ and $L_j^\star = 0$ for all $j \ne i$. The optimal depth for limit orders is price-level i, where $i = 1$ if λ_ξ satisfies

$$\lambda_\xi < \frac{\lambda_{\psi_1}\delta}{h + r + \rho_u + \theta}, \tag{4.44}$$

$i = 2$ if λ_ξ satisfies

$$\frac{\lambda_{\psi_1}\delta}{h + r + \rho_u + \theta} \le \lambda_\xi < \frac{\lambda_{\psi_2}\delta}{h + \delta + r + \rho_u + \theta}, \tag{4.45}$$

and $i = 3$ if λ_ξ satisfies

$$\lambda_\xi \ge \frac{\lambda_{\psi_2}\delta}{h + \delta + r + \rho_u + \theta}. \tag{4.46}$$

4.4 General Solution for n Price-Levels

Putting the solutions of the one, two, three and four price-level cases next to each other, we find the pattern starting to emerge (where the four price-level case is solved in Appendix section B.2). We can formulate the general solution to the Optimal Placement Problem 4.1 considering n price levels in the following proposition.

Proposition 4.5. *Assume that the queue outflows ξ and ψ_i (with $i = 1, 2, ..., (n-1)$) have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocation for the n price-level case of the Order Placement Problem 4.1 satisfies the following, based on the value of ρ_u.*

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+(n-1)\delta+f+r}{F_{\xi+\psi_1+...+\psi_{n-1}}(Q_1+...+Q_n)} - (h + (n-1)\delta + r + \theta)$, then $(M^*, L_1^*, ..., L_{n-1}^*, L_n^*) = (S, 0, ..., 0, 0)$ is an optimal allocation.

(ii) If $\rho_u \leq \underline{\rho_u} = \frac{2h+(n-1)\delta+f+r}{F_{\xi+\psi_1+...+\psi_{n-1}}(Q_1+...+Q_n+S)} - (h + (n-1)\delta + r + \theta)$, then $M^\star = 0$ and $L_i^\star$ (with $i = 1, 2, ..., (n-1)$) must satisfy

$$\begin{aligned} &\delta + (h + (i-1)\delta + r + \rho_u + \theta)F_{\xi+\psi_1+...+\psi_{i-1}}(\sum_{k=1}^{i}(Q_k + L_k^\star)) \\ &- (h + i\delta + r + \rho_u + \theta)F_{\xi+\psi_1+...+\psi_i}(\sum_{k=1}^{i}(Q_k + L_k^\star) + Q_{i+1}) = 0 \end{aligned} \tag{4.47}$$

and $L_n^\star$ satisfies

$$L_n^\star = S - M^\star - \sum_{i=1}^{n-1} L_i^\star. \tag{4.48}$$

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, then the optimal allocation is a mix of limit and market orders, where $L_i^\star$ (with $i = 1, 2, ..., (n-1)$) satisfies (4.47), $L_n^\star$ satisfies (4.48) and $M^\star$ satisfies

$$M^* = S - F^{-1}_{\xi+\psi_1+...+\psi_{n-1}}\left(\frac{2h + (n-1)\delta + f + r}{h + (n-1)\delta + r + \rho_u + \theta}\right) + Q_1 + ... + Q_n, \tag{4.49}$$

where $F_\xi(x) = \mathbb{P}(\xi \leq x)$ is the distribution of the outflow at the first price-level ξ, $F_{\xi+\psi_1}(x) = \mathbb{P}(\xi + \psi_1 \leq x)$ is the distribution of the outflow at the first and second price-level ξ and ψ_1, and $F_{\xi+\psi_1+...\psi_{n-1}}(x) = \mathbb{P}(\xi + \psi_1 + ... + \psi_{n-1} \leq x)$ is the distribution of the outflow at the first up to the n-th price-level ξ and ψ_1. $F^{-1}_{\xi+\psi_1+...+\psi_{n-1}}$ is the left-inverse of the cumulative distribution function describing the outflow at the first up to the n-th price-level.

Methods for finding the root of (4.47) are discussed in section 3.1.

4.4.1 Example: Exponentially Distributed Order Flow at All Price-Levels

To express the solution in Proposition 4.5 analytically, we again assume $\xi \sim X$, where X is exponentially distributed with rate λ_ξ and $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are independent and exponentially distributed with rate λ_{ψ_i}. We must then know the cumulative distribution function of the convolution of the random variables $\xi, \psi_1, ..., \psi_{n-1}$. This is given in the following lemma.

Lemma 4.3. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ and $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are independent and exponentially distributed with rate λ_{ψ_i}, the cumulative distribution function of the convolution of the random variables $\xi, \psi_1, ...\psi_i$ for $i = 1, 2, ...n-1$ is given by*

$$F_{\xi+\psi_1+\ldots+\psi_i}(A) = \begin{cases} \sum_{k=1}^{i} \frac{\lambda_\xi \prod_{j\neq k}^{i} \lambda_{\psi_j}}{(\lambda_\xi+\lambda_{\psi_k}) \prod_{j\neq k}^{i}(\lambda_{\psi_j}-\lambda_{\psi_k})} e^{-\lambda_{\psi_k}((\sum_{k=2}^{i+1} Q_k)-A)} & , \ A \leq \sum_{k=2}^{i+1} Q_k \\ 1 - \frac{\prod_{k=1}^{i} \lambda_{\psi_k}}{\prod_{k=1}^{i}(\lambda_\xi+\lambda_{\psi_k})} e^{-\lambda_\xi(A-(\sum_{k=2}^{i+1} Q_k))} & , \ A > \sum_{k=2}^{i+1} Q_k \end{cases} \tag{4.50}$$

The proof to this lemma can be found in section A.7 of the Proofs and can be used to derive an analytical expression for the general n price-level case of the static Optimal Placement Problem 4.1. The following corollary describes this analytical solution.

Corollary 4.3. *If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ and $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are independent and exponentially distributed with rate λ_{ψ_i}, the solution as in Proposition 4.5 is given by*

$$M^\star = S - \frac{1}{\lambda_\xi} \log \left(\frac{\prod_{i=1}^{n-1} \lambda_{\psi_i}}{\prod_{i=1}^{n-1}(\lambda_\xi + \lambda_{\psi_i})} \frac{h + (n-1)\delta + r + \rho_u + \theta}{-h - f + \rho_u + \theta} \right) + Q_1 \tag{4.51}$$

*where $L_i^\star = S - M^\star \in [0, S]$ and $L_j^\star = 0$ for all $j \neq i$. The optimal depth of limit orders, the level L_i, is reached at the **critical point** when*

$$\frac{\lambda_{\psi_{i-1}}\delta}{h + (i-2)\delta + r + \rho_u + \theta} \leq \lambda_\xi < \frac{\lambda_{\psi_i}\delta}{h + (i-1)\delta + r + \rho_u + \theta}. \tag{4.52}$$

5 | The Dynamic Optimal Placement Problem

The previous chapter examined the static case where the trader placed her limit and market orders at time $t = 0$ and could not touch them until the terminal time $t = T$, when the limit orders were either filled or had to be cancelled and replaced with market orders if they remained unfilled. However, big deviations in the calibrated parameters or distributions, caused by large incoming limit/market orders after time $t = 0$ for example, would lead to higher costs or missed potential savings for this static strategy. A dynamic model that allowed the trader to adjust her strategy at intermediate times means much more flexibility when such deviations in order flow occur.

In this chapter we will again consider the Optimal Placement Problem, but this time considering it to be dynamic, allowing the trader to place limit orders at a certain depth, cancel unfilled limit orders and place market orders at different time steps between $t = 0$ and $t = T$. We build forth upon the method used in [5], introducing mid-price drift and general penalty functions for price impact of placed orders and under-filling the target amount.

In [8] they work out an optimal stopping problem where the trader is allowed to either post market orders or limit orders at the best price-level. This can be seen as the dynamic counterpart of our one price-level static problem in section 4.1. To also include deeper price levels in our dynamic model, we choose to focus on the combination of an optimal stopping problem, that tells the trader when it is optimal to market orders, and an optimal control problem that tells the trader how deep to optimally place her limit orders. We solve the Hamilton-Jacobi-Bellman equations by iteratively increasing the inventory that the trader has to liquidate, to eventually end up at a recursive analytical solution for all inventory sizes.

Our full model contains mid-price drift and a running inventory penalty. These are essential in creating an incentive for the trader to place MOs. Without these terms, the optimal stopping time for all inventory size would be an instant before the terminal time. The underlying logic is that without price drift, the expected future mid-price is equal to the current mid-price. Therefore, a trader does not stand to gain from posting MO during the trading window. Instead she can post her orders in the form of LOs, hoping to fill them, and execute her remaining inventory at an instant before the terminal time. If there is also no term to penalise the trader for holding on to her inventory during the trading window, we indeed find all stopping times to be just before the terminal time, making the inclusion of MOs in our problem redundant. In the absence of price drift and running inventory penalty, the optimal strategy degenerates to the solution of the optimal control problem described in [20], without terminal MO penalty.

5.1 Full Model Setup

We assume that the risk-free rate is zero and define the Optimal Placement Problem model using the following terms.

- S is the number of shares the trader wishes to liquidate
- T is the terminal time at which the trader wishes to have liquidated her portfolio
- $p = (p_t)_{0 \leq t \leq T}$ is the asset's mid-price, which is described by Arithmetic Brownian Motion, namely

$$dp_t = \mu dt + \sigma_p dW_t^p \tag{5.1}$$

- $\Delta = (\Delta_t)_{0 \leq t \leq T}$ denotes the spread and in our case we will mostly consider half of the spread $\frac{\Delta}{2}$. This spread can be modelled as a mean reverting random variable (like an Ornstein-Uhlenbeck process without drift), but this makes the problem considerably more complicated to solve, so here it is assumed to be constant for the duration of the trading window
- $\delta = (\delta_t)_{0 \leq t \leq T}$ denotes the depth at which the trader posts her limit sell orders measured from the best sell order queue, i.e. the trader posts LOs at a price of $p_t + \frac{\Delta_t}{2} + \delta_t$ at time t
- $\nu = (\nu_t)_{0 \leq t \leq T}$ denotes the Poisson process (with intensity $\lambda_\nu > 0$) corresponding to the number of market buy orders (from other traders) that have arrived up until time t
- $L^\delta = (L_t^\delta)_{0 \leq t \leq T}$ denotes the (controlled) counting process corresponding to the number of the trader's limit sell-orders which are executed at a depth δ by incoming market buy orders
- $M^{\boldsymbol{\tau}} = (M_t^{\boldsymbol{\tau}})_{0 \leq t \leq T}$ is the counting process denoting the trader's placed MOs
- $\boldsymbol{\tau} = \{\tau_k : k = 1, ..., K\}$ with $K \leq S$ is the corresponding sequence of stopping times at which the trader places (and immediately executes) MOs, with $M_t = \sum_{k=1}^{K} 1_{\tau_k \leq t}$
- $P(\delta) = e^{-\kappa\delta}$ with $\kappa > 0$ is the probability that the trader's LO will be filled when a buy MO arrives
- $X^\delta = (X_t^\delta)_{0 \leq t \leq T}$ is the trader's wealth process
- $I_t^{\boldsymbol{\tau},\delta} = S - L_t^\delta - M_t^{\boldsymbol{\tau}}$ is the trader's inventory that remains to be liquidated
- $\rho_u = \rho_u(I_t^{\boldsymbol{\tau},\delta})$ is the risk premium paid for executing MOs at terminal time T for the remaining inventory the trader still has. It is a function of the size of the remaining inventory and often assumed to be a linear function, but here we will keep it general by denoting $\rho_u(I_T^{\boldsymbol{\tau},\delta})$
- $\theta = \theta(I_t^{\boldsymbol{\tau},\delta})$ is the penalty term on the running inventory, where the size of the penalty depends on the size of the inventory the trader holds during the time frame. The reason we need this running inventory penalty is that otherwise there is no incentive for the trader to place market orders until an instant before the terminal time T.
- r, f are the rebate and fee earned/paid for filling a limit or market order resp.

The trader's wealth is given by the SDE

$$dX_t^{\boldsymbol{\tau},\delta} = (p_t + \frac{\Delta}{2} + \delta_t + r)dL_t^\delta + (p_t - \frac{\Delta}{2} - f)dM_t^{\boldsymbol{\tau}} \tag{5.2}$$

where the first term describes the amount (plus rebate) earned by filling a limit sell-order (at depth δ) and the second term is the price and fee paid for placing and filling a market sell-order. The number of limit orders sold in the interval $(0, t)$ is denoted by the counting process $L^\delta = (L_t^\delta)_{0 \leq t \leq T}$ and the number of market orders denoted by the counting $M^{\boldsymbol{\tau}} = (M_t^{\boldsymbol{\tau}})_{0\le t\le T}$. We now define the trader's performance criteria, which is given by

$$H^{\boldsymbol{\tau},\delta}(t,x,p,i) = \mathbb{E}_{t,x,p,i}[X_T^{\boldsymbol{\tau},\delta} + I_T^{\boldsymbol{\tau},\delta}(p_T - \frac{\Delta}{2} - f - \rho_u(I_T^{\boldsymbol{\tau},\delta})) - \int_t^T \theta(I_u^{\boldsymbol{\tau},\delta})du] \tag{5.3}$$

where the notation $\mathbb{E}_{t,x,p,q}[\cdot]$ represents the expectation conditional on t, $X_{t^-}^{\tau,\delta} = x, p_t = p$ and $I_{t^-}^{\tau,\delta} = i$. Our control variables are the stopping time τ and depth δ, which means that the set of admissible strategies $\mathcal{A}$ consists of seeking over all $\mathcal{F}$-stopping times $\boldsymbol{\tau}$ and the set of $\mathcal{F}$-predictable, bounded from below, depths δ. In this case the value function is given by

$$H(t,x,p,i) = \sup_{(\boldsymbol{\tau},\delta)\in\mathcal{A}} H^{\boldsymbol{\tau},\delta}(t,x,p,i) \tag{5.4}$$

and can be interpreted as the expected cost function, which is maximized by the optimal strategy, in the form of the optimal stopping times $\tau^\star$ and optimal limit order depth $\delta^\star$.

Now, the Dynamic Programming Principle allows us to derive the Dynamic Programming Equation (DPE) (see section 3.2 of the Preliminaries). The DPE, which the value function should satisfy, is in the form of a quasi-variational-inequality (QVI), given by

$$0 = \max\Bigg\{\partial_t H + \mu\partial_p H + \frac{1}{2}\sigma_p^2\partial_{pp}H - \theta(i) + \sup_\delta \lambda_\nu e^{-\kappa\delta}\left[H(t,x+(p+\frac{\Delta}{2}+\delta+r),p,i-1) - H(t,x,p,i)\right] \quad ; \\ \left[H(t,x+(p-\frac{\Delta}{2}-f),p,i-1) - H(t,x,p,i)\right]\Bigg\}, \tag{5.5}$$

with boundary and terminal conditions

$$\begin{aligned} H(t,x,p,0) &= x, \\ H(T,x,p,i) &= x + i(p - \frac{\Delta}{2} - f - \rho_u(i)). \end{aligned} \tag{5.6}$$

The derivation can be found in section B.3.1 of the Appendix. The first part inside the max function in (5.5) describes the evolution of the value function when the trader chooses to "continue" and place limit orders at the optimal depth $\delta^\star$. The second part inside the max function represents when the trader decides to "stop" and place a market order. She does so when the value function H is equal to the reward function G. The reward function is given by

$$G = H(t, x + (p - \frac{\Delta}{2} - f), p, i-1), \tag{5.7}$$

which represents the trader executing one MO, increasing her wealth function by $p - \frac{\Delta}{2} - f$ and lowering her inventory by 1.

Looking at the boundary and terminal conditions in (5.6) , we choose to make the ansatz for the value function $H(t,x,p,i) = x + i(p - \frac{\Delta}{2} - f) + h(t,i)$, based on [8]. Substituting this into our QVI, we see that $h(t,i)$ satisfied the simplified QVI

$$0 = \max\Bigg\{\partial_t h + \mu i - \theta(i) + \sup_\delta \lambda_\nu e^{-\kappa\delta}\left[\delta + \Delta + r + f + h(t,i-1) - h(t,i)\right] \quad ; \\ \left[h(t,i-1) - h(t,i)\right]\Bigg\}, \tag{5.8}$$

with boundary and terminal conditions

$$\begin{aligned} h(t,0) =&0, \\ h(T,i) =&-i\rho_u(i), \qquad \text{for } \; i = 1, ..., \mathcal{R}. \end{aligned} \tag{5.9}$$

Focusing on the supremum term in (5.8), we can find the value of δ that maximizes this term, by taking the derivative w.r.t. δ and setting it equal to zero. We find that the optimal strategy $\delta^\star$ in feedback control form is given by

$$\delta^\star(t,p,i) = \frac{1}{\kappa} + [h(t,i) - h(t,i-1)] - \Delta - r - f. \tag{5.10}$$

We see from the stopping region (the second part inside the max function) in (5.8), that is is optimal to execute an MO at stopping time τ_i when

$$h(\tau_i, i-1) - h(\tau_i, i) = 0 \tag{5.11}$$

This can be interpreted as executing an MO whenever doing so does not change h, and thus increase the value function by half of the spread plus the fee (based on the ansatz). Combining this observation with (5.10) we find a lower bound for the optimal depth of

$$\delta^\star \geq \frac{1}{\kappa} - \Delta - r - f. \tag{5.12}$$

since we know $h(t,i-1) - h(t,i) \leq 0$ from the QVI (5.8). Now we assume that the trader may not post LOs inside the spread, so we require $\delta > 0$. Therefore we must require $\Delta < \frac{1}{\kappa} - r - f$.

Substituting the optimal depth into the simplified QVI (5.8), we find that $h(t,i)$ satisfies

$$0 = \max\left\{ \partial_t h + \mu i - \theta(i) + \frac{\lambda_\nu}{e}\frac{e^{\kappa(\Delta+r+f)}}{\kappa} e^{-\kappa[h(t,i)-h(t,i-1)]} \;\; ; \;\; [h(t,i-1) - h(t,i)] \right\}. \tag{5.13}$$

If the trader has inventory left at terminal time T, she must execute the remaining inventory as MOs at a cost of $\rho_u(I_T^{\boldsymbol{\tau},\delta})$ per share. At an instant time step before this terminal time, however, the trader may execute MOs at a cost of $\frac{\Delta}{2} + f < \rho_u$, so it is never optimal for her to wait until time T to execute an MO. As a result, the left-limit of the value function is not equal to its value at the terminal time, and we have

$$H(T^-,x,p,i) = -\frac{\Delta}{2} - f + H(T^-,x,p,i-1) \tag{5.14}$$

for every $i > 0$, which means that $h(T^-,i) = h(T^-,i-1) = 0$ (since $h(t,0) = 0$). The phenomenon where the left limit of the solution is different from the terminal condition is sometimes called *face-lifting*[36].

To be able to solve the QVI analytically, it must be further simplified. We apply the transformation

$$h(t,i) = \frac{1}{\kappa}\log(w(t,i)) \tag{5.15}$$

which transforms our simplified QVI into the following equation:

$$\max\left(\frac{1}{\kappa w(t,i)}(\partial_t w(t,i) + \kappa(\mu i - \theta(i))w(t,i) + \tilde{\lambda}_\nu w(t,i-1)) \;\; ; \;\; \frac{1}{\kappa}\log\left(\frac{w(t,i-1)}{w(t,i)}\right) \right) = 0 \tag{5.16}$$

where $\tilde{\lambda}_\nu = \frac{\lambda_\nu}{e} e^{\kappa(\Delta+r+f)}$.

We can rewrite QVI (5.16) by multiplying the left term inside the max function by $\kappa w(t,i)$ and setting the right term equal to zero and taking the exponent. We find the simplified QVI

$$\max\Big(\partial_t w(t,i) + \kappa(\mu i - \theta(i))w(t,i) + \tilde{\lambda}_\nu w(t,i-1)\ ;\ w(t,i-1) - w(t,i)\Big) = 0 \tag{5.17}$$

with boundary and terminal condition

$$\begin{aligned} w(t,0) &= 1, \\ w(T,i) &= e^{-i\kappa\rho_u(i)} \quad \text{for } i = 1, ..., \mathcal{R}. \end{aligned} \tag{5.18}$$

The optimal depth $\delta^\star$ at which to place limit orders, when the trader holds an inventory of size i is now given by

$$\delta^\star(t,i) = \frac{1}{\kappa} - \Delta - r - f + \frac{1}{\kappa}\log\left(\frac{w(t,i)}{w(t,i-1)}\right). \tag{5.19}$$

The left part inside the max function in equation (5.17) represents the continuation value, and describes evolution of the value function for a trader holding an inventory of size i at a time t, assuming the trader follows the optimal depth scheme in (5.19). The stopping time τ_i, at which the trader should (optimally) place an MO instead of continuing trying to liquidate her inventory of size i using LOs, is then defined as the moment where this continuation value is equal to the stopping value (the right part inside the max function in equation (5.17)). In that case the trader chooses to execute an MO, as this guarantees this value, rather than the uncertain continuation value. Setting the left term inside the max function in QVI (5.17) equal to zero gives us a differential equation (DE) of the form

$$\partial_t g_i(t) + \kappa(\mu i - \theta(i))g_i(t) + \tilde{\lambda}_\nu g_{i-1}(t) = \tag{5.20}$$

where $g_i(t)$ and $g_{i-1}(t)$ are functions that solve this DE. Then $w(t,i)$, that solves (5.17), is given by

$$w(t,i) = g_i(t)\cdot 1_{\{t\le\tau_i\}} + g_{i-1}(t)\cdot 1_{\{t>\tau_i\}} \tag{5.21}$$

where τ_i is the optimal stopping time at an inventory of size i. We can now solve QVI iteratively, first plugging in the known boundary condition for $i=0$, to find $w(t,1)$, then plugging this into the QVI to find $w(t,2)$ and continuing until we have a solution for a general inventory size S.

5.2 Analytical Solution

5.2.1 The $i = 1$ Case:

Let us assume that fees, rebates and spread are constant, then the problem we are solving becomes

$$\begin{aligned} \max\Big(\partial_t w(t,1) + \kappa(\mu - \theta(1))w(t,1) + \tilde{\lambda}_\nu\ ;\ 1 - w(t,1)\Big) &= 0 \\ w(T,1) &= e^{-\kappa\rho_u(1)} \end{aligned} \tag{5.22}$$

where $\tilde{\lambda}_\nu = \frac{\lambda_\nu}{e}e^{\kappa(\Delta+r+f)}$. Using the fact that it is never optimal to stop at terminal time T (since executing an MO at that time will induce extra penalty cost $\rho_u(I_t^{\tau,\delta})$), we see that we have to solve the ODE

$$\partial_t g_1(t) + \kappa(\mu - \theta(1))g_1(t) + \tilde{\lambda}_\nu = 0, \quad g_1(T^-) = 1. \tag{5.23}$$

The solution is given by

$$g_1(t) = e^{\kappa(\mu-\theta(1))(T-t)} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))}\left(e^{\kappa(\mu-\theta(1))(T-t)} - 1\right). \tag{5.24}$$

We see that $g_1(t) \geq 1$ for all $t \in (0,T)$ when

$$\mu \geq \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1), \tag{5.25}$$

which means that the continuation value is always larger than the execution value (the price increase is larger than the running inventory costs and the execution value), hence the solution to the QVI (5.22) is

$$w(t,1) = g_1(t) \cdot 1_{t<T} + e^{-\kappa\rho_u(1)} \cdot 1_{t=T} \tag{5.26}$$

where $g_1(t)$ is given by (g1-full-model) and it is never optimal to execute an MO, except for an instant before the terminal time T.

Moreover, based on equation (5.19), the optimal depth to post the limit order is given by

$$\delta^\star(t,1) = \frac{1}{\kappa} - \Delta - r - f + \frac{1}{k}\log\left(e^{\kappa(\mu-\theta(1))(T-t)} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))}\left(e^{\kappa(\mu-\theta(1))(T-t)} - 1\right)\right). \tag{5.27}$$

We also find that if

$$\mu < \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1), \tag{5.28}$$

that $g_1(t) < 1$ for all $t \in (0,T)$, i.e. the continuation value is always less than the execution value, and the solution to the QVI (5.22) is

$$w(t,1) = 1_{t<T} + e^{-\kappa\rho_u(1)} \cdot 1_{t=T}, \tag{5.29}$$

meaning it is always optimal to execute a market order, so $\tau_1 = 0$.

5.2.2 The $i = 2$ Case:

In this case we are solving the system

$$\begin{cases} \max\left(\partial_t w(t,2) + \kappa(2\mu - \theta(2))w(t,2) + \tilde{\lambda}_\nu w(t,1)\ ;\ w(t,1) - w(t,2)\right) = 0 \\ w(T,2) = e^{-\kappa\rho_u(2)}. \end{cases} \tag{5.30}$$

We only consider the non-trivial case where $\mu \geq \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1)$ and we use the trivial optimal strategy to execute MOs an instant before the terminal time as our terminal condition face-lifted to $w(T^-,2) = 1 \cdot w(T^-,2) = 1$. We must now determine the time τ_2 when the solution to the QVI switches from its immediate execution value to the continuation value, namely where $w(t,2) = w(t,1)$.To guarantee continuity of the QVI, we need to ensure that $w(t,2)$ and its derivative are continuous at this point, i.e.

$$w(\tau_2^-,2) = w(\tau_2,2) \quad \text{and} \quad \partial_t w(\tau_2^-,2) = \partial_t w(\tau_2,2) = \partial_t w(\tau_2,1). \tag{5.31}$$

From the QVI (5.30), we know that in the continuation region, we have

$$\begin{aligned} 0 =& \partial_t w(\tau_2,2) + \kappa(2\mu - \theta(2))w(\tau_2,2) + \tilde{\lambda}_\nu w(\tau_2,1) \\ =& \partial_t w(\tau_2,1) + \kappa(2\mu - \theta(2))w(\tau_2,1) + \tilde{\lambda}_\nu w(\tau_2,1) \\ =& \left(-\kappa(\mu - \theta(1))w(\tau_2,1) - \tilde{\lambda}_\nu\right) + \kappa(2\mu - \theta(2))w(\tau_2,1) + \tilde{\lambda}_\nu w(\tau_2,1) \\ =& w(\tau_2,1)\left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right) - \tilde{\lambda}_\nu, \end{aligned} \tag{5.32}$$

where the third equality follows from the continuation region of (5.22).

We find that the optimal stopping time, at which the trader should execute a market order when she holds an inventory of size 2, solves

$$w(\tau_2, 1) = \frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu}. \tag{5.33}$$

This can be solved explicitly for τ_2, since $w(t,1) = g_1(t)$ is given in equation (5.26). In section B.3.3 we obtain

$$\tau_2 = T - \frac{\log\left[\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu+\theta(1)-\theta(2))+\tilde{\lambda}_\nu} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))}\right)\right] - \log\left[\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))} + 1\right)\right]}{\kappa(\mu - \theta(1))}. \tag{5.34}$$

Furthermore, we have that τ_2 only exists when $w(\tau_2, 1) > 0$, so we require

$$\mu > \frac{-\tilde{\lambda}_\nu}{\kappa} - \theta(1) + \theta(2). \tag{5.35}$$

Comparing that with the requirement for the non-trivial case at the beginning of this subsection, we find a condition on running inventory penalty function, namely $\theta(2) > 2\theta(1)$.

Now that we know the optimal stopping time, we can find the full solution for $w(t,2)$ and for that we must solve the continuation equation backwards from τ_2. This is done by solving the ODE

$$\partial_t g_2(t) + \kappa(2\mu - \theta(2))g_2(t) + \tilde{\lambda}_\nu g_1(t) = 0, \quad g_2(\tau_2) = \Gamma_2 \tag{5.36}$$

where $\Gamma_2 = 1 \cdot w(\tau_2, 1) = \tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1}$ and $g_1(t)$ is given by equation (5.24). The solution is given by

$$\begin{aligned} g_2(t) =& \frac{\tilde{\lambda}_\nu^2}{\kappa(2\mu - \theta(2))\kappa(\mu - \theta(1))} \\ &+ e^{\kappa(2\mu-\theta(2))(\tau_2 - t)}\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu} - \frac{\tilde{\lambda}_\nu^2}{\kappa(2\mu - \theta(2))\kappa(\mu - \theta(1))}\right. \\ &+ \left.\frac{\tilde{\lambda}_\nu^2}{\kappa(\mu + \theta(1) - \theta(2))\kappa(\mu - \theta(1))} e^{\kappa(\mu-\theta(1))(T-\tau_2)} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(1) - \theta(2))} e^{\kappa(\mu-\theta(1))(T-\tau_2)}\right) \\ &+ e^{\kappa(\mu-\theta(1))(T-t)}\left(-\frac{\tilde{\lambda}_\nu^2}{\kappa(\mu + \theta(1) - \theta(2))\kappa(\mu - \theta(1))} - \frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(1) - \theta(2))}\right). \end{aligned} \tag{5.37}$$

So the solution of the QVI (5.30) for $w(t,2)$ is

$$w(t,2) = g_2(t) \cdot 1_{\{t<\tau_2\}} + g_1(t) \cdot 1_{\{t \geq \tau_2\}} \tag{5.38}$$

where $g_2(t)$ is given by (5.37). Finally the optimal depth at which the trader should post her LOs is given by

$$\delta^\star(t,2) = \frac{1}{\kappa} - \Delta - r - f + \frac{1}{\kappa}\log\left(\frac{w(t,2)}{w(t,1)}\right). \tag{5.39}$$

5.2.3 The $i = 3$ Case:

In this case, we are solving the system

$$\begin{cases} \max\left(\partial_t w(t,3) + \kappa(3\mu - \theta(3))w(t,3) + \tilde{\lambda}_\nu w(t,2)\ ;\ w(t,2) - w(t,3)\right) = 0 \\ w(T,3) = e^{-\kappa\rho_u(3)}. \end{cases} \tag{5.40}$$

Let us only consider the non-trivial case where $\mu \geq \frac{-\tilde{\lambda}_\nu}{\kappa} - \theta(1) + \theta(2)$ and we use the trivial optimal strategy to execute MOs an instant before the terminal time as our terminal condition face-lifted to $w(T^-, 3) = 1 \cdot w(T^-, 2) = 1$. We must now determine the time τ_3 when the solution to the QVI switches from its immediate execution value to the continuation value, namely where $w(t,3) = w(t,2)$. To guarantee continuity of the QVI, we need to ensure that $w(t,2)$ and its derivative are continuous at this point, i.e.

$$w(\tau_3^-, 3) = w(\tau_3, 3) \quad \text{and} \quad \partial_t w(\tau_3^-, 3) = \partial_t w(\tau_3, 3) = \partial_t w(\tau_3, 2). \tag{5.41}$$

From the QVI (5.40), we know that in the continuation region, we have

$$\begin{aligned}
0 =& \partial_t w(\tau 3, 3) + \kappa(3\mu - \theta(3)) w(\tau_3, 3) + \tilde{\lambda}_\nu w(\tau_3, 2) \\
=& \partial_t w(\tau 3, 2) + \kappa(3\mu - \theta(3)) w(\tau_3, 2) + \tilde{\lambda}_\nu w(\tau_3, 2) \\
=& -\kappa(2\mu - \theta(2)) w(\tau 3, 2) - \tilde{\lambda}_\nu w(\tau 3, 1) + \kappa(3\mu - \theta(3)) w(\tau_3, 2) + \tilde{\lambda}_\nu w(\tau_3, 2) \\
=& (\kappa(\mu + \theta(2) - \theta(3)) + \tilde{\lambda}_\nu) w(\tau_3, 2) - \tilde{\lambda}_\nu w(\tau 3, 1),
\end{aligned} \tag{5.42}$$

where the third equality follows from the continuation region of (5.30).

We find that the optimal stopping time τ_3, at which the trader should execute an MO when she holds an inventory of size 3, solves

$$w(\tau_3, 2) = \frac{\tilde{\lambda}_\nu}{(\kappa(\mu + \theta(2) - \theta(3)) + \tilde{\lambda}_\nu)} w(\tau 3, 1). \tag{5.43}$$

This can be used to find τ_3, since $w(t,2)$ is given by equation (5.38) and $w(t,1)$ is given in equation (5.26). We try this in the section B.3.5, but find an $e^{\kappa(2\mu - \theta(2))(\tau_2 - \tau_3)}$ term and an $e^{\kappa(\mu - \theta(1))(T - \tau_3)}$ term in the equation, which leads to a transcendental equation that cannot be solved analytically, as described in section 3.1. We can solve (5.43) explicitly for the special case where $\theta(2) = 2\theta(1)$, as it then becomes a quadratic equation. This is also shown in section B.3.5.

Once we know the optimal stopping time, we can find the full solution for $w(t,3)$ and for that we must solve the continuation equation backwards from τ_3. We can this by solving the ODE

$$\partial_t g_3(t) + \kappa(3\mu - \theta(3)) g_3(t) + \tilde{\lambda}_\nu g_2(t) = 0, \quad g_3(\tau_3) = \Gamma_3 \tag{5.44}$$

where $\Gamma_3 = \frac{\tilde{\lambda}_\nu}{(\kappa(\mu + \theta(2) - \theta(3)) + \tilde{\lambda}_\nu)} w(\tau 3, 1)$ and $g_2(t)$ is given in equation (5.37). Solving ODE (5.44) in section B.3.4, we find

$$\begin{aligned}
g_3(t) =& e^{\kappa(3\mu - \theta(3))(\tau_3 - t)} \left(\left(1 + \frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(2) - \theta(3))}\right) g_2(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{\kappa(3\mu - \theta(3))\kappa(\mu + \theta(2) - \theta(3))} \right) \\
& - \left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(2) - \theta(3))} g_2(t) + \frac{\tilde{\lambda}_\nu^2}{\kappa(3\mu - \theta(3))\kappa(\mu + \theta(2) - \theta(3))} \right),
\end{aligned} \tag{5.45}$$

where $g_2(t)$ is given in equation (5.37) and τ_3 solves equation (5.43). So the solution of the QVI (5.40) for $w(t,3)$ is given by

$$w(t,3) = g_3(t) \cdot 1_{\{t < \tau_3\}} + g_2(t) \cdot 1_{\{t \geq \tau_3\}}. \tag{5.46}$$

Finally the optimal depth at which the trader should post her LOs is given by

$$\delta^\star(t,3) = \frac{1}{\kappa} - \Delta - r - f + \frac{1}{\kappa} \log\left(\frac{w(t,3)}{w(t,2)} \right). \tag{5.47}$$

5.2.4 The General $i = S$ Case:

The $i = 4$ case can be found in section B.3.6 and our value formulas $g_1(t), g_2(t), g_3(t)$ and $g_4(t)$ are given by

$$\begin{aligned}
g_1(t) =& e^{A_1(T-t)}\left(1+\frac{\tilde{\lambda}_\nu}{A_1}\right)-\frac{\tilde{\lambda}_\nu}{A_1},\\
g_2(t) =& e^{A_2(\tau_2-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_2-A_1})g_1(\tau_2)+\frac{\tilde{\lambda}_\nu^2}{A_2(A_2-A_1)}\right)-\left(\frac{\tilde{\lambda}_\nu}{A_2-A_1}g_1(t)+\frac{\tilde{\lambda}_\nu^2}{A_2(A_2-A_1)}\right),\\
g_3(t) =& e^{A_3(\tau_3-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_3-A_2})g_2(\tau_3)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(\tau_3)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right)\\
&-\left(\frac{\tilde{\lambda}_\nu}{A_3-A_2}g_2(t)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(t)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right),\\
g_4(t) =& e^{A_4(\tau_4-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_4-A_3})g_3(\tau_4)+\frac{\tilde{\lambda}_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(\tau_4)\right.\\
&\left.+\frac{\tilde{\lambda}_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(\tau_4)+\frac{\tilde{\lambda}_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\right)\\
&-\left(\frac{\tilde{\lambda}_\nu}{A_4-A_3}g_3(t)+\frac{\tilde{\lambda}_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(t)\right.\\
&\left.+\frac{\tilde{\lambda}_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(t)+\frac{\tilde{\lambda}_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\right)\Bigg),
\end{aligned} \tag{5.48}$$

where $A_1 = \kappa(\mu - \theta(1))$, $A_2 = \kappa(2\mu - \theta(2))$ and $A_3 = \kappa(3\mu - \theta(3))$ and where we have the stopping times τ_i satisfying

$$\begin{aligned}
g_1(\tau_2) =& \frac{\tilde{\lambda}_\nu}{A_2-A_1+\tilde{\lambda}_\nu},\\
g_2(\tau_3) =& \frac{\tilde{\lambda}_\nu}{A_3-A_2+\tilde{\lambda}_\nu}g_1(\tau_3),\\
g_3(\tau_4) =& \frac{\tilde{\lambda}_\nu}{A_4-A_3+\tilde{\lambda}_\nu}g_2(\tau_4).
\end{aligned} \tag{5.49}$$

As explained in the $i = 3$ case, τ_i cannot be expressed explicitly for $i \geq 3$, since the formula in (5.49) leads to transcendental equations as explained in section 3.1 and shown for $i = 3$ in section B.3.5. This also means the general formula for $g_S(t)$ cannot be expressed explicitly for $S \geq 3$, since it always contains τ_S. In (5.48) it is crucial to note that $A_i = 0$ for $i = 1, 2, ...$ or $A_i - A_j = 0$ for $i \neq j$ would lead to infinite "blow-up" values, as it would mean dividing by zero. This means certain combinations of mid-price drift and running inventory penalties cannot be combined using our formula (like $\mu = 0.001$ and $\theta(i) = 0.001 \cdot i^2$, since this would mean $A_1 = 0$), but this should not be a problem in real-world situations, since these types of combinations for calibrated parameters are unlikely to occur.

We see the general formula for $g_S(t)$ can be expressed in the following proposition

Proposition 5.1. *The value function of the dynamic Optimal Placement Problem as defined in equation (5.4), which solves QVI (5.5), is given by*

$$H(t, x, p, i) = x + i(p - \frac{\Delta}{2} - f) + \frac{1}{\kappa}\log(w(t, i)), \tag{5.50}$$

where $w(t, i) = g_i(t) \cdot 1_{\{t<\tau_i\}} + g_{i-1}(t) \cdot 1_{\{t\geq\tau_i\}}$ *for* $t \in (0, T)$ *and* $g_i(t)$ *is given by*

$$\begin{aligned} g_S(t) =& e^{A_S(\tau_S - t)}\left(g_{S-1}(\tau_S) + \sum_{i=1}^{S} \frac{\tilde{\lambda}_\nu^i}{\prod_{j=1}^{i}(A_S - A_{S-j})} g_{S-i}(\tau_S)\right) - \left(\sum_{i=1}^{S} \frac{\tilde{\lambda}_\nu^i}{\prod_{j=1}^{i}(A_S - A_{S-j})} g_{S-i}(t)\right) \\ =& e^{A_S(\tau_S - t)} g_{S-1}(\tau_S) + \sum_{i=1}^{S} \frac{\tilde{\lambda}_\nu^i}{\prod_{j=1}^{i}(A_S - A_{S-j})} \left(e^{A_s(\tau_S - t)} g_{S-i}(\tau_S) - g_{S-i}(t)\right) \end{aligned} \tag{5.51}$$

where $A_i = \kappa(i\mu - \theta(i))$ *for* $i \in \mathbb{N}^+$, $A_0 = 0$ *and* $g_0(t) = 1$. *The the stopping time* τ_S *solves the equation*

$$g_{S-1}(\tau_S) = \frac{\tilde{\lambda}_\nu}{A_S - A_{S-1} + \tilde{\lambda}_\nu} g_{S-2}(\tau_S) \tag{5.52}$$

which can be solved using numerical methods, like Newton-Raphson's method, as described in section 3.1. The optimal depth to place limit orders at a certain time t *with remaining inventory* S *is given by*

$$\delta^\star(t, S) = \frac{1}{\kappa} - \Delta - r - f + \frac{1}{\kappa}\log\left(\frac{g_S(t)}{g_{S-1}(t)}\right) \cdot 1_{t<\tau_S} \tag{5.53}$$

and at $t = \tau_S$ *a market order should be placed and the inventory size decreases to* $i = S - 1$.

6 | Numerical Results

In this chapter, we aim to assess and compare the solutions found throughout this thesis, using numerical experiments. Firstly the solution for the static model is reviewed, analysing the optimal order allocation for the 1, 2, 3 and n price-level solution. To asses its performance we compare expected cost of the optimal order allocation to benchmark models. For the dynamic solution we look at the optimal depth and optimal stopping times and test their behaviour to different parameter changes. Finally, we compare the average price per share the dynamic optimal strategy produces to benchmark strategies.

6.1 Static Model

The static model allows the trader to place her S orders in the form of market and limit orders (at different price levels) time $t = 0$. Any orders unfilled at the terminal time $t = T$ are then filled as market orders. The following default parameter values are chosen (similar to [15]): $h = 0.02$, $r = 0.002$, $f = 0.003$, $\delta = 0.01$, $\theta = 0.0005$, $\rho_u = 0.043$, $Q_1 = 2000$, $S = 1000$, and $\lambda_\xi = 1/\bar{\xi} = 1/2200$ with $\xi \sim \text{Exp}(\lambda_\xi)$ and $\psi_i = Q_{i+1} - Y_i$ with $Y_i \sim \text{Exp}(\lambda_\xi)$, to allow for fair benchmark comparison

First we will analyse the optimal order allocation for 1,2 and 3 price-levels, and then move on to the general n price-level solution. Then we examine the performance of the optimal solution in terms of expected costs, comparing it to benchmark strategies that only place 1 type of order (only MOs or only LOs at a certain price-level) and the strategy from the paper by Cont and Kukanov [2017][15].

6.1.1 Optimal MO-LO Split for the One, Two and Three Price-Level Case

To see the evolution of the solution for the static model as the number of considered price-levels increases, we start by analysing the 1 price-level solution and gradually consider deeper and deeper levels.

One Price-Level

Our solution allows for any distribution of the random variable ξ. Obviously the chosen distribution is crucial in determining our optimal solution. Figure 6.1 shows the optimal split between market and limit orders for one price-level in the case where ξ is exponentially, normally or Pareto distirbuted, all with parameters such that their mean is 2200. Furthermore, we see that the under-filling penalty coefficient ρ_u also plays a big part in determining the optimal split. If under-filling is penalized heavily (corresponding to higher ρ_u values) the trader is more conservative with placing limit orders and instead opts for the more certain market orders. In the remainder of this section ξ is assumed to follow an exponential distribution, as this is most common in the literature and allows for an analytical expression of the optimal split, as seen in subsection 4.1.1. The behaviour of our cost function for different values of ρ_u is shown in Figure 6.2. It illustrates why the bounds stated in Proposition 4.2 are so important, as above $\overline{\rho_u}$ or below $\underline{\rho_u}$ the cost function is strictly increasing or decreasing w.r.t. L_1. Figure 6.3 shows how each parameter influences the optimal split between market and limit orders.

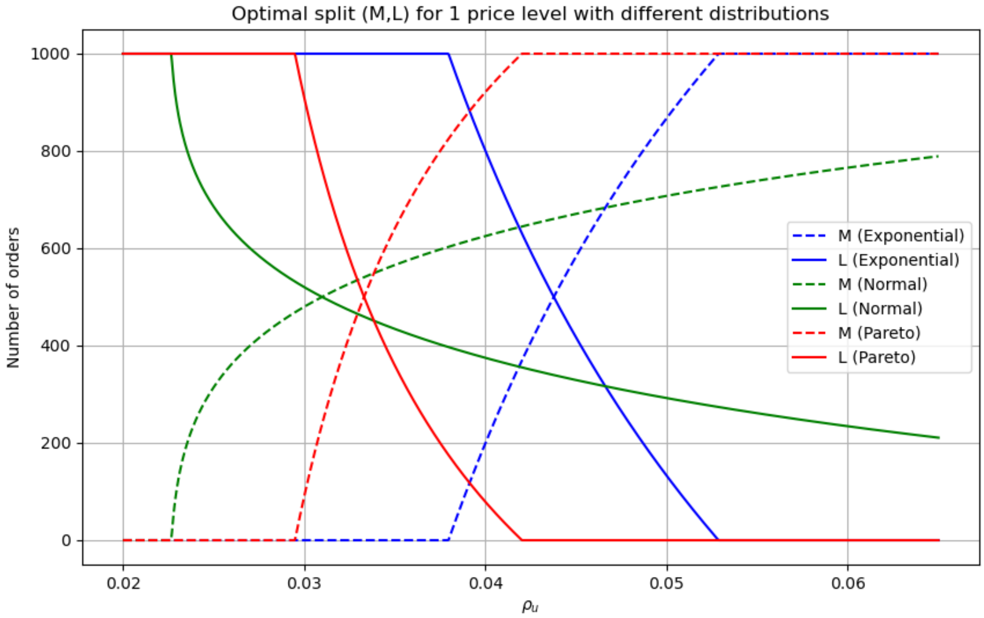

Figure 6.1: Optimal split between market orders and limit orders at the best buy quote for different values of ρ_u and different distributions for ξ.

Figure 6.2: The expected cost for one price-level as a function of the number of placed limit orders L_1 (where $M = S - L_1$) for $\rho_u = \{\underline{\rho_u}, 0.043, \overline{\rho_u}\}$.

We see that as the target quantity S increases, the number of limit orders to optimally place plateaus, where any extra shares the trader wishes to buy should be placed as market orders.

Two Price-Levels

Next, the optimal allocation considering two price-levels, described in section 4.2, is analysed. Although our solutions don't specify a certain distribution for the stochastic variables ξ and ψ_i, from now on we assume ξ is exponentially distributed and ψ_i is given by the queue size minus an exponential distribution, i.e. $\xi \sim \text{Expon}(\lambda_\xi)$ and $\psi_i \sim Q_{i+1} - \text{Expon}(\lambda_{\psi_i})$. This assumption on these distributions allow for analytical solutions for the 2 price-level case, as seen in section 4.2, and they are common in the literature (as in [32]). We set the target number of shares $S = 1000$ and the first two price queues at $Q_1 = 2000$ and $Q_2 = 1500$.

Figure 6.4 shows how the static optimal allocation goes from only MOs to a mix of MOs and L_1 orders, as $\bar{\xi}$ increases, until the optimal allocation consists of only L_1 orders(i.e. $L_1^\star = S$). As $\bar{\xi}$ continues to increase, we eventually reach a point where the optimal solution jumps from $L_1^\star = S$ and $L_2^\star = 0$ to $L_1^\star = 0$ and $L_2^\star = S$. The moment where this happens is, at the value described in Corollary 4.1. The standard parameter values $h = 0.02, r = 0.002, \rho_u = 0.05, \theta = 0.0005$ and $\delta = 0.01$, mean that $\lambda_{\psi_1} = \frac{1}{0.5 \cdot Q_2} = 1/750$. Plugging these values into Corollary 4.1, we find the critical value of λ_ξ to be $\lambda_\xi = 1/5437.5 = 1/\bar{\xi}$, which is indeed the value at which the optimal allocation switches in Figure 6.4.

These findings align well with our expectations. As $\bar{\xi}$ increases, the probability of L_1 limit orders being filled reaches such a high value, that this certainty is worth the trade-off with the uncertain limit orders at L_2 at a slightly better price. As $\bar{\xi}$ increases even more another tipping point is reached: the increase in certainty of execution for L_1 order plateaus, whereas the fill probability of L_2 orders grows, causing the optimal allocation to jump to $L_2^\star = S$.

To see what happens to the expected cost function around this critical tipping point, the expected cost function for this value $\bar{\xi}$ is plotted using 10,000 Monte Carlo simulations of ξ and ψ_1 in Figure 6.5. The plot highlights that when $\bar{\xi}$ is exactly at this tipping point the expected cost is almost the same for all splits between L_1 and L_2, when we set $M = 0$.

It is also important to see how the optimal solution behaves under changes in λ_{ψ_1}. Figure 6.6 shows the optimal split for $\bar{\xi} = 1/\lambda_\xi = 1/2800$ and different $\bar{\psi}_1 = 1/\lambda_{\psi_1}$, where the outflow from the second queue is given by $\psi_1 \sim Q_2 - \text{Expon}(\lambda_{\psi_1})$. We see that for low $\bar{\psi}_1$ (i.e. "high" order outflow at the second price-level) L_2 orders are always preferred to L_1 orders. This is according to our intuition, since high second order outflow means a small queue before the L_2 orders, this means that L_1 and L_2 orders have a very similar probability of being filled within the time frame, so due to the better price, L_2 orders are preferred. At some point, however, there is not enough outflow from the second queue, meaning L_2 have a significantly smaller chance of being filled. This sudden switch in optimal LOS price-level occurs at the tipping point described in Corollary 4.1, namely $1/\lambda_{\psi_1} = \frac{1}{\lambda_\xi} \frac{\delta}{h+r+\rho_u+\theta} \approx 386.2$.

Three Price-Levels

The optimal allocation considering 3 price-levels (described in section 4.3) shows similar results. Setting $Q_3 = 1500$ and $\lambda_{\psi_2} = 0.6 \cdot Q_3$, Figure 6.7 again shows the optimal static order allocation switching from L_2 orders to L_3 at the critical point $1/\lambda_\xi = \frac{h+\delta+r+\rho_u+\theta}{\lambda_{\psi_2}\delta} = 7425$ described in Corollary 4.2.

6.1.2 Optimal Strategy for n Price-Levels

We now proceed with the optimal allocation when considering any amount price-levels, as in section 4.4. Figure 6.8 shows the optimal order allocation when 8 price-levels are considered, for different mean

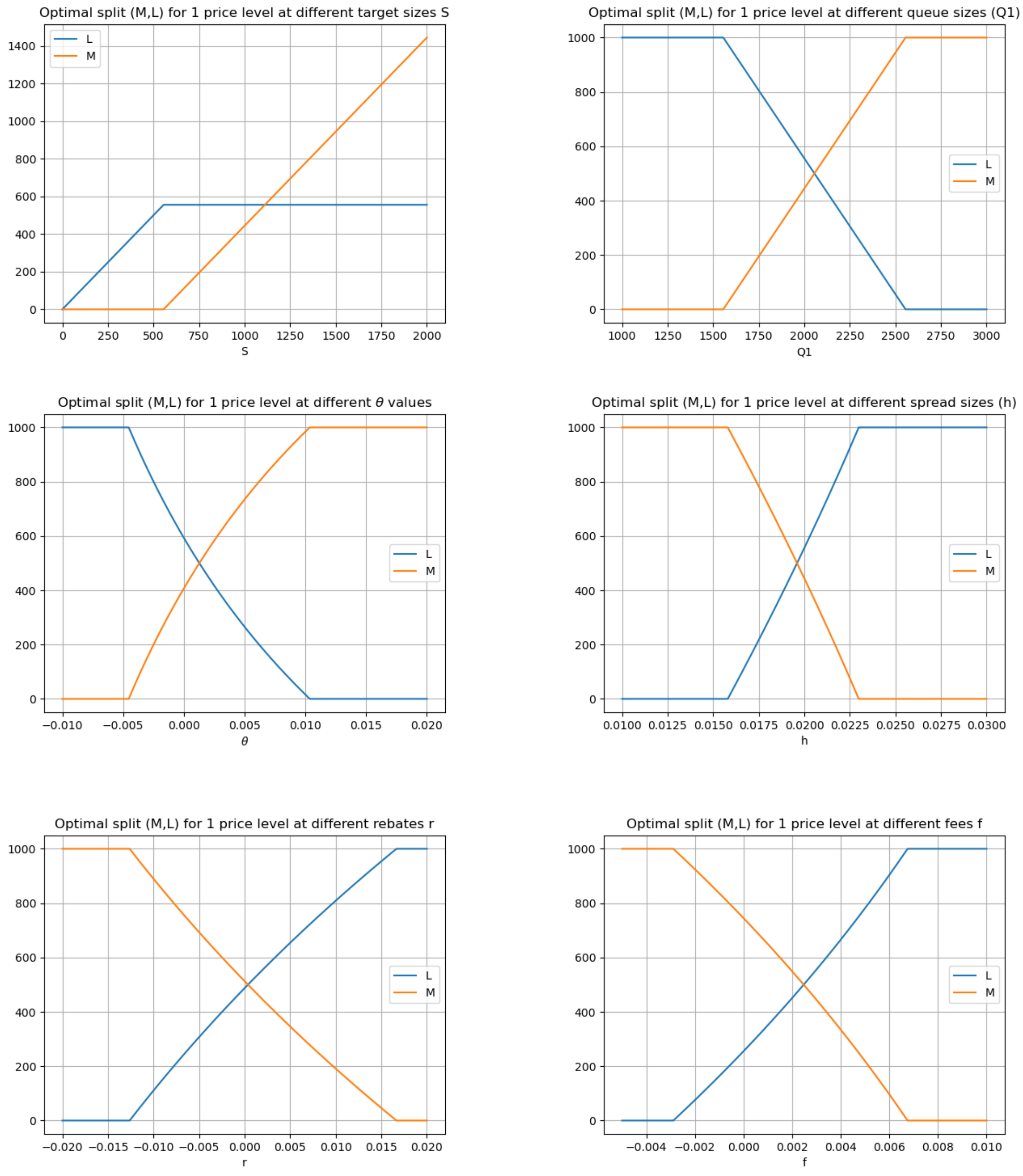

Figure 6.3: The optimal split for one price-level for different parameter values, for the parameters S, Q_1, θ, h, r and f.

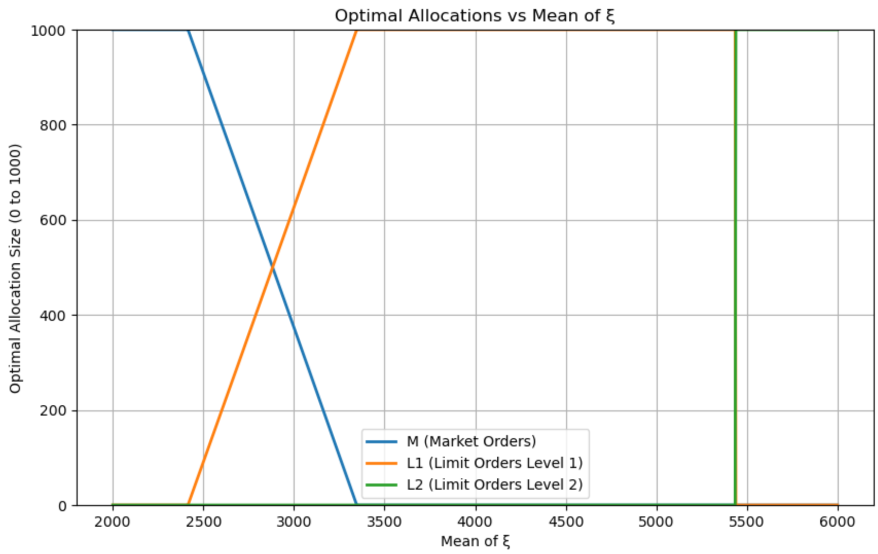

Figure 6.4: Optimal split between market orders and limit orders at the best and second-best buy quotes for different values of $\bar{\xi} = 1/\lambda_\xi$

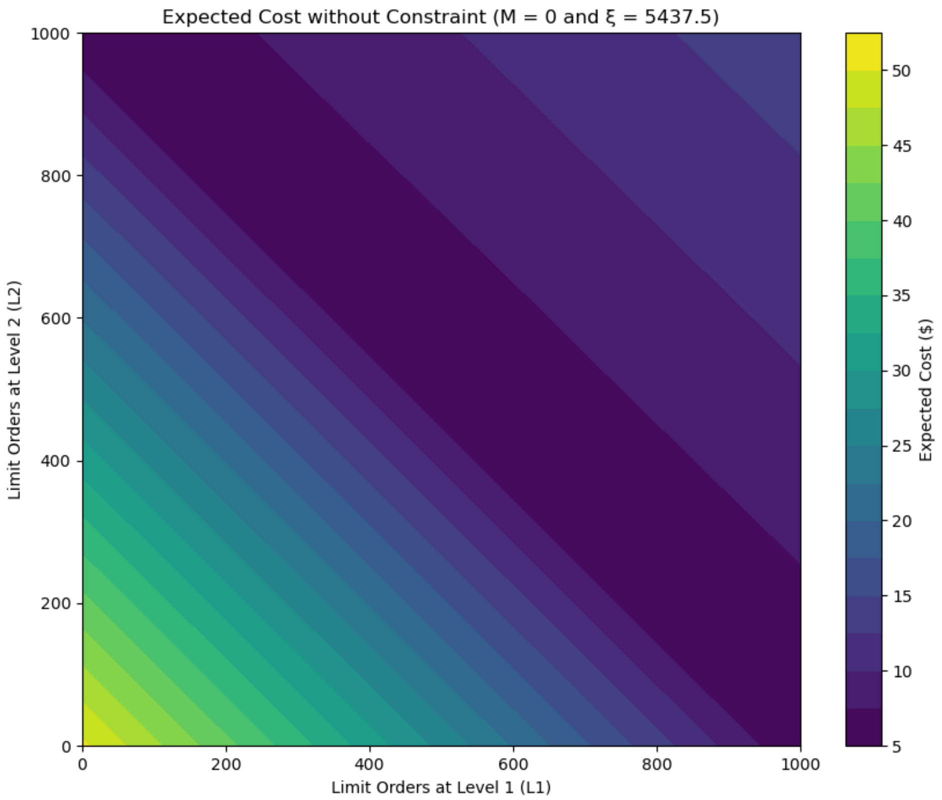

Figure 6.5: Expected cost function for $\bar{\xi} = 5437.5$ with L_1 and L_2 ranging from 0 to S

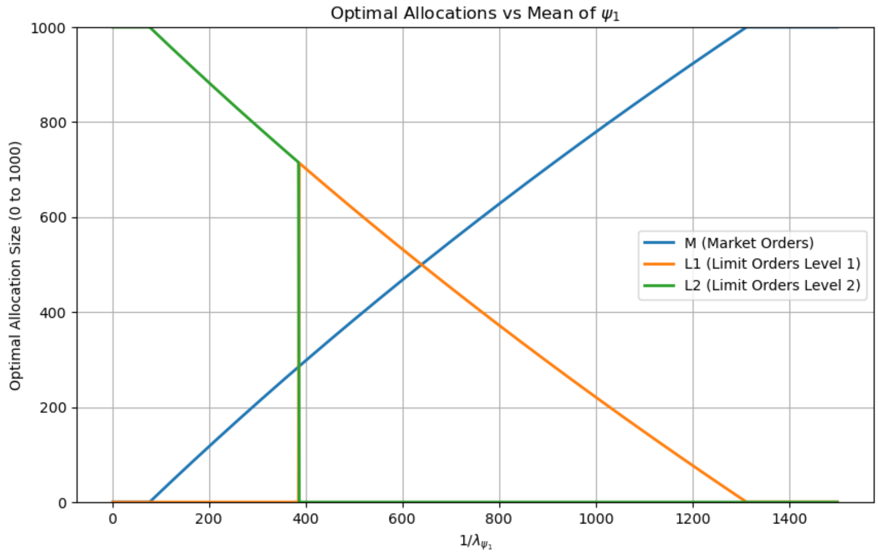

Figure 6.6: Optimal split between market orders and limit orders at the best and second-best buy quotes for different values of $\bar{\psi}_1 = 1/\lambda_{\psi_1}$

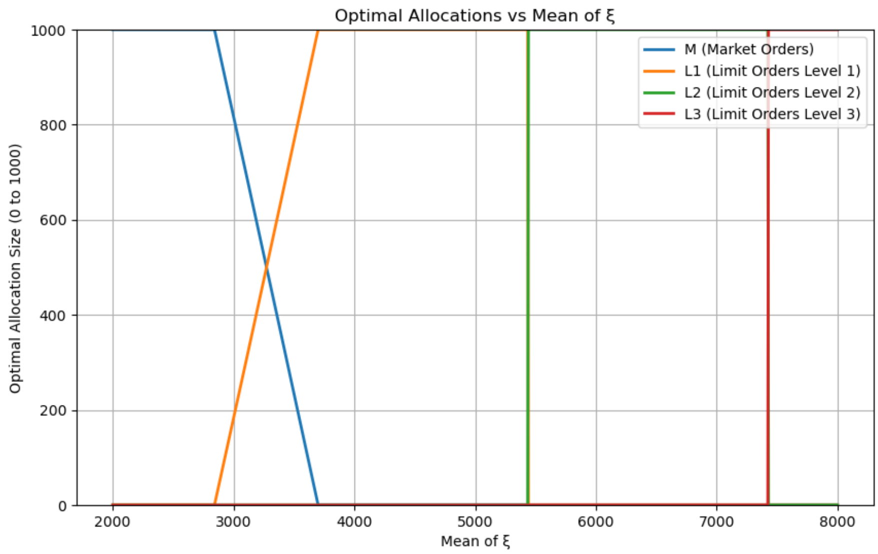

Figure 6.7: Optimal split between market orders and limit orders at the best three buy quotes for different values of $\bar{\xi} = 1/\lambda_{\xi}$

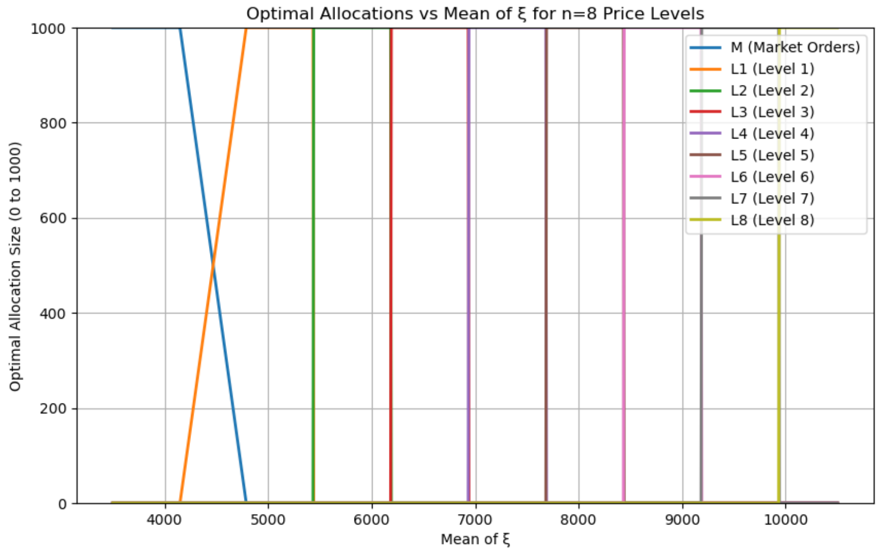

Figure 6.8: Optimal split between market orders and limit orders at the best eight buy quotes for different values of $\bar{\xi} = 1/\lambda_\xi$

arrival rates of ξ, $\bar{\xi} = 1/\lambda_\xi$. It is again assumed that $\xi \sim X$, where X is exponentially distributed with rate λ_ξ and $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are independent and exponentially distributed with rate λ_{ψ_i} and set $Q_2, ..., Q_9 = 1500$ and $1/\lambda_{\psi_i} = 0.5 \cdot Q_{i+1}$. Considering 8 price-levels, the optimal allocation is indeed as described in Corollary 4.3, with it jumping from S limit orders at price-level i to S LOs at price-level $i+1$ at the critical points described in (4.52). For a larger tick size δ, assuming the filling probabilities remain the same, we expect the optimal allocation to shift to deeper price-levels more quickly, since those orders will result in a higher reward with the same execution risk. Figure 6.9 confirms this expectation.

6.1.3 Cost Analysis

To see if our optimal order allocation strategy does indeed lead to lower expected cost, we compare it to simple benchmark strategies, that either place only MOs or only LOs at a certain price level and the one price-level strategy from Cont and Kukanov [2017] [15]. The average cost for each strategy is calculated using 200.000 Monte Carlo simulations, with the following parameter values: $h = 0.02$, $r = 0.002$, $f = 0.003$, $\delta = 0.01$, $\theta = 0.0005$, $\rho_u = 0.05$, $Q_1 = 2000$, $Q_2, Q_3, .. = 1500$ $S = 1000$, and $\bar{\xi} = 1/\lambda_\xi \in (0, 10.000)$ with $\xi \sim \text{Exp}(\lambda_\xi)$ and $\bar{\psi}_i = 1/\lambda_{\psi_i} = Q_{i+1}/2$ with $\psi_i \sim Q_{i+1} - \text{Exp}(\lambda_{\psi_i})$.

Figure 6.10 illustrates that for high values of $\bar{\xi}$ the optimal allocation leads to a lower expected cost than the benchmark strategies. For lower values of $\bar{\xi}$, however, the one price-level optimal strategy from [15] leads to a lower expected cost.

In Figure 6.11 we zoom into the graph for the values of $\bar{\xi}$ where the optimal allocation reaches the critical point where the optimal strategy jumps from all L_2 orders, to all L_3 orders. The figure shows the model accurately switches price-level at the point when the expected cost for this strategy starts to become lower than the previous strategy, out performing the strategies that only post L_2 or L_3 orders. Compared to the strategy from [15], however, our optimal strategy switches from L_1 orders to L_2 orders prematurely. Consequently, around $\bar{\xi} = 5450$ the one price-level strategy achieves slightly lower expected costs than the static optimal strategy.

The static optimal strategy performs well for higher values of $\bar{\xi}$, but the one price-level strategy

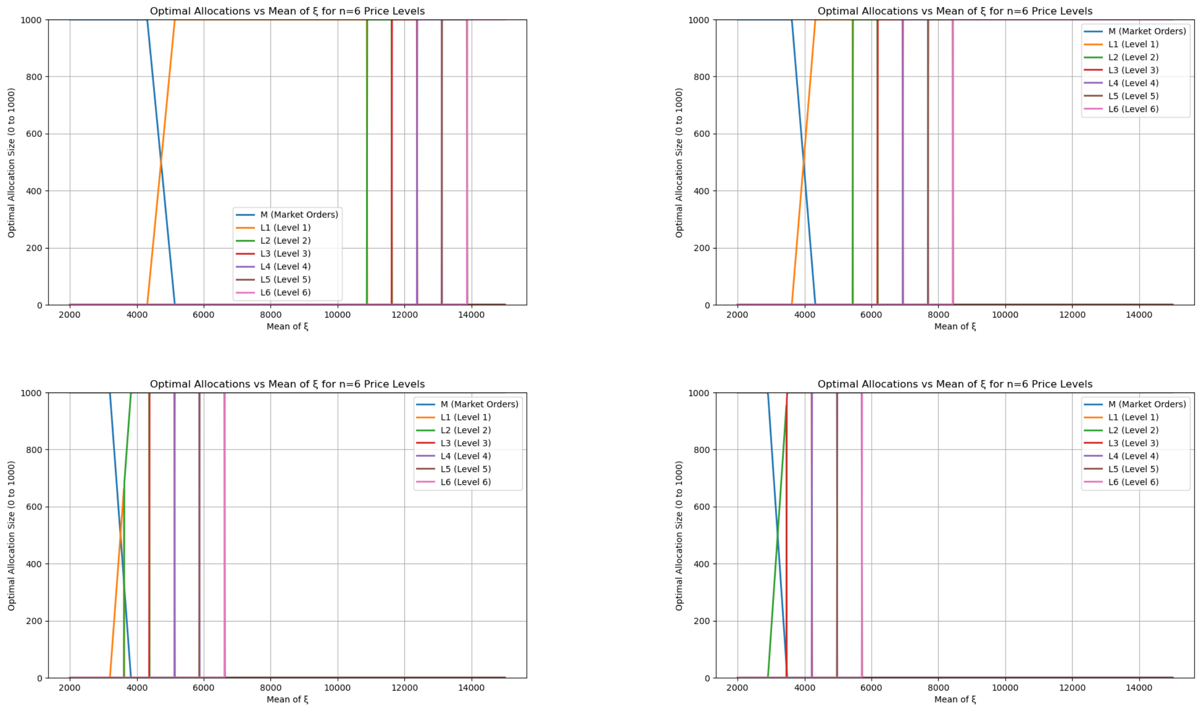

Figure 6.9: The optimal split for 6 price-levels for different different tick sizes $\delta = 0.005, 0.010, 0.015, 0.020$ from top left to bottom right.

from [15] performs better for lower values $\bar{\xi}$, where the optimal solution is to balance market orders and limit orders at the best price-level. To improve the performance of our strategy, we let our choice of the number of price levels to consider be guided by the critical points described in Corollary 4.3. If the value of $\bar{\xi}$ lies below the critical point for L_i, there is no need to consider that price-level, as our optimal allocation will not place any limit orders there. Adjusting the static optimal strategy accordingly we get the expected cost seen in Figure 6.12. It illustrates that our strategy indeed matches the performance of the one price-level strategy from [15]. Zooming in in Figure 6.13, it becomes evident that for higher values of $\bar{\xi}$, the adjusted optimal strategy jumps to the two price-level strategy and then the three price-level strategy, at the points where these strategies have lowest expected cost.

In the previous examples optimal allocation was only a mix of MOs and L_1 orders. It is crucial to examine how the static optimal strategy behaves when deeper price-levels enter this mix. Changing the spread size from 0.04 to 0.02, so $h = 0.02$ to $h = 0.01$, causes our optimal solution to behave completely differently. Figure 6.14 shows that for our 5 price-level solution it is no longer optimal to place L_1 orders at all, instead at a certain point the number of MOs starts to go down and the trader should start placing L_2 orders immediately. Figure 6.15 and Figure 6.16 show that the $n = 5$ price-level optimal allocation leads to slightly higher expected cost than the one price-level allocation (from [15]) when the optimal solution involves L_1 limit orders. Although it produces significantly lower costs at higher values of $\bar{\xi}$, we must use the critical points to determine the number of price-levels to consider for our static optimal order allocation.

This adjustment, using the critical points from Corollary 4.3 to decide how many price levels to consider, enables our optimal order allocation to achieve equal or lower average costs than the strategies posting LOs at a certain price-level and the strategy from [15] for all values of $\bar{\xi}$, as seen in Figure 6.17. Zooming in, Figure 6.18 highlights that our adjusted optimal solution is able to adapt to $\bar{\xi}$ and

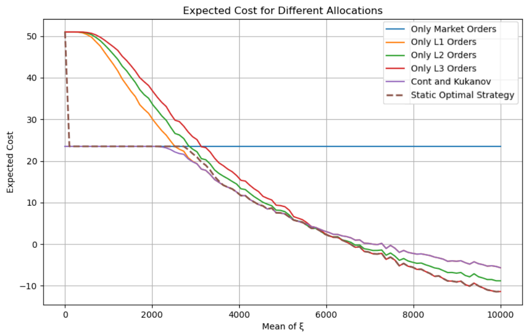

Figure 6.10: Expected costs of different allocation strategies for different values of $\bar{\xi} = 1/\lambda_\xi$

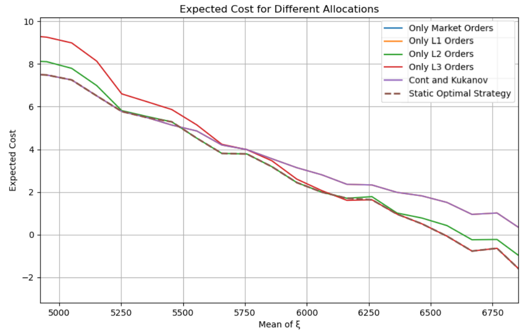

Figure 6.11: Expected costs of different allocation strategies for different values of $\bar{\xi} = 1/\lambda_\xi$, zoomed in where the optimal strategy switches from L_1 to L_2 to L_3 LOs

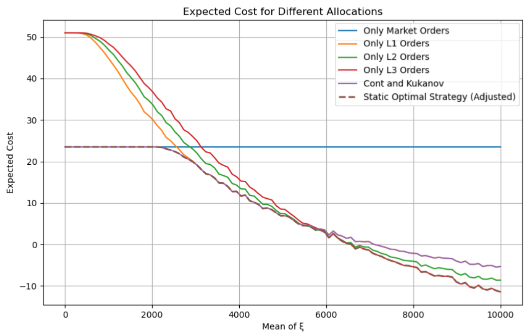

Figure 6.12: Expected costs of different allocation strategies compared to our adjusted order allocation for different values of $\bar{\xi} = 1/\lambda_\xi$

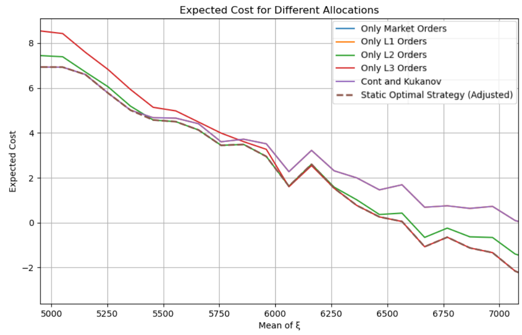

Figure 6.13: Expected costs of different allocation strategies compared to our adjusted order allocation for different values of $\bar{\xi} = 1/\lambda_\xi$, zoomed in where the optimal strategy switches from L_1 to L_2 to L_3 LOs

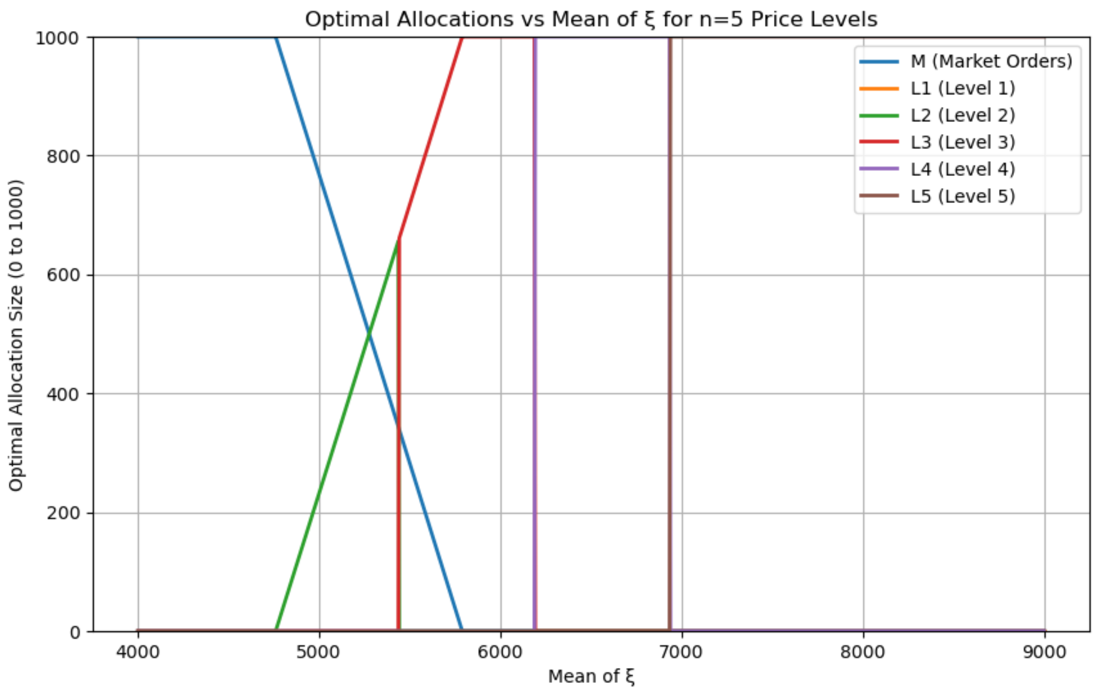

Figure 6.14: Optimal order allocation for 5 price-levels for different values of $\bar{\xi} = 1/\lambda_\xi$, with $h = 0.01$

change the number of price-levels it considers, to find the allocation that leads to the lowest expected cost in all cases.

To summarize, under the assumption that order flow is exponentially distributed our static optimal allocation provides an intuitive and analytically tractable trading strategy. The critical points in Corollary 4.3 allow for adjustment of the number of price-levels considered in our static optimal allocation, leading to lower expected cost under all order flow values when compared to benchmark strategies only considering MOs and LOs at the first price-level, or only LOs at any one price-level.

6.2 Dynamic Model

This section highlights the key results of the dynamic model in chapter 5. To allow for fair comparison the parameter values are set based on the values in the paper by Cartea and Jaimungal [2015] [5]: $p_0 = 60$, $\mu = 0, \sigma_p = 0.01$, the starting inventory $\mathcal{R} = 10$, $T = 60$ seconds, $\kappa = 100, \Delta = 0.01, \lambda_\nu = 50/\text{min}, \theta = 0.0001$, but also introducing a rebate for LOs and fee for MOs by setting $r = 0.003$ and $f = 0.002$.

First we will analyse the optimal depth at which a trader should place her limit orders, and then the optimal stopping times when a trader should place market orders. Lastly the average cost per share our strategy is able to realise is compared to the benchmark models that don't consider mid-price drift (as in [5]) or don't consider MOs (as in [20]), and the Time Weighted Average Price (TWAP).

6.2.1 Optimal Depth for Limit Orders

As mentioned in the beginning of Chapter 5, in the absence of a running inventory penalty and price drift, there is no incentive for the trader to place market orders until an instant before the terminal time, setting the optimal stopping time for all inventories to $t = T^-$. There is, however, an incentive

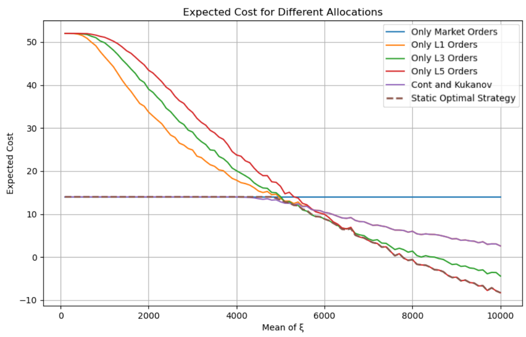

Figure 6.15: Expected costs of different allocation strategies for different values of $\bar{\xi} = 1/\lambda_\xi$, with $h = 0.01$

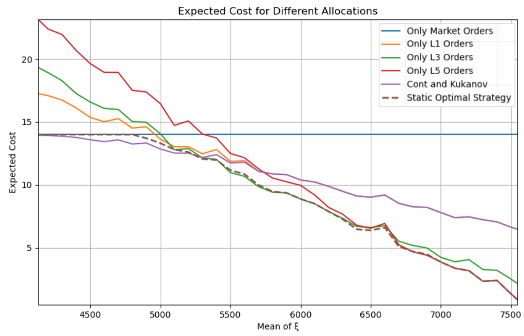

Figure 6.16: Expected costs of different allocation strategies for different values of $\bar{\xi} = 1/\lambda_\xi$, with $h = 0.01$, zoomed in where the optimal strategy switches from L_2 to L_3 to L_4 to L_5 LOs

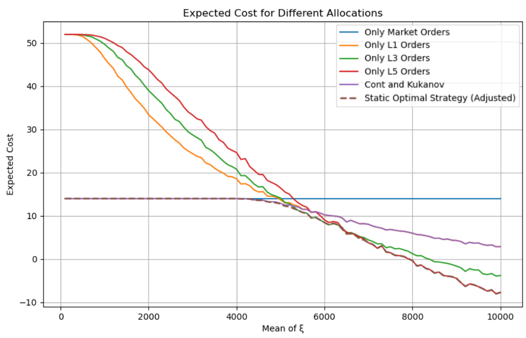

Figure 6.17: Expected costs of different allocation strategies compared to our adjusted order allocation for different values of $\bar{\xi} = 1/\lambda_\xi$, with $h = 0.01$

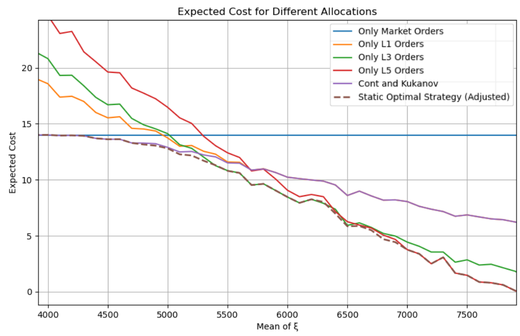

Figure 6.18: Expected costs of different allocation strategies compared to our adjusted order allocation for different values of $\bar{\xi} = 1/\lambda_\xi$, with $h = 0.01$, zoomed in where the optimal strategy switches from L_2 to L_3 to L_4 to L_5 LOs

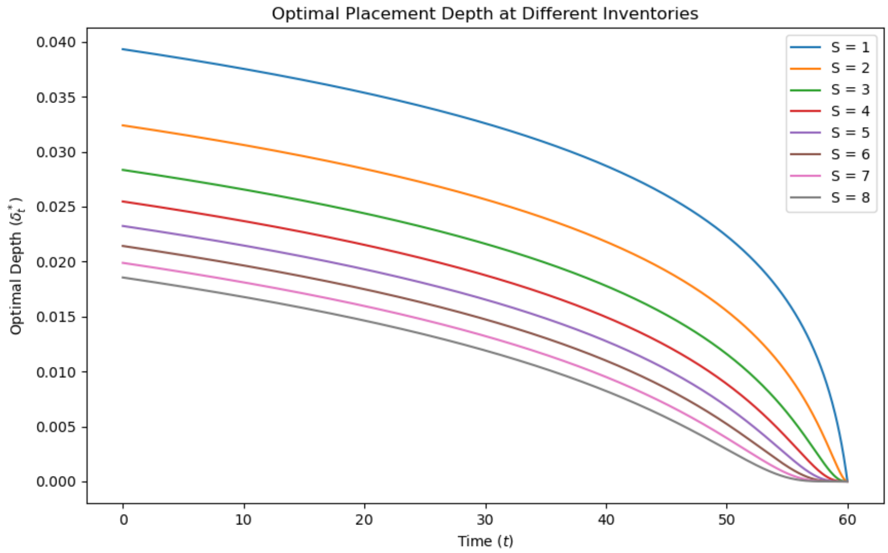

Figure 6.19: Optimal LO placement depth as function of inventory size without running inventory penalty or mid-price drift ($\theta = 0, \mu = 0$)

to liquidate her inventory in the form of limit orders before the terminal time. Figure 6.19 shows how the optimal depth for the trader's limit orders decreases as the trader's remaining inventory increases, showing an higher urgency to fill limit orders when the remaining inventory is higher, based on the solution from [20].

A crucial incentive for a trader to quickly liquidate her inventory is the running inventory penalty. Figures 6.20, 6.21, 6.22 and 6.23 illustrates that as the penalty term over the remaining inventory increases, the optimal depth at which the trader should place her limit order decreases, highlighting this sense of urgency. For certain inventory sizes, the optimal depth hits zero before the terminal time, this indicates the optimal strategy has determined a stopping time. In the case where $\theta = 0.0005$, the optimal depth is always at the touch ($\delta = 0$) for $i = 5$ and $i = 6$, since in those cases the stopping time is equal to 0.

Another important incentive for a trader to have more urgency in liquidating her inventory is when the mid-price is expected to go down. Figure 6.24 shows what happens when we introduce a drift term to the mid-price. In line with our intuition, the higher the drift term μ (i.e. the more the price will go up), the less urgent the trader is to liquidate her portfolio quickly, and the more she can afford to place her limit orders deeper into the order book. All-in-all our dynamic optimal strategy is effective at adapting the depth at which it places limit order to (un)favourable price drifts and running inventory penalty.

6.2.2 Optimal Stopping Times for Market Orders

When a trader deems it suboptimal to wait for her limit orders to be filled (at any depth), she chooses to place market orders to liquidate (part of) her inventory. When the penalty on her running inventory is low, or the mid-price is expected to move up, she has less urgency in liquidating her inventory quickly, so we expect she can afford to keep a larger inventory for a longer period of time. In Figure 6.25 shows that for a lower running inventory penalty term, the optimal stopping time (at which to place a market order) when the trader holds an inventory of size 10, is at slightly over 6 seconds before the terminal time. As this penalty term is increased, however, it becomes optimal to start placing

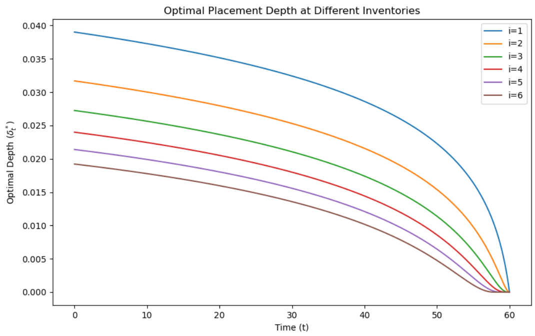

Figure 6.20: Optimal LO placement depth as function of inventory size at $\theta = 0.00001$

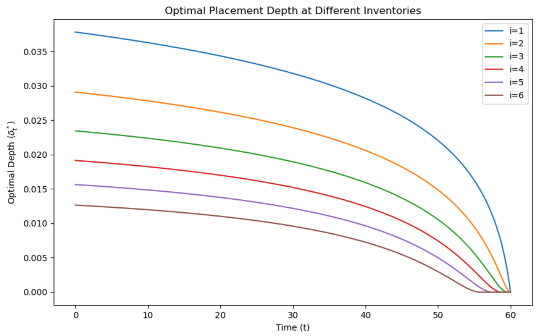

Figure 6.21: Optimal LO placement depth as function of inventory size at $\theta = 0.00005$

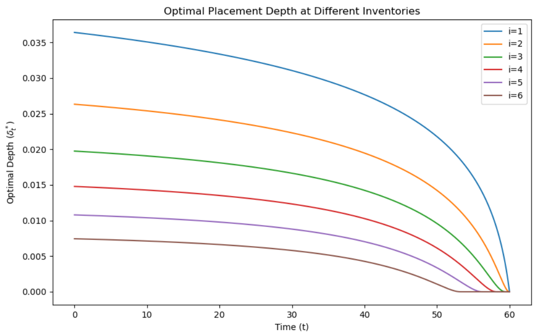

Figure 6.22: Optimal LO placement depth as function of inventory size at $\theta = 0.0001$

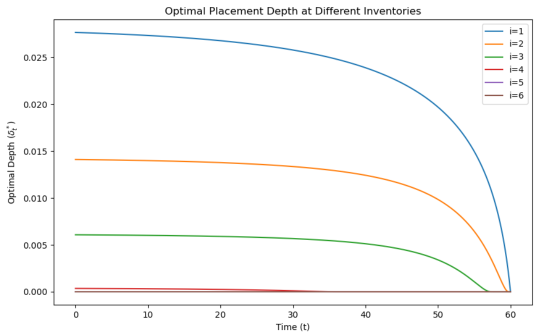

Figure 6.23: Optimal LO placement depth as function of inventory size at $\theta = 0.0005$

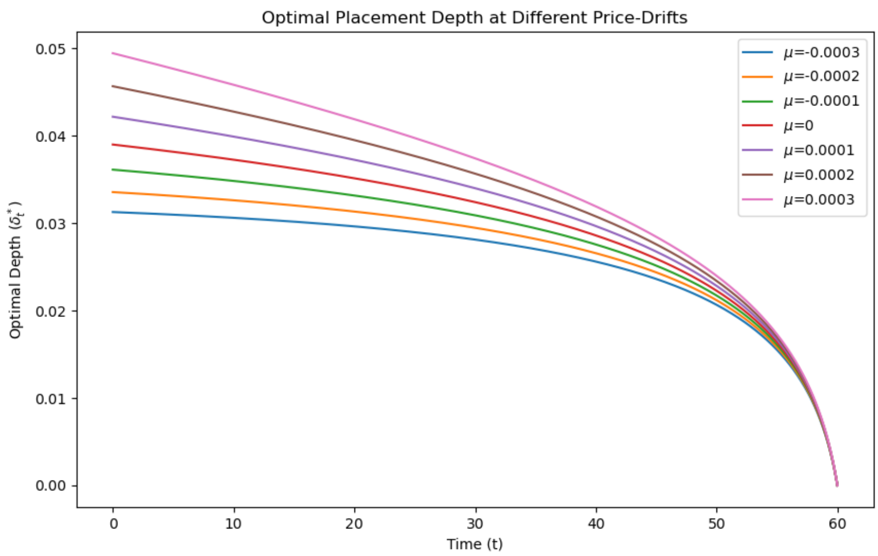

Figure 6.24: Optimal LO placement depth for inventory of size 1 and different mid-price drifts μ with $\theta = 0.00001$

market orders immediately (at $t = 0$) and continue with an inventory of size 8 (for $\theta = 0.0002$) or size 5 (for $\theta = 0.0005$).

The reason we will find most strategies execute several MOs at the beginning and then post limit orders at decreasing depth is logical: the depth optimizes so that we execute our limit orders with a good probability and hopefully at attractive prices, but the MO execution strategy tells us what is a realistic inventory size to start with so that the trader is able to liquidate them all as limit orders with high probability.

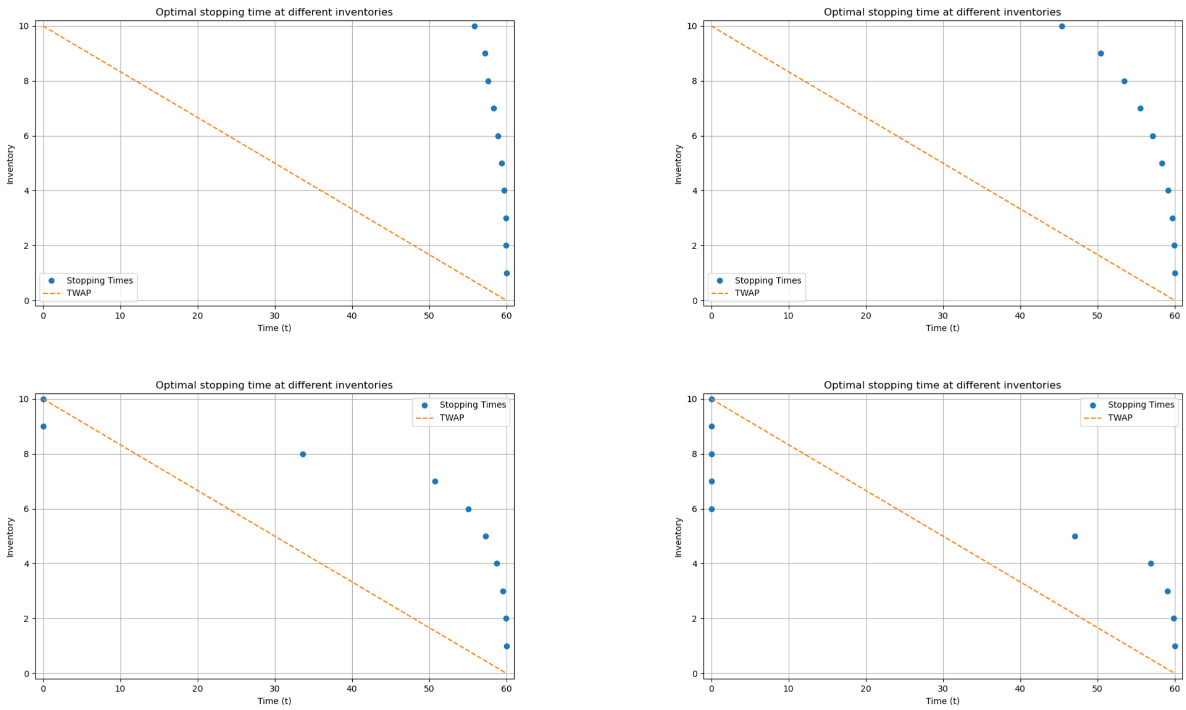

Figure 6.25: The optimal stopping time for market orders for different running inventory penalty values, $\theta = 0.00001, 0.0001, 0.0002, 0.0005$ from top left to bottom right, with the TWAP as comparison

As mentioned in the previous section, another parameter driving a trader's incentive to place market orders is price drift. We saw a negative price drift, makes the trader more eager to liquidate her inventory, causing the optimal depth to be lower. Figure 6.26 illustrates the same is true for stopping times, with negative drift leading to faster stopping times. Next, in Figure 6.27 we see the exact opposite holds for positive μ values. As the drift in mid-price increase, a trading hoping to optimally liquidate her inventory should wait to place market orders longer and longer. This is logical, since the trader wishes to sell the shares in her inventory when the price is as high as possible, but with enough time left to allow her to sell them as limit orders before the terminal time $t = T$. The reason we chose $\mu = 0, 0.0008, 0.00018, 0.00028, 0.00038, 0.0048, 0.0058$ is that whole values like $\mu = 0.001$ results in zeros in the denominator of the formula for $g_S(t)$, resulting in errors. All-in-all our dynamic optimal strategy is effective at adapting to (un)favourable price drifts and running inventory penalty through the timing of market orders.

6.2.3 Cost Analysis

To see if our optimal allocation strategy actually leads to cost savings, we compare it to benchmark strategies. The first benchmark strategy is the Time Weighted Average Price (TWAP) strategy, which

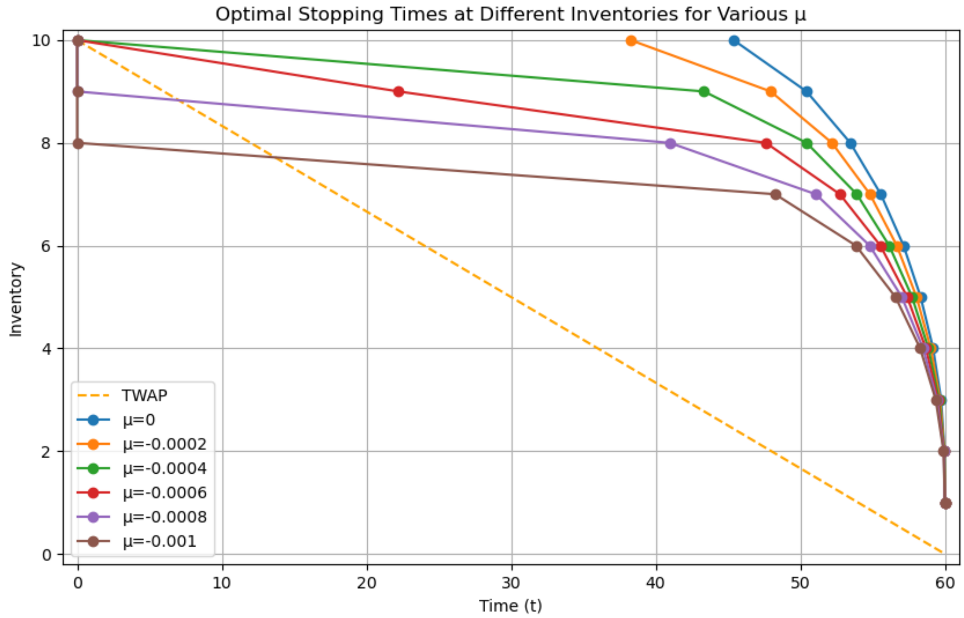

Figure 6.26: Optimal stopping time for different price drifts $\mu = 0, -0.0002, -0.0004, -0.0006, -0.0008, -0.001$ with $\theta = 0.0001$

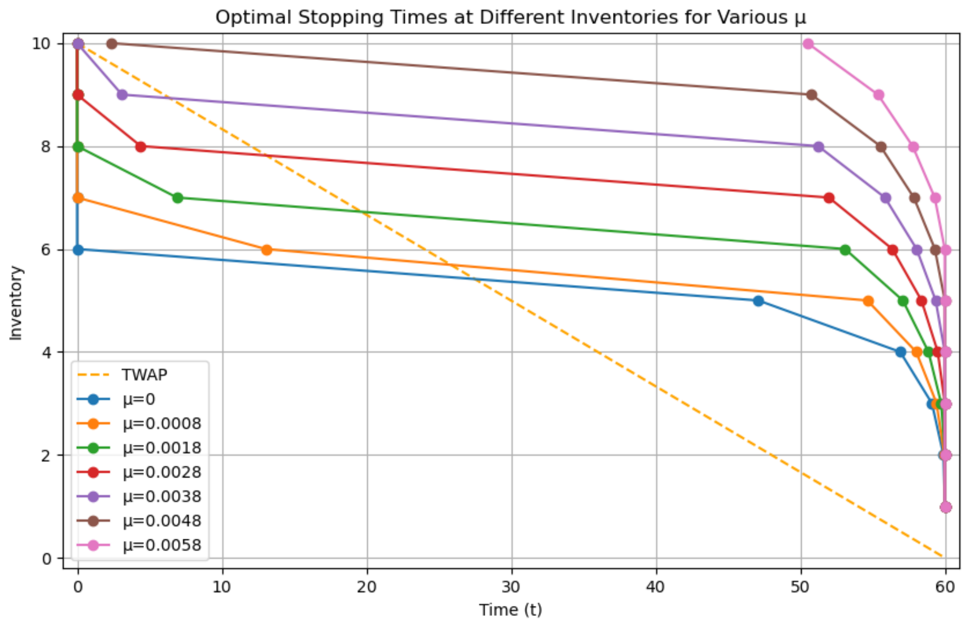

Figure 6.27: Optimal stopping time for different price drifts $\mu = 0, 0.0008, 0.00018, 0.00028, 0.00038, 0.0048, 0.0058$ with $\theta = 0.0005$

places only market orders, spread out equally over the time interval. This strategy relies on the certainty of market orders being filled and spreads them out to avoid exerting impact on the mid-price. The second benchmark strategy we compare to our solution was proposed by Cartea and Jaimungal [2015] [5], who used the same optimal stopping and control model, but did not incorporate mid-price drift. Finally our strategy is compared to the benchmark strategy from [20], which does consider mid-price drift, but doesn't allow the trader to post MOs, only optimising for the optimal LO depth.

The average earnings per share of the strategies is compared based on 10,000 Monte Carlo simulations of the mid-price. For these simulations we set the following parameter values: $p_0 = 60$ and $\mu = -0.001, \sigma_p = 0.01$, the starting inventory $\mathcal{R} = 10$, $T = 60$ seconds, $\kappa = 100, \Delta = 0.01, \lambda_\nu = 50/\text{min}, \theta = 0.0001, r = 0.003, f = 0.002$.

Optimal Strategy vs. TWAP

Figure 6.28 shows three Monte Carlo price simulations for $\mu = -0.001$ and $\sigma_p = 0.01$. Although the drift term is negative, due to the stochastic nature of Brownian Motion we see that the price does not necessarily go down (simulation 2). Figure 6.29 shows the depth at which the trader must place her limit orders according to the optimal strategy. The figure highlights that as her inventory size decreases (due to her limit orders being filled) she can afford to place her limit orders deeper into the book. Next, in Figure 6.30, it is evident that the trader's inventory gradually decrease over time in all 3 simulations, with the average liquidation scheme being relatively close to the TWAP scheme. Furthermore, for the second and third simulation the trader was not able to liquidate her entire inventory using limit orders according to the optimal strategy, requiring her to place a market order at stopping time $t = \tau_2$ (for simulation 2) and a market order at time $t = T$ (for simulation 2 and 3) to liquidate the last share in her inventory. Finally, the average price the trader receives for the shares she has sold is compared in Figure 6.31. It shows that the optimal strategy achieves higher average prices (which means higher earnings for the trader) than the TWAP in all three simulations. This is to be expected, since the optimal strategy allows the placement (and therefore the filling) of LOs, which are always favourable above MOs. What makes the result interesting to see, however, is the liquidation scheme that the inventory follows over time. It resembles that of a TWAP strategy, meaning the optimal strategy also benefits from avoiding mid-price impact (like TWAP) while achieving a more favourable average price per share.

Optimal Strategy With Drift vs. No-Drift Strategy

Now our dynamic optimal strategy as described in Proposition 5.1 is compared to the strategy proposed by Cartea and Jaimungal [2015][5]. Figure 6.32 again shows see 3 Monte Carlo price simulations for $\mu = -0.001$ and $\sigma_p = 0.01$ and Figure 6.33 shows the depth at which the trader must place her limit orders according to the optimal strategy. Evidently the trader following the strategy that does not consider mid-price drift places her limit orders deeper into the book than our optimal strategy, as the no-drift optimal solution does not take into account the effect of the negative price drift. From the liquidation scheme, in Figure 6.34, we see that the trader's inventory gradually decrease over time in all 3 simulations, under both strategies. For the strategy that doesn't consider the mid-price drift, the orders are filled later than those under our optimal strategies, with two of the no-drift simulations requiring the placement of 1 or 2 MOs at terminal time $t = T$ to liquidate the remaining inventory (compared to only 1 simulation of our optimal strategy requiring 1 MO at $t = T$). This untimely filling of LOs is reflected in the average price per share, as seen in Figure 6.35. Upon closer examination, the average price for simulation 1 and 3 of the no-drift strategy decreases sharply at the terminal time, because in those simulations the trader is forced to place MOs at $t = T$ to fill her remaining inventory. To see if (and when) our dynamic optimal strategy does lead to a trader selling her shares at a higher average price per share, our strategy is executed on 10,000 Monte Carlo simulations of the mid-price.

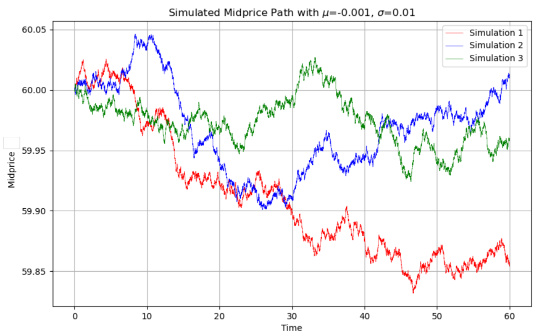

Figure 6.28: Price simulations for $\mu = -0.001$ and $\sigma_p = 0.01$

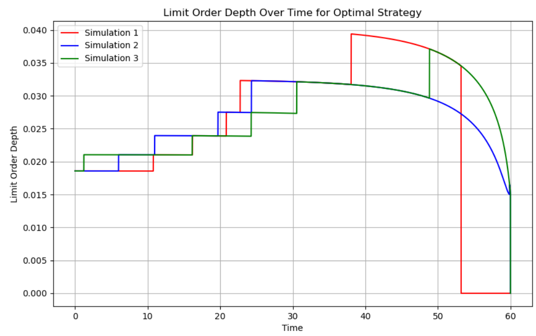

Figure 6.29: Optimal LO placement depth over time for all 3 simulations

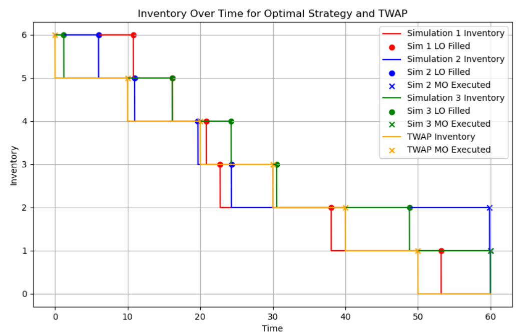

Figure 6.30: Inventory over time for all 3 simulations and the TWAP schedule, indicating when LOs were filled and MOs were executed

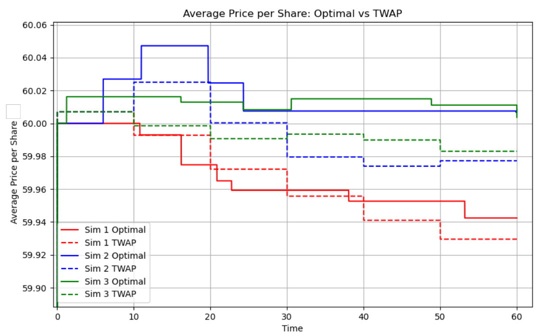

Figure 6.31: Average price paid per share in every simulation with the dynamic optimal strategy vs. the TWAP schedule

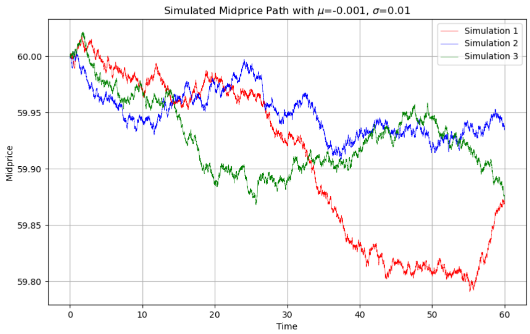

Figure 6.32: Price simulations for $\mu = -0.001$ and $\sigma_p = 0.01$

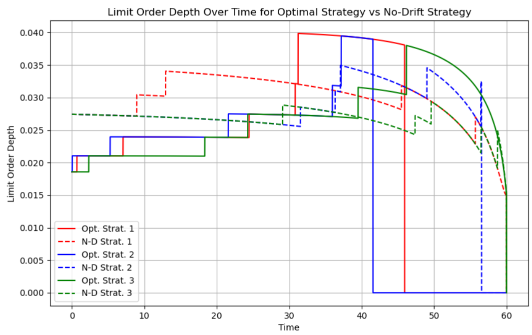

Figure 6.33: Optimal LO placement depth over time for the optimal strategy with drift and the no-drift strategy, for all 3 simulations

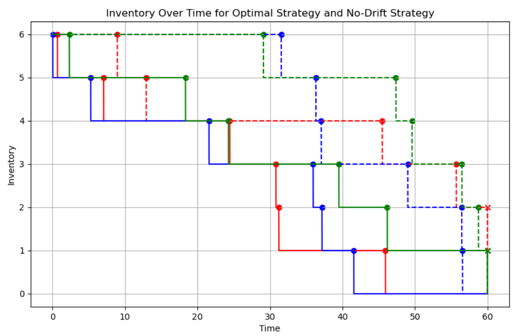

Figure 6.34: Inventory over time for all 3 simulations of the optimal strategy with drift (solid lines) and the no-drift strategy (dashed lines), indicating when LOs were filled (O) and MOs were executed (X)

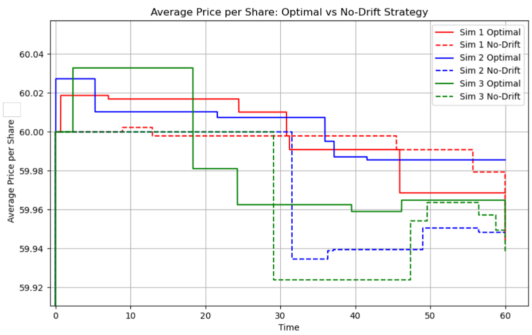

Figure 6.35: Average price paid per share in every simulation with the optimal strategy with drift vs. the no-drift strategy

For each price path we first apply our optimal strategy and then the strategy not considering drift, and find the average price paid per share. Table 6.1 shows the average price per share earned using each strategy, for different mid-price drift values and across 10,000 price simulations, with the trader holding a starting inventory of size $i = 5$. As per our expectations, in the absence of drift both strategies perform the same, earning slightly above the mid-price ($p_0 = 60$) on average. For negative drift, we see the more drift there is, the more the new optimal strategy is able to earn compared to the no-drift strategy. For $\mu = -0.003$, our optimal strategy tells us that any inventory above size 4 should be sold as market orders at time $t = 0$, and this adjustment leads to an extra earnings of 0.086 per share. For positive price drift, our optimal strategy still leads to a higher earnings per share, although the difference between both strategies is not as big as for negative drift. We find that for smaller/larger starting inventories the savings per share are similar. All-in-all, our dynamic optimal strategy is effective at adapting to positive and negative price drift, leading to higher average earnings per compared to dynamic strategies that don't consider mid-price drift.

Mid-price drift (μ)	**No-Drift Strategy**	**Strategy With Drift**	**Extra earnings per share**
0.003	60.152	60.190	0.038
0.002	60.110	60.131	0.021
0.001	60.070	60.075	0.005
0	60.029	60.029	0
-0.001	59.988	60.004	0.016
-0.002	59.945	59.993	0.048
-0.003	59.904	59.990	0.086

Table 6.1: Average price earned per share using the no-drift strategy and our dynamic optimal strategy (with drift) across 10,000 simulations, for different mid-price drifts

Optimal Strategy With Limit and Market Orders vs. Only Limit Orders

Finally we compare our found optimal strategy to the one from [20], which also produces a dynamic trading strategy considering mid-price movement, but only considers limit orders at the the optimal depth. Furthermore in [20] no running inventory penalty is considered, so we set $\theta = 0$. Our interest solely lie where our optimal strategy says the trader should execute market orders, since that is where the strategies differ. For non-negative price drifts and low running inventory penalty (i.e. $\theta = 0$) the stopping times are all close to the terminal time as can be seen in Figure 6.26 and 6.27, so mid-price drifts are set at $\mu = 0.001, -0.001, -0.002, -0.003, -0.004, -0.005$ and inventory the trader seeks to liquidate is set at size $i = 6$. For $\mu = -0.001$, we have stopping time $\tau_6 = 56$ seconds, and for $\mu = -0.005$ our optimal strategy tells the trader to liquidate any inventory size above $i = 2$ (i.e. $\tau_3, \tau_4, ... = 0$) as market orders immediately and only place 2 orders as limit orders. Lastly, to make a fair comparison we set $\rho_u = 0$, since our optimal strategy, which can place market orders, will never pay an extra fee at the terminal, but instead choose to execute its remaining inventory as market orders and instant before the terminal time T, as mentioned in the beginning of Chapter 5. Table 6.2 highlights that considering market orders does in fact lead to higher average earnings per share. For a positive price drift ($\mu = 0.001$) the stopping times are all equal to the terminal time ($\tau_1, ..., \tau_6 = T$), so our optimal strategy is exactly the same as the optimal strategy only considering limit orders. As soon as our strategy starts to set stopping times below T, however, we see that across the 10,000 simulations our optimal strategy results in higher earnings per share, with the extra earnings growing as our found stopping times are closer to the initial time $t = 0$. All-in-all the inclusion of market orders in our dynamic optimal strategy leads to higher average earnings per share compared to strategies that only consider limit orders, with the extra earnings highest when the trader's urgency to liquidate her inventory is higher (e.g. for negative price-drift).

Mid-price drift (μ)	**Only Limit Orders**	**LO and MO Strategy**	**Extra earnings per share**
0.001	60.0686	60.0686	0
-0.001	60.0082	60.0083	0.0001
-0.002	60.0013	60.0014	0.0001
-0.003	59.9966	59.9980	0.0014
-0.004	59.9920	59.9963	0.0043
-0.005	59.9878	59.9954	0.0076

Table 6.2: Average price earned per share using the limit-order-only strategy and our dynamic optimal strategy (placing LOs and MOs) across 10,000 simulations, for different mid-price drifts

7 | Conclusion and Discussion

Algorithmic trading has become an indispensable part of electronic trading systems, with all types of market participants using it to execute trades at the most favourable prices. At the heart of trade execution lies the Optimal Placement Problem (OPP): a trader looking to buy or sell a target quantity of shares must decide what types of orders to place (market or limit) at what time and what price-level. This thesis aims to provide a general analytical solution to this OPP, considering both market and limit orders, at all price levels.

To achieve this we considered two models, which we extended to include market orders and limit orders at all price levels, consider any target quantity of shares and possible drifts in the mid-price. The first static model formulated the OPP as a static convex optimisation problem, where a trader must place her orders optimally at the initial time to minimise expected costs. This model created an intuitive understanding of the effect of order book events on the optimal solution to the OPP and allowed us to derive a tractable analytical optimal solution which is widely applicable, not just in the case where only limit orders at the best buy/sell price-level are considered.

Although the static model was widely applicable, its solution is not practical when order flow and other LOB parameters are volatile. Therefore, we implemented a second model, which formulated the OPP as a dynamic stochastic optimal stopping and control problem, where a trader optimally adapts their placement strategy over time in response to evolving market conditions. This model allowed for a dynamic optimal strategy that we were able to express analytically for any inventory (target quantity) and considering possible drift in the mid-price, allowing for an improvement in adapting the trader's strategy to (unfavourable) price drift.

In the Chapter 6 we presented a detailed numerical assessment of both the static and dynamic formulations of the OPP. In this chapter, we summarise the key findings, highlighting the most promising results offered by the static and dynamic model solutions. We then compare the two models, analysing how they relate and their respective pros and cons. Lastly, we propose additional analyses that could be further investigated in future research and end with some final remarks.

7.1 Promising Results

7.1.1 The Static Model

The static model was first proposed by Cont and Kukanov [2017] [15], but only considered market orders and limit orders at the best buy quote. We were able to extend the optimal solution for 1 price-level to 2, 3, and eventually n price-levels. When analysing this general solution, we saw the optimal strategy behaved intuitively, with the optimal allocation shifting to price-levels deeper into the book, as the number of incoming market orders increased. We saw similar intuitive behaviour of our general optimal strategy in cases where the number of incoming limit orders or the tick size was changed, indicating correctness and an ability to translate changing LOB conditions into a new optimal order allocation.

Next, we assumed the order outflows at each price-level to be exponentially distributed. This allowed us to derive an analytical expression for the solution for n price-levels. We found the static optimal order allocation to jump from limit orders at one price-level to the next at critical points, which can be described explicitly in terms of the arrival rates of market orders and limit orders after time $t = 0$. These critical points balance the marginal cost (of not filling the deeper LO) with the marginal gain (if the deeper LO is filled). At these critical tipping points, we find the expected cost for a strategy of placing limit orders at a certain level or limit orders at one price level lower to be exactly the same, confirming the intuition of our found solution and indicating robustness. The fact that these transitions happen at analytically tractable thresholds is a promising result, because it offers traders a simple rule of thumb for when to switch price-levels based solely on observable queue depletion rates.

When compared to benchmark strategies, that place only market orders, only limit orders or adhere to the 1 price-level solution proposed by Cont and Kukanov [2017] [15], we see our optimal allocation for n price levels by by itself does not achieve lower expected cost in cases where the optimal solution is to combine market orders with limit orders at the first (few) price-level(s). However, when we adjust our optimal allocation to only consider the number of price-levels that are of interest to the optimal solution, using the critical points, we see the adjusted optimal allocation is able to outperform the before mentioned benchmark strategies in every case, achieving a lower expected cost and being much more widely applicable. This makes the optimal solution considering n price-levels, combined with the critical values we found more effective at allowing the trader to acquire any amount of shares at the best price in a wider variety of cases.

7.1.2 The Dynamic Model

The dynamic model was based on a paper by Cartea and Jaimungal [2015] [5], which considered both market and limit orders at any depth. In [5] they derived an analytical solution for a trader looking to liquidate an inventory of size 1 or 2, but did not take into account possible drifts in the mid-price. We were able to derive a recursive formula describing the optimal strategy, for any inventory size S, that took into account important parameters like running inventory penalty and mid-price drift. This strategy consisted of two parts: the optimal depth at which to place limit orders and the optimal stopping times at which to place market orders. The dynamic optimal strategy proved to behave very intuitively for different inventory sizes, running inventory penalties and drifts in the mid-price, validating its ability to create a sense of urgency for a trader looking to liquidate her portfolio, when the mid-price drift is negative or the running inventory penalty is large.

We find that the optimal strategy achieves higher average prices (which means higher profits for the trader) than the Time Weighted Average Price across all three mid-price simulations. This is to be expected, since the optimal strategy allows the placement (and therefore the filling) of limit orders, which are always favourable above market orders. What makes the result interesting to see however, is the liquidation scheme which the inventory follows over time, which is similar to that of the TWAP, which means the optimal strategy also has the advantage of avoiding mid-price impact that the TWAP scheme has, at a more favourable average price per share.

When compared to the solution by Cartea and Jaimungal [2015] [5] that does not consider mid-price drift, we find our optimal strategy is also able to consistently generate higher sell prices for the trader. In the presence of negative mid-price drift, our model is effective at adapting the trader's optimal strategy to post market orders sooner and limit orders at lower depths, prioritising a timely liquidation of her inventory. This leads to a higher average price per share, across 10,000 simulations. For positive mid-price drift, our optimal strategy tells the trader to post deeper limit orders, as market orders filled at the terminal time trade at favourable prices, also leading to higher earnings per share compared to the solution not considering mid-price drift.

Finally, we compared to the solution in [20] that only considers the optimal depth to place limit orders, and doesn't allows the trader to place market orders. We see that when our strategy tells us it is optimal to place market orders before the terminal time, in the presence of negative price drift for example, it leads to higher average earnings per share across 10,000 simulations. The earlier our optimal stopping times are, the higher the average earnings per share for our trader, highlighting the importance of the consideration of market orders in this formulation of the OPP.

7.2 Comparing the Static and Dynamic Model

The key difference between the two models is that the static model did not allow for adjustments to the strategy mid-trajectory. As a result, if the market or limit order arrivals differ from the expected distribution, the static strategy can result in significant extra costs. By contrast, the dynamic model allows a trader to re-optimise and adjust their strategy at intermediate times and correct for discrepancies between expected and actual LOB events. As seen in the numerical experiments, the dynamic strategy delays market orders when the mid-price drifts favourably, or accelerates MO usage when inventory penalties or negative drifts increase. This flexibility consistently leads to lower realised costs in environments where order flow is more volatile or drift is non-zero.

The static model seems to more accurately represent the true workings and structure of a limit order book. The dynamic model estimates fill probability of an LO at a certain depth based on a pre-determined distribution and a constant MO arrival rate. The static model, on the other hand, captures the queueing system underlying this fill probability and accurately tries to capture the three events that completely determine the dynamics and evolution of a real-world LOB: the arrival of MOs, the arrival of LOs and the cancellation of LOs.

For larger inventories and longer time windows the static model is a better fit. The recursive nature of the dynamic solution makes computations for large inventories over longer time windows more cumbersome, which means you lose the ability to quickly react and adapt to changing market conditions. In a high-frequency environment, where every microseconds counts, this speed is crucial for an effective trading algorithm. It is more effective at liquidating small inventories in short time frames at an optimal price. The static model works especially well for large inventories and long periods of time, as this means more market orders arriving, for which we have seen the static optimal solution to be robust.

The static model's optimal split is highly sensitive to the under-fill penalty coefficient. If this coefficient is miss-estimated, placing too many LOs (which the static solution might do) can force costly, last-minute MOs. In the dynamic model, the trader can wait to see whether LOs actually fill and only place MOs when necessary. When the mid-price exhibits drift, the static model cannot adapt to changing expectations of price movement. In contrast, the dynamic model explicitly incorporates drift in its value function, allowing it to "lean into" favourable drifts (placing deeper LOs) or pre-empt unfavourable drifts (placing earlier MOs). Because the dynamic strategy's stopping times and LO depths adjust continuously to this mid-price drift, it is able to consistently ensure timely execution, avoiding having to place MOs at the terminal time.

To conclude, whereas the static model offers an intuitive solution to the OPP, the dynamic model is better at playing into changing LOB conditions (such as deviation in arrival rates or mid-price drift), allowing for re-evaluation and re-optimisation of the trader's strategy at intermediary time steps. At larger inventory levels, however, the recursive nature of the solution for the dynamic model leads to longer computational time, which can be a major drawback.

7.3 Discussion

While our static and dynamic models yield clear insights and improvements to benchmark order placement strategies, several limitations and opportunities for further research remain.

The static model currently outperforms the studied benchmark models, when the arrival rate of market orders and limit orders is assumed to be exponentially distributed. Although the literature proves this to be a reasonable assumption, interesting results might appear when these arrivals are described differently, such as (Compound) Hawkes processes.

Secondly, the static model does not consider the order in which MOs and LOs arriving during the time window. This might miss occurrences like the arrival of a large MO "walking up the book" before new LOs are posted later in the time window. In reality, such a large number of incoming MOs could potentially fill deeper LOs. A more realistic model could incorporate the order in which MOs and LOs arrive, looking at the intensity of the arrivals.

Moreover, we show that under economically reasonable assumptions, the optimal static solution consists of placing as many orders in total as the trader's target quantity. Then the optimal static solution is found by taking the derivative of the cost function w.r.t. the order placement variables. For the one price-level solution we can use the constraint on the total order size to capture the fact that placing more MOs means that the traders can place less LOs. When the number of order placement variables is then increased for the two, three and n price-level solutions, this fact is no longer captured in the optimal solution. This leads to a discrepancy between the found optimal static solution and the true solution that minimises the cost function. This discrepancy can be removed by adjusting the number of price-levels our static solution considers. What might also be interesting, however, is to analyse the problem as a stochastic linear programming, to make sure the solution always adheres to the optimality constraints, without need for adjustments.

Lastly the static model also makes some key assumptions. It assumes that no matter how many MOs are placed, they never walk the book, meaning the best sell quote queue is greater than our target number of shares. Although this assumption is simple to check, it could be an interesting addition, to take into account what effect it has on our optimal solution, if placing too many market orders means walking up the book. Furthermore it does not take into account price levels inside the spread. Doing this could allow the model to be extended to also include price drifts or changes in the spread size.

The dynamic model currently assumes a simplified, depth-based fill probability for any LO and a Poisson arrival for MOs. In reality, fill probabilities depend on more detailed queue dynamics, including order size, queue imbalance, and cancellation rates. It is important to note that the distribution of the fill probability plays a crucial role in whether the dynamic model allows for an analytical solution. Future work could, however, look into fill-probability functions, perhaps using survival analysis or empirical estimation, to assess how sensitive the dynamic solution is to this assumption.

Moreover, the dynamic model assumes that a trader can place a limit order at any "continuous" depth. However, in real exchanges, prices are discrete multiples of a tick size. Consequently, the optimal "continuous" depth must be rounded to the nearest tick, potentially increasing expected cost. Future work should look to discretise the LO depth and evaluate how rounding affects the dynamic strategy. This could be done by solving a discrete dynamic programming problem over integer tick increments. The dynamic strategy also re-places limit orders at the optimal depth every time step. In a real LOB this would send the trader's order to the back of the queue at that price-level, which is very impactful for the fill probability of that order. Accounting for these LOB mechanics more accurately could improve the practical applicability and performance of the optimal strategy.

In the functions describing the dynamic optimal strategy certain combinations of values for the mid-price drift and running inventory penalty lead to asymptotic "blow-up" values in our solution. Intuitively this happens when the price drift pushes the dynamic optimal strategy one way and the running inventory penalty pushes the other way, causing our equations to break down. Interpolation from values above and below the problematic combinations or implementing smoothing adjustments could help resolve this issue.

As noted, the dynamic strategy's computational runtime grows exponentially with inventory size, due to the recursive nature of the solution. To make it viable for larger inventories (e.g., inventories greater than 15), one could explore approximate dynamic programming methods like Value Function Approximation (using basis functions or neural networks), Model Predictive Control or Grid Coarsening (solving the dynamic problem on a coarse grid of inventory and time). Implementing these methods might make the dynamic optimal strategy more practically applicable for a real-world trader in a high-speed environment.

Lastly, although our numerical experiments use parameter values inspired by realistic scenarios, its performance remains to be tested on real LOB data. Further research on calibration of parameters could be very interesting and will be crucial to make the found solutions practically applicable.

7.4 Final Remarks

This work offers two different approaches to the general Optimal Placement Problem, considering market orders and limit orders at all depths. The static convex optimisation model offers intuitive insights in the workings of the queueing mechanism in a limit order book. Its n price-level solution is computationally efficient and, when combined with exponential distribution functions for order arrivals, results in a explicit optimal order allocation that has consistently lower expected cost than benchmark strategies and is more widely applicable. The dynamic optimal stopping and control problem excels when the LOB environment is volatile and mid-price drift is likely. The dynamic optimal strategy also leads to higher average earnings per share compared to benchmark strategies, but may require approximate or heuristic methods to scale to large inventories in practice. Future extensions, like the testing on different arrival distributions, depth discretisation and parameter calibration, could further improve the performance and practical applicability of these approaches in real-world trading systems.

Bibliography

[1] Robert Almgren and Neil Chriss. "Optimal execution of portfolio transactions". In: *The Journal of Risk* 3.2 (Jan. 2001), pp. 5–39. DOI: 10.21314/jor.2001.041.

[2] Yang An and Ngai Hang Chan. "Short-Term Stock Price Prediction Based on Limit Order Book Dynamics". en. In: *Journal of Forecasting* 36.5 (2017), pp. 541–556. ISSN: 1099-131X. DOI: 10.1002/for.2452. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2452 (visited on 09/04/2024).

[3] Burcu Aydoğan, Ömür Uğur, and Ümit Aksoy. "Optimal Limit Order Book Trading Strategies with Stochastic Volatility in the Underlying Asset". en. In: *Computational Economics* 62.1 (June 2023), pp. 289–324. ISSN: 1572-9974. DOI: 10.1007/s10614-022-10272-4. URL: https://doi.org/10.1007/s10614-022-10272-4 (visited on 11/06/2024).

[4] Francesco Capponi and Rama Cont. "Multi-Asset Market Impact and Order Flow Commonality". In: *SSRN Electronic Journal* (Jan. 2020). DOI: 10.2139/ssrn.3706390.

[5] Alvaro Cartea and Sebastian Jaimungal. "Optimal execution with limit and market orders". In: *Quantitative Finance* 15.8 (2015), pp. 1279–1291. ISSN: 1469-7688. DOI: 10.1080/14697688.2015.1032543.

[6] Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. "Enhancing trading strategies with order book signals". In: *Applied Mathematical Finance* 25 (Feb. 2018), pp. 1–35. DOI: 10.1080/1350486X.2018.1434009.

[7] Álvaro Cartea and Sebastian Jaimungal. "Incorporating order-flow into optimal execution". en. In: *Mathematics and Financial Economics* 10.3 (June 2016). Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer Berlin Heidelberg, pp. 339–364. ISSN: 1862-9660. DOI: 10.1007/s11579-016-0162-z. URL: https://link.springer.com/article/10.1007/s11579-016-0162-z (visited on 05/13/2025).

[8] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and High-Frequency Trading.* Cambridge, UK: Cambridge University Press, 2015. ISBN: 9781107091146.

[9] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. "Buy Low, Sell High: A High Frequency Trading Perspective". In: *SIAM Journal on Financial Mathematics* 5.1 (2014), pp. 415–444. DOI: 10.1137/130911196. URL: https://doi.org/10.1137/130911196.

[10] George H. Chen. *Survival Kernets: Scalable and Interpretable Deep Kernel Survival Analysis with an Accuracy Guarantee.* arXiv:2206.10477 [cs, stat]. Feb. 2024. DOI: 10.48550/arXiv.2206.10477. URL: http://arxiv.org/abs/2206.10477 (visited on 09/04/2024).

[11] Jingnan Chen et al. "Optimal portfolio execution with a Markov chain approximation approach". In: *IMA Journal of Management Mathematics* 34 (Aug. 2021). DOI: 10.1093/imaman/dpab025.

[12] R. Cont, A. Kukanov, and S. Stoikov. "The Price Impact of Order Book Events". In: *Journal of Financial Econometrics* 12.1 (June 2013), pp. 47–88. DOI: 10.1093/jjfinec/nbt003.

[13] R. Cont, A. Kukanov, and S. Stoikov. “The Price Impact of Order Book Events”. In: *Journal of Financial Econometrics* 12.1 (June 2013), pp. 47–88. ISSN: 1479-8417. DOI: 10.1093/jjfinec/nbt003. URL: http://dx.doi.org/10.1093/jjfinec/nbt003.

[14] Rama Cont, Pierre Degond, and Lifan Xuan. *A Mathematical Framework for Modelling Order Book Dynamics*. en. SSRN Scholarly Paper. Rochester, NY, Dec. 2022. DOI: 10.2139/ssrn.4304211. URL: https://papers.ssrn.com/abstract=4304211 (visited on 10/17/2024).

[15] Rama Cont and Arseniy Kukanov. “Optimal order placement in limit order markets”. In: *Quantitative Finance* 17.1 (Jan. 2017), pp. 21–39. ISSN: 1469-7688. DOI: 10.1080/14697688.2016.1190030. URL: https://doi.org/10.1080/14697688.2016.1190030 (visited on 10/17/2024).

[16] Rama Cont and Marvin S. Mueller. *A stochastic partial differential equation model for limit order book dynamics*. en. Apr. 2019. URL: https://arxiv.org/abs/1904.03058v2 (visited on 10/09/2024).

[17] Jin Fang et al. *Imitate then Transcend: Multi-Agent Optimal Execution with Dual-Window Denoise PPO*. 2022. arXiv: 2206.10736 [cs.LG]. URL: https://arxiv.org/abs/2206.10736.

[18] Joaquin Fernandez-Tapia. “Modeling, optimization and estimation for the on-line control of trading algorithms in limit-order markets”. PhD thesis. Pierre and Marie Curie Universtiy, Sept. 2015. DOI: 10.13140/RG.2.1.1490.5684.

[19] José E. Figueroa-López, Hyoeun Lee, and Raghu Pasupathy. *Optimal placement of a small order in a diffusive limit order book*. arXiv:1708.04337. Aug. 2017. DOI: 10.48550/arXiv.1708.04337. URL: http://arxiv.org/abs/1708.04337 (visited on 11/05/2024).

[20] L. Fornaro. “A Novel Optimal Execution Strategy Using Data-Driven Methods and Stochastic Modeling, with Application in the FX Spot Market”. Retrieved from TU Delft repository. Master’s thesis. Delft University of Technology, 2024. URL: https://resolver.tudelft.nl/uuid:b9504af7-0490-43d0-ac4c-9c5693b5f95d.

[21] Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. “Optimal Portfolio Liquidation with Limit Orders”. In: *SIAM Journal on Financial Mathematics* 3.1 (2012), pp. 740–764. ISSN: 1945-497X. DOI: 10.1137/110850475.

[22] Fabien Guilbaud and Huyen Pham. *Optimal High Frequency Trading with limit and market orders*. 2011. arXiv: 1106.5040 [q-fin.TR]. URL: https://arxiv.org/abs/1106.5040.

[23] Xin Guo. “Optimal Placement in a Limit Order Book”. en. In: *Theory Driven by Influential Applications*. Ed. by Huseyin Topaloglu, J. Cole Smith, and Harvey J. Greenberg. INFORMS, Sept. 2013, pp. 191–200. ISBN: 978-0-9843378-4-2. DOI: 10.1287/educ.2013.0113. URL: http://pubsonline.informs.org/doi/abs/10.1287/educ.2013.0113 (visited on 10/30/2024).

[24] Xin Guo, Adrien Larrard, and Zhao Ruan. “Optimal Placement in a Limit Order Book”. In: *SSRN Electronic Journal* (Jan. 2013). DOI: 10.2139/ssrn.2318220.

[25] Yanhong Guo and Xinxin Chen. “Forecasting the Mid-price Movements with High-Frequency LOB: A Dual-Stage Temporal Attention-Based Deep Learning Architecture”. en. In: *Arabian Journal for Science and Engineering* 48.8 (Aug. 2023), pp. 9597–9618. ISSN: 2191-4281. DOI: 10.1007/s13369-022-07197-3. URL: https://doi.org/10.1007/s13369-022-07197-3 (visited on 09/04/2024).

[26] Ben Hambly, Jasdeep Kalsi, and James Newbury. *Limit order books, diffusion approximations and reflected SPDEs: from microscopic to macroscopic models*. 2019. arXiv: 1808.07107 [q-fin.MF]. URL: https://arxiv.org/abs/1808.07107.

[27] Nikolaus Hautsch and Ruihong Huang. "The market impact of a limit order". In: *Journal of Economic Dynamics and Control* 36.4 (Apr. 2012), pp. 501–522. ISSN: 0165-1889. DOI: 10.1016/j.jedc.2011.09.012. URL: https://www.sciencedirect.com/science/article/pii/S0165188911002375 (visited on 08/07/2025).

[28] Konark Jain et al. *Limit Order Book Dynamics and Order Size Modelling Using Compound Hawkes Process*. en. Dec. 2023. URL: https://arxiv.org/abs/2312.08927v5 (visited on 10/09/2024).

[29] Michaël Karpe et al. "Multi-agent reinforcement learning in a realistic limit order book market simulation". In: *Proceedings of the First ACM International Conference on AI in Finance*. ICAIF '20. ACM, Oct. 2020. DOI: 10.1145/3383455.3422570. URL: http://dx.doi.org/10.1145/3383455.3422570.

[30] Kissell. *Algorithmic Trading Methods*. Elsevier, Jan. 2021. DOI: 10.1016/c2017-0-03456-0.

[31] Petter N. Kolm, Jeremy Turiel, and Nicholas Westray. "Deep order flow imbalance: Extracting alpha at multiple horizons from the limit order book". en. In: *Mathematical Finance* 33.4 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mafi.12413, pp. 1044–1081. ISSN: 1467-9965. DOI: 10.1111/mafi.12413. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12413 (visited on 09/04/2024).

[32] Felix Lokin and Fenghui Yu. *Fill Probabilities in a Limit Order Book with State-Dependent Stochastic Order Flows*. 2024. arXiv: 2403.02572 [q-fin.TR]. URL: https://arxiv.org/abs/2403.02572.

[33] Maxime Morariu-Patrichi and Mikko S. Pakkanen. *State-dependent Hawkes processes and their application to limit order book modelling*. arXiv:1809.08060. Sept. 2021. URL: http://arxiv.org/abs/1809.08060 (visited on 10/28/2024).

[34] Max Oliveberg. *Optimal Order Placement Using Markov Models of Limit Order Books*. eng. 2023. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-342769 (visited on 06/01/2025).

[35] Matthias Schnaubelt. "Deep reinforcement learning for the optimal placement of cryptocurrency limit orders". In: *European Journal of Operational Research* 296 (May 2021). DOI: 10.1016/j.ejor.2021.04.050.

[36] H. Mete Soner and Nizar Touzi. "Hedging under Gamma Constraints by Optimal Stopping and Face-lifting". In: *Mathematical Finance* 17.1 (2006), pp. 59–79. DOI: 10.1111/j.1467-9965.2007.00294.x.

[37] Weijing Tang et al. *SODEN: A Scalable Continuous-Time Survival Model through Ordinary Differential Equation Networks*. arXiv:2008.08637 [cs, stat]. Dec. 2021. DOI: 10.48550/arXiv.2008.08637. URL: http://arxiv.org/abs/2008.08637 (visited on 09/04/2024).

[38] Zihao Zhang, Stefan Zohren, and Stephen Roberts. "DeepLOB: Deep Convolutional Neural Networks for Limit Order Books". In: *IEEE Transactions on Signal Processing* 67.11 (June 2019). Conference Name: IEEE Transactions on Signal Processing, pp. 3001–3012. ISSN: 1941-0476. DOI: 10.1109/TSP.2019.2907260. URL: https://ieeexplore.ieee.org/document/8673598/?arnumber=8673598 (visited on 09/11/2024).

A | Proofs

A.1 Proposition 4.1

Under assumption A1-A3 all optimal allocations belong to the set

$$\mathcal{C} = \{(M^\star, L_1^\star, L_2^\star, ...) :\ 0 \le M^\star \le S,\ \ 0 \le L_i^\star \le S,\ \ M^\star + \sum L_i^\star = S\} \tag{A.1}$$

Proof. Since market orders are always filled, the second inequality in assumption **A2** tells us that $0 \le M^\star \le S$. Now, suppose we have an allocation $\boldsymbol{X}$ with $0 \le M \le S$, $L_1 \le S - M$, $L_2 = S - M - L_1$,..., but there is a price-level k such that

$$\{\min_{k>0} k : L_k > S - M - \sum_{i=1}^{k-1} L_i\} \tag{A.2}$$

which means $\sum L_i^\star + M^\star > S$. We then make a new allocation $\boldsymbol{X}'$ with all the same values except $L_k' = S - M - \sum_{i=1}^{k-1} L_i$. Then the difference in cost between the two allocations is given by

$$\begin{aligned} v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = &- (h + (k-1)\delta + r)(\mathrm{OF}_k - \mathrm{OF}_k') + \rho_u\big((S-A)_+ - (S-A')_+\big) \\ &+ \rho_o\big((A-S)_+ - 0\big) + \theta\big(L_2 - L_2' + (S-A)_+ - (S-A')\big). \end{aligned} \tag{A.3}$$

We must now differentiate between four different scenarios:

(i) $\xi + \psi_1 + ... + \psi_{k-1} < \sum_{i=1}^{k-1}(Q_i + L_i) + Q_k$:

We have $\mathrm{OF}_k = \mathrm{OF}_k' = 0$, $(S-A)_+ = (S-A')_+ = S - M - \sum_{i=1}^{k-1} \mathrm{OF}_i$ and $(A-S)_+ = 0$, so $v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = \theta(L_k - L_k') > 0$.

(ii) $\sum_{i=1}^{k-1}(Q_i + L_i) + Q_k \le \xi + \psi_1 + ... + \psi_{k-1} < \sum_{i=1}^{k-1}(Q_i + L_i) + Q_k + L_k'$:

We have $\mathrm{OF}_k = \mathrm{OF}_k' = \xi + \psi_1 + ... + \psi_{k-1} - \sum_{i=1}^{k-1}(Q_i + L_i) - Q_k \le S - M - \sum_{i=1}^{k-1} L_i$, $(S-A)_+ = (S-A')_+ = S - M - \sum_{i=1}^{k-1} \mathrm{OF}_i - \mathrm{OF}_k'$ and $(A-S)_+ = 0$, so $v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = \theta(L_k - L_k') > 0$.

(iii) $\sum_{i=1}^{k-1}(Q_i + L_i) + Q_k + L_k' \le \xi + \psi_1 + ... + \psi_{k-1} < \sum_{i=1}^{k-1}(Q_i + L_i) + Q_k + L_k$:

We have $L_k' = \mathrm{OF}_k' \le \mathrm{OF}_k = \xi + \psi_1 + ... + \psi_{k-1} - \sum_{i=1}^{k-1}(Q_i + L_i) - Q_k$, $(S-A)_+ = (S-A')_+ = 0$ and $(A-S)_+ = \xi + \psi_1 + ... + \psi_{k-1} - \sum_{i=1}^{k-1}(Q_i + L_i) - Q_k - L_k' = \mathrm{OF}_k - \mathrm{OF}_k' \ge 0$, so $v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = (\rho_o - h - (k-1)\delta - r)(\mathrm{OF}_k - \mathrm{OF}_k') + \theta(L_k - L_k')$. We know from **A2** that $\rho_o + \theta \ge h + (k-1)\delta + r$ so we have $v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = (\rho_o - h - (k-1)\delta - r)(\mathrm{OF}_k - \mathrm{OF}_k') + \theta(L_k - L_k') > (\rho_o + \theta - h - (k-1)\delta - r)(\mathrm{OF}_k - \mathrm{OF}_k') \ge 0$, where the first inequality follows from **A2** and the fact that $\theta > 0$.

(iv) $\xi + \psi_1 + ... + \psi_{k-1} \ge \sum_{i=1}^{k-1}(Q_i + L_i) + Q_k + L_k$:

We have $L_k' = \mathrm{OF}_k' < \mathrm{OF}_k = L_k$, $(S-A)_+ = (S-A')_+ = 0$ and $(A-S)_+ = L_k - L_k' \ge 0$, so by **A2** $v(\boldsymbol{X}, \xi, \boldsymbol{\psi}) - v(\boldsymbol{X}', \xi, \boldsymbol{\psi}) = (\rho_o + \theta - h - (k-1)\delta - r)(L_k - L_k') > 0$, since $\rho_o + \theta \ge h + (k-1)\delta + r$.

We have proven an allocation $\boldsymbol{X} = (M^\star, L_1^\star, L_2^\star, ...)$ is suboptimal if $\sum L_i^\star + M^\star > S$. The fact that an allocation is suboptimal if $\sum L_i^\star + M^\star < S$ simply follows from the fact that setting $M' = S - \sum L_i^\star$ for a new allocation $\boldsymbol{X}'$ will always result in lower costs due to **A2**. □

A.2 Proposition 4.2

Assume that ξ has a continuous distribution and Assumptions (A1-A3) hold. Then the following allocations are optimal for the $n = 1$ case of the Order Placement Problem 4.1

(i) If $\rho_u \leq \underline{\rho_u} = \frac{2h+f+r}{F(Q_1+S)} - (h+r+\theta)$, then $(M^*, L_1^*) = (0, S)$ is an optimal allocation.

(ii) If $\rho_u \geq \overline{\rho_u} = \frac{2h+f+r}{F(Q_1)} - (h+r+\theta)$, then $(M^*, L_1^*) = (S, 0)$ is an optimal allocation.

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, an optimal allocation is a mix of limit and market orders, given by

$$M^* = S - F^{-1}\left(\frac{2h+f+r}{\rho_u+h+r+\theta}\right) + Q_1, \tag{A.4}$$

$$L_1^* = F^{-1}\left(\frac{2h+f+r}{\rho_u+h+r+\theta}\right) - Q_1, \tag{A.5}$$

where $F(x) = \mathbb{P}(\xi \leq x)$ is the distribution of the outflow at the first price-level $\xi = \mu + \gamma_1$ and F^{-1} is its left-inverse.

Proof. The expected costs is given by

$$\begin{aligned}\mathbb{E}[v(\boldsymbol{X}, \mu, \gamma)] :=& \mathbb{E}[(h+f)M - (h+r)\text{OF} \\ &+ \rho_u(S - A(\boldsymbol{X}, \mu, \gamma))_+ + \rho_o(A(\boldsymbol{X}, \mu, \gamma) - S)_+ \\ &+ \theta\,(M + L + (S - A(\boldsymbol{X}, \mu, \gamma))_+)].\end{aligned} \tag{A.6}$$

We know that $S = M + L$ and we set $\xi = \mu + \gamma$. In that case, we have $\{A(\boldsymbol{X}, \xi) > S\} = \emptyset$ and $\{A(\boldsymbol{X}, \xi) < S | \xi > Q + L\} = \emptyset$, which means that $(S - A(\boldsymbol{X}, \xi))_+ = (S - A(\boldsymbol{X}, \xi))1_{\{\xi < Q+L\}}$. This allows us to rewrite the expected cost as a function of M as

$$\begin{aligned}\mathbb{E}[v(\boldsymbol{X}, \xi)] = V(M) :=& \mathbb{E}[(h+f)M - (h+r)((\xi - Q)_+ - (\xi - Q - S + M)_+) \\ &+ \rho_u(S - M - ((\xi - Q)_+ - (\xi - Q - S + M)_+))1_{\{\xi < Q+S-M\}} \\ &+ \theta\left(S + (S - M - ((\xi - Q)_+ - (\xi - Q - S + M)_+))1_{\{\xi < Q+S-M\}}\right)] \\ =& \mathbb{E}[(h+f)M - (h+r)((\xi - Q)_+ - (\xi - Q - S + M)_+) \\ &+ (\rho_u + \theta)(S - M - ((\xi - Q)_+ - (\xi - Q - S + M)_+))1_{\{\xi < Q+S-M\}} + \theta S].\end{aligned} \tag{A.7}$$

For the amount of our orders filled (OF_1) is given by

$$\text{OF}_1 = (\xi - Q)_+ - (\xi - Q - S + M)_+ = (\xi - Q)1_{\{Q < \xi < Q+S-M\}} + (S - M)1_{\{\xi \geq Q+S-M\}}. \tag{A.8}$$

This allows us to rewrite the expected cost function as

$$\begin{aligned}
V(M) = \mathbb{E}[(h+f)M - (h+r)((\xi - Q)1_{\{Q<\xi<Q+S-M\}} + (S-M)1_{\{\xi\geq Q+S-M\}}) \quad (A.9)\\
+ (\rho_u+\theta)\left(S - M - (\xi - Q)1_{\{Q<\xi<Q+S-M\}} + (S-M)1_{\{\xi\geq Q+S-M\}}\right)1_{\{\xi<Q+S-M\}} + \theta S] \quad (A.10)\\
= \mathbb{E}[(h+f)M - (h+r)(\xi-Q)1_{\{Q<\xi<Q+S-M\}} - (h+r)(S-M)1_{\{\xi\geq Q+S-M\}} \quad (A.11)\\
+ (\rho_u+\theta)(S-M)1_{\{\xi<Q+S-M\}} - (\rho_u+\theta)(\xi-Q)1_{\{Q<\xi<Q+S-M\}} + \theta S] \quad (A.12)\\
= \mathbb{E}[(h+f)M - (h+r+\rho_u+\theta)(\xi-Q)1_{\{Q<\xi<Q+S-M\}} - (h+r)(S-M)1_{\{\xi\geq Q+S-M\}} \quad (A.13)\\
+ (\rho_u+\theta)(S-M)1_{\{\xi<Q+S-M\}} + \theta S]. \quad (A.14)
\end{aligned}$$

We are allowed to differentiate inside the expectation, due to the **dominated convergence theorem**. Doing so, we find the derivative of the cost function w.r.t. M, given by

$$\begin{aligned}
\frac{\partial \mathbb{E}[v(\boldsymbol{X},\xi)]}{\partial M} = V'(M) &= \mathbb{E}[(h+f) - 0\cdot 1_{\{Q<\xi<Q+S-M\}} + (h+r)1_{\{\xi\geq Q+S-M\}} - (\rho_u+\theta)1_{\{\xi<Q+S-M\}}]\\
&= \mathbb{E}[(h+f) + (h+r)1_{\{\xi>Q+S-M\}} - (\rho_u+\theta)1_{\{\xi<Q+S-M\}}]\\
&= \int_{-\infty}^{\infty}((h+f) + (h+r)1_{\{\xi>Q+S-M\}} - (\rho_u+\theta)1_{\{\xi<Q+S-M\}})f_\xi(x)dx\\
&= (h+f)\int_{-\infty}^{\infty} f_\xi(x)dx + (h+r)\int_{Q+S-M}^{\infty} f_\xi(x)dx - (\rho_u+\theta)\int_{-\infty}^{Q+S-M} f_\xi(x)dx\\
&= (h+f) + (h+r)(1 - F_\xi(Q+S-M)) - (\rho_u+\theta)F_\xi(Q+S-M)\\
&= 2h+f+r - (h+r+\rho_u+\theta)F_\xi(Q+S-M) = 0.
\end{aligned} \quad (A.15)$$

We see that if $\rho_u \leq \frac{2h+f+r}{F(Q+S)} - (h+r+\theta)$, then $V'(M) \geq 0$ for $M \in (0,S)$ and therefore V is non-decreasing at these points. Checking that $V(S)-V(0) \geq (h+f-\rho_u-\theta)S + (h+r+\rho_u+\theta)S(1-F(Q+S)) \geq 0$ we conclude that $M^\star = 0$ (since it is non-decreasing across the interval $M \in (0,S)$). Similarly, if $\rho_u \geq \frac{2h+f+r}{F(Q)} - (h+r+\theta)$, then $V'(M) \leq 0$ for all $M \in (0,S)$ and $V(M)$ is non-increasing at these points. Checking that $V(S)-V(0) \leq (h+f-\rho_u-\theta)S + (h+r+\rho_u+\theta)S(1-F(Q)) \leq 0$ we conclude that $M^\star = S$. Finally, if ρ_u is between these two values, $\exists\epsilon > 0$, such that $V'(\epsilon) < 0$, $V'(S-\epsilon) > 0$ and by continuity of V' there is a point where $V'(M^\star) = 0$. This $M^\star$ is optimal by convexity of $V(M)$ and Proposition A.2 solves the Optimal Placement Problem 4.1 for one price-level. □

A.3 Proposition 4.3

Assume that ξ and ψ have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocation for the $n = 2$ case of the Order Placement Problem 4.1 satisfies the following equations

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+\delta+f+r}{F_{\xi+\psi}(Q_1+Q_2)} - (h+r+\delta+\theta)$, then $(M^*, L_1^*, L_2^*) = (S,0,0)$ is an optimal allocation.

(ii) If $\rho_u \leq \underline{\rho_u} = \frac{2h+\delta+f+r}{F_{\xi+\psi}(Q_1+Q_2+S)} - (h+r+\delta+\theta)$, then $M^* = 0$ and L_1^* satisfies,

$$\delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1^\star) - (h+\delta+r+\rho_u+\theta)F_{\xi+\psi}(Q_1+Q_2+L_1^\star) = 0, \quad (A.16)$$

and L_2^* satisfies

$$L_2^\star = S - M^\star - L_1^\star. \quad (A.17)$$

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, then the optimal allocation is a mix of limit (L_1^*, L_2^*) and market $(M^\star)$ orders, where $L_1^\star$ satisfies (A.16), $L_2^\star$ satisfies (A.17), and $M^\star$ satisfies

$$M^\star = S - F_{\xi+\psi}^{-1}\left(\frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta}\right) + Q_1 + Q_2, \tag{A.18}$$

where $F_\xi(x) = \mathbb{P}(\xi \leq x)$ is the distribution of the outflow at the first price-level $\xi = \mu + \gamma_1$, $F_{\xi+\psi}(x) = \mathbb{P}(\xi + \psi \leq x)$ is the distribution of the outflow at the first and second price-level ξ, $\psi = -\eta_1 + \gamma_2$ and $F_\xi^{-1}, F_{\xi+\psi}^{-1}$ are their left-inverses.

Proof. Using (A.8) and (4.20) we can again write the expected cost function as a function of M and this time also L_1, so that $L_2 = S - M - L_1$ (again adhering to Proposition 4.1). The expected cost function is then given by

$$\begin{aligned}
\mathbb{E}[v(\boldsymbol{X},\xi,\psi)] = V(M,L_1) :=& \mathbb{E}[(h+f)M - (h+r)((\xi - Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} + L_1 \cdot 1_{\{\xi \geq Q_1+L_1\}}) \\
& - (h+\delta+r)(\ (\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+L_1+Q_2+S-M-L_1\}} \\
& \quad + (S-M-L_1)\cdot 1_{\{\xi+\psi>Q_1+L_1+Q_2+S-M-L_1\}}\ ) \\
& + (\rho_u+\theta)(\ S - (M + ((\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} + L_1\cdot 1_{\{\xi\geq Q_1+L_1\}}) \\
& \quad + ((\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+L_1+Q_2+S-M-L_1\}} \\
& \quad + (S-M-L_1)\cdot 1_{\{\xi+\psi>Q_1+L_1+Q_2+S-M-L_1\}}))\ )\cdot 1_{\{\xi+\psi<Q_1+L_1+Q_2+S-M-L_1\}}\cdot \\
& + \theta S]
\end{aligned} \tag{A.19}$$

Separating the terms and using $Q_1 + L_1 + Q_2 + S - M - L_1 = Q_1 + Q_2 + S - M$ we obtain

$$\begin{aligned}
V(M,L_1) =& \mathbb{E}[(h+f)M - (h+r)(\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} - (h+r)L_1\cdot 1_{\{\xi\geq Q_1+L_1\}}) \\
& - (h+\delta+r)(\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}} \\
& - (h+\delta+r)(S-M-L_1)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}} \\
& + (\rho_u+\theta)(\ S - M - (\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} - L_1\cdot 1_{\{\xi\geq Q_1+L_1\}} \\
& \quad - (\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}} \\
& \quad - (S-M-L_1)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}}\ )\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}} + \theta S] \\
=& \mathbb{E}[(h+f)M - (h+r)(\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} - (h+r)L_1\cdot 1_{\{\xi\geq Q_1+L_1\}}) \\
& - (h+\delta+r)(\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}} \\
& - (h+\delta+r)(S-M-L_1)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}} \\
& + (\rho_u+\theta)(S-M)\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}} - (\rho_u+\theta)(\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} \\
& - (\rho_u+\theta)L_1\cdot 1_{\{\xi\geq Q_1+L_1\}}\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}} \\
& - (\rho_u+\theta)(\xi+\psi-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}} + \theta S].
\end{aligned} \tag{A.20}$$

Taking the derivative of V w.r.t. the number of market orders M gives

$$\begin{aligned}
\frac{\partial V}{\partial M} &= \mathbb{E}[(h+f) + (h+\delta+r)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}} - (\rho_u+\theta)\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}}] \\
&= \int_{-\infty}^{\infty} ((h+f) + (h+\delta+r)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}} - (\rho_u+\theta)\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}}) f_{\xi+\psi}(x)dx \\
&= (h+f)\int_{-\infty}^{\infty} f_{\xi+\psi}(x)dx + (h+\delta+r)\int_{Q_1+Q_2+S-M}^{\infty} f_{\xi+\psi}(x)dx - (\rho_u+\theta)\int_{-\infty}^{Q_1+Q_2+S-M} f_{\xi+\psi}(x)dx \\
&= h + f + (h+\delta+r)(1 - F_{\xi+\psi}(Q_1+Q_2+S-M)) - (\rho_u+\theta)F_{\xi+\psi}(Q_1+Q_2+S-M) \\
&= 2h + \delta + f + r - (h+\delta+r+\rho_u+\theta)F_{\xi+\psi}(Q_1+Q_2+S-M).
\end{aligned} \tag{A.21}$$

Setting (A.21) equal to zero and using similar reasoning as in the 1 price-level case, we see that if $\rho_u \leq \frac{2h+\delta+f+r}{F(Q_1+Q_2+S)} - (h+\delta+r+\theta)$, then $\frac{\partial V}{\partial M} \geq 0$ for $M \in (0, S)$, so $M^\star = 0$. Similarly we also see for $\rho_u \geq \frac{2h+\delta+f+r}{F(Q_1+Q_2)} - (h+\delta+r+\theta)$ that $\frac{\partial V}{\partial M} \leq 0$ for $M \in (0, S)$, so $M^\star = S$. For $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$ the optimal number of market orders $M^\star$ is given by

$$M^\star = S - F^{-1}_{\xi+\psi}\left(\frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta}\right) + Q_1 + Q_2. \tag{A.22}$$

Taking the derivative of V w.r.t. the number of limit orders at the best quote L_1 gives

$$\begin{aligned}
\frac{\partial V}{\partial L_1} =& \mathbb{E}[-(h+r)\cdot 1_{\{\xi \geq Q_1+L_1\}} \\
&+ (h+\delta+r)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}} \\
&+ (h+\delta+r)\cdot 1_{\{\xi+\psi>Q_1+Q_2+S-M\}} \\
&- (\rho_u+\theta)\cdot 1_{\{\xi\geq Q_1+L_1\}}\cdot 1_{\{\xi+\psi<Q_1+Q_2+S-M\}} \\
&+ (\rho_u+\theta)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi<Q_1+Q_2+S-M\}}] \\
=& -(h+r)(1-F_\xi(Q_1+L_1)) \\
&+ (h+\delta+r)(F_{\xi+\psi}(Q_1+Q_2+S-M) - F_{\xi+\psi}(Q_1+Q_2+L_1)) \\
&+ (h+\delta+r)(1-F_{\xi+\psi}(Q_1+Q_2+S-M)) \\
&- \int_{-\infty}^{\infty}(\rho_u+\theta)\cdot 1_{\{x\geq Q_1+L_1\}}\cdot 1_{\{x+y<Q_1+Q_2+S-M\}} f_{\xi,\psi}(x,y)d(x,y) \\
&+ (\rho_u+\theta)(F_{\xi+\psi}(Q_1+Q_2+S-M) - F_{\xi+\psi}(Q_1+Q_2+L_1)).
\end{aligned} \tag{A.23}$$

We evaluate the integral separately, since we have to be careful with the boundaries. The integral above can be rewritten to

$$\begin{aligned}
&= -\int_{-\infty}^{\infty}(\rho_u+\theta)\cdot 1_{\{x\geq Q_1+L_1\}}\cdot 1_{\{x+y<Q_1+Q_2+S-M\}} f_{\xi,\psi}(x,y)d(x,y) \\
&= -(\rho_u+\theta)\int_{y=-\infty}^{y=Q_2}\int_{x=Q_1+L_1}^{x=Q_1+Q_2+S-M-y} f_\xi(x)f_\psi(y)dxdy \\
&= -(\rho_u+\theta)\int_{y=-\infty}^{y=Q_2}(F_\xi(Q_1+Q_2+S-M-y) - F_\xi(Q_1+L_1))f_\psi(y)dy \\
&= -(\rho_u+\theta)\left(\int_{y=-\infty}^{y=Q_2} F_\xi(Q_1+Q_2+S-M-y)f_\psi(y)dy - F_\psi(Q_2)F_\xi(Q_1+L_1)\right) \\
&= (\rho_u+\theta)\left(F_\psi(Q_2)F_\xi(Q_1+L_1) - F_{\xi+\psi}(Q_1+Q_2+S-M)\right)
\end{aligned} \tag{A.24}$$

where $F_\psi(Q_2) = 1$ since $\psi \in (-\infty, Q_2]$. We know from convolution of probability functions, that if we have two **independent** random variables X, Y, then the probability of their sum $X+Y$ is given by

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \tag{A.25}$$

which can be seen as simply saying, the probability that the sum is equal to z is found by adding up all probabilities where $X = x$ and $Y = z - x$. Now if we were to integrate this function, to find its CDF, we get

$$\begin{aligned}
F_{X+Y}(A) &= \int_{-\infty}^{A} f_{X+Y}(z)dz = \int_{-\infty}^{A}\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dxdz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{A} f_X(x)f_Y(z-x)dzdx = \int_{-\infty}^{\infty} F_Y(A-x)f_X(x)dx.
\end{aligned} \tag{A.26}$$

We find the derivative of the expected cost function is given by

$$\begin{aligned}\frac{\partial V}{\partial L_1} =& -(h+r)(1-F_\xi(Q_1+L_1)) \\ &+(h+\delta+r)(F_{\xi+\psi}(Q_1+Q_2+S-M)-F_{\xi+\psi}(Q_1+Q_2+L_1)) \\ &+(h+\delta+r)(1-F_{\xi+\psi}(Q_1+Q_2+S-M)) \\ &+(\rho_u+\theta)(F_\xi(Q_1+L_1)-F_{\xi+\psi}(Q_1+Q_2+L_1)) \\ =&\delta+(h+r+\rho_u+\theta)F_\xi(Q_1+L_1))-(h+\delta+r+\rho_u+\theta)F_{\xi+\psi}(Q_1+Q_2+L_1)=0.\end{aligned} \tag{A.27}$$

This can be simplified to

$$AF_{\xi+\psi}(Q_1+Q_2+L_1)+BF_\xi(Q_1+L_1)+C=0 \tag{A.28}$$

with

$$A=-(h+\delta+r+\rho_u+\theta), \quad B=h+r+\rho_u+\theta, \quad C=\delta. \tag{A.29}$$

With the assumptions made earlier, the optimal number of limit orders in the second best price level, should simply be $L_2^\star = S - M^\star - L_1^\star$, but this is only the case when (A.28) has root $L_1^\star \in (0, S)$. If this root does not exist in this interval, it means it would be optimal to place $L_1 = 0$ (when $\frac{\partial V}{\partial L_1} > 0$ for $L_1 \in (0, S)$) or $L_1 = S$ (when $\frac{\partial V}{\partial L_1} < 0$ for $L_1 \in (0, S)$).

□

A.4 Lemma 4.1

If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, and $\psi \sim Q_2 - Y$, where Q_2 is the length of the second price-level queue and Y is independent and exponentially distributed with rate λ_ψ, the cumulative distribution function of the convolution of the random variables ξ and ψ is given by

$$F_{\xi+\psi}(A) = \begin{cases} \frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi} e^{-\lambda_\psi(Q_2-A)} & , \; A \le Q_2 \\ 1 - \frac{\lambda_\psi}{\lambda_\xi+\lambda_\psi} e^{-\lambda_\xi(A-Q_2)} & , \; A > Q_2. \end{cases} \tag{A.30}$$

Proof. For the exponential distribution the cumulative distribution function (CDF) is given by $F_X(x) = 1 - e^{-\lambda_X x}$. We set $\xi \sim X$ and $\psi \sim Q_2 - Y$ where X is exponentially distributed with rate λ_ξ , Q_2 is the length of our second price-level queue and Y is exponentially distributed with rate λ_ψ. An analysis of limit order arrivals in [32] shows that the exponential distribution is a reasonable approximation of actual limit orders arriving in FX spot markets. So for ψ we have

$$f_\psi(\psi = y) = \lambda_\psi e^{-\lambda_\psi(Q_2-y)}, \tag{A.31}$$

meaning the CDF is given by

$$F_\psi(\psi = y) = e^{-\lambda_\psi(Q_2-y)} \tag{A.32}$$

where $y \in (-\infty, Q_2]$. To find the CDF of the convolution of ξ and ψ, which is denoted by $F_{\xi+\psi}(A)$, we look at two different cases. Firstly, when $A > Q_2$, the CDF of the convolution of ξ and ψ is given by

$$\begin{aligned}F_{\xi+\psi}(A) =& \int_{-\infty}^{Q_2} F_\xi(A-y)f_\psi(y)dy\\ =& \int_{-\infty}^{Q_2}(1-e^{-\lambda_\xi(A-y)})\lambda_\psi e^{-\lambda_\psi(Q_2-y)}dy\\ =&\lambda_\psi\int_{-\infty}^{Q_2} e^{-\lambda_\psi(Q_2-y)}dy - \lambda_\psi e^{-\lambda_\xi A-\lambda_\psi Q_2}\int_{-\infty}^{Q_2} e^{(\lambda_\xi+\lambda_\psi)y}dy\\ =&\left[e^{-\lambda_\psi(Q_2-y)}\right]_{-\infty}^{Q_2} - \frac{\lambda_\psi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\xi A-\lambda_\psi Q_2}\left[e^{(\lambda_\xi+\lambda_\psi)y}\right]_{-\infty}^{Q_2}\\ =&1-\frac{\lambda_\psi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\xi(A-Q_2)}.\end{aligned} \tag{A.33}$$

Now if $A \leq Q_2$, the CDF of the convolution of ξ and ψ is given by

$$\begin{aligned}F_{\xi+\psi}(A) =& \int_0^\infty F_\psi(A-x)f_\xi(x)dx\\ =& \int_0^\infty e^{-\lambda_\psi(Q_2-A+x)}\lambda_\xi e^{-\lambda_\xi x}dx\\ =&\lambda_\xi e^{-\lambda_\psi(Q_2-A)}\int_0^\infty e^{-(\lambda_\xi+\lambda_\psi)x}dx\\ =&-\frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\psi(Q_2-A)}\left[e^{-(\lambda_\xi+\lambda_\psi)x}\right]_0^\infty\\ =&\frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\psi(Q_2-A)}.\end{aligned} \tag{A.34}$$

We see that these two cases indeed coincide at the point when $A = Q_2$, so we find that the convolution of ξ and ψ is given by

$$F_{\xi+\psi}(A) = \begin{cases}\frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\psi(Q_2-A)} & , \ A \leq Q_2\\ 1-\frac{\lambda_\psi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\xi(A-Q_2)} & , \ A > Q_2.\end{cases} \tag{A.35}$$

The inverse of this function is given by

$$F^{-1}_{\xi+\psi}(x) = \begin{cases}Q_2 - \frac{1}{\lambda_\psi}\log\left(\frac{\lambda_\xi}{(\lambda_\xi+\lambda_\psi)x}\right) & , \ x \leq F_{\xi+\psi}(Q_2) = \frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}\\ Q_2 + \frac{1}{\lambda_\xi}\log\left(\frac{\lambda_\psi}{(\lambda_\xi+\lambda_\psi)(1-x)}\right) & , \ x > F_{\xi+\psi}(Q_2) = \frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}.\end{cases} \tag{A.36}$$

We can use (A.35) and (A.36) to find a closed-form expression of the optimal solution $M^\star$ and $L_1^\star$. For $M^\star$, given by equation (4.25), we find

$$M^\star = \begin{cases}S + \frac{1}{\lambda_\psi}\log\left(\frac{\lambda_\xi}{(\lambda_\xi+\lambda_\psi)}\frac{h+\delta+r+\rho_u+\theta}{2h+\delta+f+r}\right)+Q_1 & , \ \frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta} \leq \frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}\\ S - \frac{1}{\lambda_\xi}\log\left(\frac{\lambda_\psi}{(\lambda_\xi+\lambda_\psi)}\frac{h+\delta+r+\rho_u+\theta}{-h-f+\rho_u+\theta}\right)+Q_1 & , \ \frac{2h+\delta+f+r}{h+\delta+r+\rho_u+\theta} > \frac{\lambda_\xi}{\lambda_\xi+\lambda_\psi}\end{cases} \tag{A.37}$$

and to derive a closed form expression $L_1^\star$, we insert (A.35) and (A.32) into equation (4.23), to find

$$\delta+(h+r+\rho_u+\theta)(1-e^{-\lambda_\xi(Q_1+L_1)})-(h+\delta+r+\rho_u+\theta)(1-\frac{\lambda_\psi}{\lambda_\xi+\lambda_\psi}e^{-\lambda_\xi((Q_1+Q_2+L_1)-Q_2)}) = 0, \tag{A.38}$$

since $Q_1 + Q_2 + L_1 \geq Q_2$. We find that equation (A.38) holds if

$$\begin{aligned} 0 =& - e^{-\lambda_\xi (Q_1+L_1)} \left((h + r + \rho_u + \theta) - (h + \delta + r + \rho_u + \theta) \frac{\lambda_\psi}{\lambda_\xi + \lambda_\psi} \right), \\ \frac{\lambda_\xi + \lambda_\psi}{\lambda_\psi} =& \frac{h + \delta + r + \rho_u + \theta}{h + r + \rho_u + \theta}, \\ \frac{\lambda_\xi}{\lambda_\psi} =& \frac{\delta}{h + r + \rho_u + \theta}, \\ \lambda_\xi =& \frac{\lambda_\psi \delta}{h + r + \rho_u + \theta}. \end{aligned} \tag{A.39}$$

We find that if $\lambda_\xi > \frac{\lambda_\psi \delta}{h+r+\rho_u+\theta}$, the derivative $\frac{\partial V}{\partial L_1} > 0$, so it is no longer optimal to place L_1 limit orders and the trader should instead place L_2 limit orders. The fact that this sudden switch in optimal limit order price-level can be expressed as a closed-form critical point is a nice result.

□

A.5 Proposition 4.4

Assume that ξ, ψ_1, and ψ_2 have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocation for the $n = 3$ case of the Order Placement Problem 4.1 satisfies the following, based on the valuw of ρ_u.

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+2\delta+f+r}{F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3)} - (h + 2\delta + r + \theta)$, then $(M^*, L_1^*, L_2^*, L_3^*) = (S, 0, 0, 0)$ is an optimal allocation.

(ii) If $\rho_u \leq \underline{\rho_u} = \frac{2h+2\delta+f+r}{F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S)} - (h + 2\delta + r + \theta)$, then $M^* = 0$ and L_1^* satisfies

$$\delta + (h + r + \rho_u + \theta) F_\xi (Q_1 + L_1^\star) - (h + \delta + r + \rho_u + \theta) F_{\xi+\psi_1} (Q_1 + L_1^\star + Q_2) = 0, \tag{A.40}$$

L_2^* satisfies

$$\begin{aligned} &\delta + (h + \delta + r + \rho_u + \theta) F_{\xi+\psi_1} (Q_1 + L_1^\star + Q_2 + L_2^\star) \\ &\quad - (h + 2\delta + r + \rho_u + \theta) F_{\xi+\psi_1+\psi_2} (Q_1 + L_1^\star + Q_2 + L_2^\star + Q_3) = 0, \end{aligned} \tag{A.41}$$

and L_3^* satisfies

$$L_3^\star = S - M - L_1^\star - L_2^\star. \tag{A.42}$$

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, then the optimal allocation is a mix of limit (L_1^*, L_2^*, L_3^*) and market $(M^\star)$ orders, where $L_1^\star$ satisfies (A.40), $L_2^\star$ satisfies (A.41), $L_3^\star$ satisfies (A.42) and $M^\star$ satisfies

$$M^\star = S - F^{-1}_{\xi+\psi_1+\psi_2} \left(\frac{2h + 2\delta + f + r}{h + 2\delta + r + \rho_u + \theta} \right) + Q_1 + Q_2 + Q_3, \tag{A.43}$$

where $F_\xi(x) = \mathbb{P}(\xi \leq x)$ is the distribution of the outflow at the first price-level ξ, $F_{\xi+\psi_1}(x) = \mathbb{P}(\xi + \psi_1 \leq x)$ is the distribution of the outflow at the first and second price-level ξ, ψ_1, $F_{\xi+\psi_1+\psi_2}(x) = \mathbb{P}(\xi + \psi_1 + \psi_2 \leq x)$ is the distribution of the outflow at the first, second and third price-level ξ, ψ_1,ψ_2 and $F_\xi^{-1}, F^{-1}_{\xi+\psi_1}, F^{-1}_{\xi+\psi_1+\psi_2}$ their left-inverses.

Proof. Using (A.8), (4.20) and (4.35) we can again write the expected cost function as a function of M, L_1 and this time also L_2, so that $L_3 = S - M - L_1 - L_2$ (again adhering to Proposition 4.1). The expected cost function is then given by

$$\begin{aligned}
\mathbb{E}[v(\boldsymbol{X}, \xi, \psi_1, \psi_2)] =& V(M, L_1, L_2) \\
:=& \mathbb{E}[(h+f)M - (h+r)((\xi - Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} + L_1 \cdot 1_{\{\xi \geq Q_1+L_1\}}) \\
& - (h+\delta+r)((\xi + \psi_1 - Q_1 - L_1 - Q_2) \cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
& + L_2 \cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}}) \\
& - (h+2\delta+r)((\xi + \psi_1 + \psi_2 - Q_1 - L_1 - Q_2 - L_2 - Q_3) \\
& \cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+S-M-L_1-L_2\}} \\
& + (S - M - L_1 - L_2) \cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+S-M-L_1-L_2\}}) \\
& + (\rho_u + \theta)(S - (M + ((\xi - Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} + L_1 \cdot 1_{\{\xi \geq Q_1+L_1\}}) \\
& + ((\xi + \psi_1 - Q_1 - L_1 - Q_2) \cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
& + L_2 \cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}}) \\
& + ((\xi + \psi_1 + \psi_2 - Q_1 - L_1 - Q_2 - L_2 - Q_3) \\
& \cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+S-M-L_1-L_2\}} \\
& + (S - M - L_1 - L_2) \cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+S-M-L_1-L_2\}}))) \\
& \cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+S-M-L_1-L_2\}} + \theta S].
\end{aligned} \tag{A.44}$$

Separating the terms and using $Q_1 + L_1 + Q_2 + L_2 + Q_3 + S - M - L_1 - L_2 = Q_1 + Q_2 + Q_3 + S - M$ we obtain

$$\begin{aligned}
V(M, L_1, L_2) :=& \mathbb{E}[(h+f)M - (h+r)(\xi - Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} - (h+r)L_1 \cdot 1_{\{\xi \geq Q_1+L_1\}} \\
& - (h+\delta+r)(\xi + \psi_1 - Q_1 - L_1 - Q_2) \cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
& - (h+\delta+r)L_2 \cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} \\
& - (h+2\delta+r)(\xi + \psi_1 + \psi_2 - Q_1 - L_1 - Q_2 - L_2 - Q_3) \\
& \quad \cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} \\
& - (h+2\delta+r)(S - M - L_1 - L_2) \cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}} \\
& + (\rho_u + \theta)(S - (M + (\xi - Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} + L_1 \cdot 1_{\{\xi \geq Q_1+L_1\}} \\
& \quad + (\xi + \psi_1 - Q_1 - L_1 - Q_2) \cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
& \quad + L_2 \cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} \\
& \quad + (\xi + \psi_1 + \psi_2 - Q_1 - L_1 - Q_2 - L_2 - Q_3) \\
& \quad \quad \cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} \\
& \quad + (S - M - L_1 - L_2) \cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}})) \\
& \cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} + \theta S].
\end{aligned} \tag{A.45}$$

Taking the derivative of $V(M, L_1, L_2)$ w.r.t. the number of market orders M gives

$$\begin{aligned}
\frac{\partial V}{\partial M} &= \mathbb{E}[(h+f) + (h+2\delta+r)\cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}} - (\rho_u+\theta)\cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}}] \\
&= \int_{-\infty}^{\infty}((h+f) + (h+2\delta+r)\cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}} \\
&- (\rho_u+\theta)\cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}})f_{\xi+\psi_1+\psi_2}(x)dx \\
&= (h+f)\int_{-\infty}^{\infty} f_{\xi+\psi_1+\psi_2}(x)dx + (h+2\delta+r)\int_{Q_1+Q_2+Q_3+S-M}^{\infty} f_{\xi+\psi_1+\psi_2}(x)dx \\
&- (\rho_u+\theta)\int_{-\infty}^{Q_1+Q_2+Q_3+S-M} f_{\xi+\psi_1+\psi_2}(x)dx \\
&= h+f+(h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M)) \\
&- (\rho_u+\theta)F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M) \\
&= 2h+2\delta+f+r-(h+2\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M).
\end{aligned} \tag{A.46}$$

Setting (A.46) equal to zero and using similar reasoning as in the 1 price-level case, we see that if $\rho_u \leq \frac{2h+2\delta+f+r}{F(Q_1+Q_2+Q_3+S)} - (h+2\delta+r+\theta)$, then $\frac{\partial V}{\partial M} \geq 0$ for $M \in (0, S)$, so $M^\star = 0$. Similarly we also see for $\rho_u \geq \frac{2h+2\delta+f+r}{F(Q_1+Q_2+Q_3)} - (h+2\delta+r+\theta)$ that $\frac{\partial V}{\partial M} \leq 0$ for $M \in (0, S)$, so $M^\star = S$. For $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$ the optimal number of market orders $M^\star$ is given by

$$M^\star = S - F^{-1}_{\xi+\psi_1+\psi_2}\left(\frac{2h+2\delta+f+r}{h+2\delta+r+\rho_u+\theta}\right) + Q_1+Q_2+Q_3 \tag{A.47}$$

and we clearly see a pattern starting to emerge.

Next, we look at what happens when we take the derivative of $V(M, L_1, L_2)$ with respect to L_1, which is given by

$$\begin{aligned}
\frac{\partial V}{\partial L_1} =& \mathbb{E}[-(h+r)\cdot 1_{\{\xi\geq Q_1+L_1\}} + (h+\delta+r)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
&+ (h+2\delta+r)\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} \\
&+ (h+2\delta+r)\cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}} \\
&+ (\rho_u+\theta)(\ -1_{\{\xi\geq Q_1+L_1\}} + 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
&\quad + 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}}\ )\cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}}] \\
=& \mathbb{E}[-(h+r)\cdot 1_{\{\xi\geq Q_1+L_1\}} + (h+\delta+r)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
&+ (h+2\delta+r)\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2\}} \\
&+ (\rho_u+\theta)(\ -1_{\{\xi\geq Q_1+L_1\}} + 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}}\ )\cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} \\
&+ (\rho_u+\theta)1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}}].
\end{aligned} \tag{A.48}$$

Using the definition of the expectation, we find the derivative of V w.r.t. L_1 to exist of the following integrals and be given by

$$\begin{aligned}\frac{\partial V}{\partial L_1} =& -(h+r)(1-F_\xi(Q_1+L_1)) + (h+\delta+r)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\ &+ (h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\ &- (\rho_u+\theta)\int 1_{\{x\geq Q_1+L_1\}} \cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\ &+ (\rho_u+\theta)\int 1_{\{Q_1+L_1+Q_2<x+y<Q_1+L_1+Q_2+L_2\}} \cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\ &+ (\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)).\end{aligned} \tag{A.49}$$

With similar reasoning as in the 2 price-level case, we find the first integral to be equivalent to the following

$$\begin{aligned}=& -(\rho_u+\theta)\int 1_{\{x\geq Q_1+L_1\}} \cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\ =& -(\rho_u+\theta)\int_{y+z=-\infty}^{y+z=Q_2+Q_3+S-M-L_1}\int_{x=Q_1+L_1}^{x=Q_1+Q_2+Q_3+S-M-(y+z)} f_\xi(x)f_{\psi_1+\psi_2}(y+z)dxd(y+z) \\ =& -(\rho_u+\theta)\int_{y+z=-\infty}^{y+z=Q_2+Q_3}\int_{x=Q_1+L_1}^{x=Q_1+Q_2+Q_3+S-M-(y+z)} f_\xi(x)f_{\psi_1+\psi_2}(y+z)dxd(y+z) \\ =& -(\rho_u+\theta)\int_{y+z=-\infty}^{y+z=Q_2+Q_3} (F_\xi(Q_1+Q_2+Q_3+S-M-(y+z)) - F_\xi(Q_1+L_1))f_{\psi_1+\psi_2}(y+z)d(y+z) \\ =& -(\rho_u+\theta)\Bigg(\int_{y+z=-\infty}^{y+z=Q_2+Q_3} F_\xi(Q_1+Q_2+Q_3+S-M-(y+z))f_{\psi_1+\psi_2}(y+z)d(y+z) \\ &- F_{\psi_1+\psi_2}(Q_2+Q_3)F_\xi(Q_1+L_1)\Bigg) \\ =&(\rho_u+\theta)\left(F_\xi(Q_1+L_1) - F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M)\right)\end{aligned} \tag{A.50}$$

and the second integral to be equal to

$$\begin{aligned}=&(\rho_u+\theta)\int 1_{\{Q_1+L_1+Q_2<x+y<Q_1+L_1+Q_2+L_2\}} \cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\ =&(\rho_u+\theta)\Bigg(\int_{z=Q_3+S-M-L_1-L_2}^{z=Q_3+S-M-L_1}\int_{x+y=Q_1+L_1+Q_2}^{x+y=Q_1+Q_2+Q_3+S-M-z} f_{\xi+\psi_1}(x+y)f_{\psi_2}(z)dzd(x+y) \\ &+ \int_{z=-\infty}^{z=Q_3+S-M-L_1-L_2}\int_{x+y=Q_1+L_1+Q_2}^{x+y=Q_1+L_1+Q_2+L_2} f_{\xi+\psi_1}(x+y)f_{\psi_2}(z)dzd(x+y)\Bigg) \\ =&(\rho_u+\theta)\Bigg(0 + \int_{z=-\infty}^{z=Q_3+S-M-L_1-L_2}(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2))f_{\psi_2}(z)dzd(x+y)\Bigg) \\ =&(\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2))\end{aligned} \tag{A.51}$$

since $\psi_2 \in (-\infty, Q_3]$ and $Q_3 \leq Q_3+S-M-L_1-L_2 \leq Q_3+S-M-L_1$ which means $\int_{Q_3+S-M-L_1-L_2}^{Q_3+S-M-L_1} f_{\psi_2}(z)dz = 0$.

This gives allows us to write the derivative of $V(M, L_1, L_2)$ w.r.t. L_1 as

$$
\begin{aligned}
\frac{\partial V}{\partial L_1} =& -(h+r)(1-F_\xi(Q_1+L_1)) + (h+\delta+r)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+ (\rho_u+\theta)\left(F_\xi(Q_1+L_1) - F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M)\right) \\
&+ (\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
=& 2\delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1) \\
&+ (h+\delta+r+\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&- (h+2\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3) = 0.
\end{aligned} \tag{A.52}
$$

We now move on to taking the derivative of $V(M, L_1, L_2)$ w.r.t. L_2 and see it is given by

$$
\begin{aligned}
\frac{\partial V}{\partial L_2} =& \mathbb{E}[-(h+\delta+r)\cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} + (h+2\delta+r)\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2\}} \\
&+ (\rho_u+\theta)(-1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} + 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}} \\
&+ 1_{\{\xi+\psi_1+\psi_2>Q_1+Q_2+Q_3+S-M\}}) \cdot 1_{\{\xi+\psi_1+\psi_2<Q_1+Q_2+Q_3+S-M\}}].
\end{aligned} \tag{A.53}
$$

Taking the expectation by integrating we obtain

$$
\begin{aligned}
\frac{\partial V}{\partial L_2} =& -(h+\delta+r)(1-F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2)) \\
&+ (h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&- (\rho_u+\theta)\int 1_{\{x+y>Q_1+L_1+Q_2+L_2\}}\cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\
&+ (\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+ (\rho_u+\theta)\cdot 0.
\end{aligned} \tag{A.54}
$$

We find the integral to be equal to

$$
\begin{aligned}
=& -(\rho_u+\theta)\int 1_{\{x+y>Q_1+L_1+Q_2+L_2\}}\cdot 1_{\{x+y+z<Q_1+Q_2+Q_3+S-M\}} f_{\xi,\psi_1,\psi_2}(x,y,z)d(x,y,z) \\
=& -(\rho_u+\theta)\int_{z=-\infty}^{z=Q_3+S-M-L_1-L_2}\int_{x+y=Q_1+L_1+Q_2+L_2}^{x+y=Q_1+Q_2+Q_3+S-M-z} f_{\xi+\psi_1}(x+y)f_{\psi_2}(z)d(x+y)dz \\
=& -(\rho_u+\theta)\int_{z=-\infty}^{z=Q_3}\int_{x+y=Q_1+L_1+Q_2+L_2}^{x+y=Q_1+Q_2+Q_3+S-M-z} f_{\xi+\psi_1}(x+y)f_{\psi_2}(z)d(x+y)dz \\
=& -(\rho_u+\theta)\int_{z=-\infty}^{z=Q_3}(F_{\xi+\psi_1}(Q_1+Q_2+Q_3+S-M-z) - F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2))f_{\psi_2}(z)dz \\
=& (\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1+\psi_2}(Q_1+Q_2+Q_3+S-M)).
\end{aligned} \tag{A.55}
$$

Inserting this into (A.53) and rearranging the terms we find the derivative of $V(M, L_1, L_2)$ w.r.t. L_2 is given by

$$
\begin{aligned}
\frac{\partial V}{\partial L_2} =& \delta + (h+\delta+r+\rho_u+\theta)F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) \\
&- (h+2\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3) = 0.
\end{aligned} \tag{A.56}
$$

We see (A.52) and (A.56) both depend on L_1 and L_2, which means we will find a line on which $V(M, L_1, L_2)$ is minimized. We can substitute (A.56) into (A.52) to find the derivative of $V(M, L_1, L_2)$ w.r.t. L_1 is given by

$$\frac{\partial V}{\partial L_1} = \delta + (h + r + \rho_u + \theta)F_\xi(Q_1 + L_1) - (h + \delta + r + \rho_u + \theta)F_{\xi+\psi_1}(Q_1 + L_1 + Q_2) = 0. \tag{A.57}$$

This proves our result found in the proposition.

□

A.6 Lemma 4.2

If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$, where Q_2 and Q_3 is the length of the second and third price-level queues and Y_1 and Y_2 are independent and exponentially distributed with rates λ_{ψ_1} and λ_{ψ_2}, the cumulative distribution function of the convolution of the random variables ξ, ψ_1 and ψ_2 is given by

$$F_{\xi+\psi_1+\psi_2}(A) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} & , \ A \le Q_2 + Q_3 \\ 1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} e^{-\lambda_\xi(A-(Q_2+Q_3))} & , \ A > Q_2 + Q_3 \end{cases} \tag{A.58}$$

Proof. For the exponential distribution the CDF is given by $F_X(x) = 1 - e^{-\lambda_X x}$. We set $\xi \sim X$, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$ where X is exponentially distributed with rate λ_ξ. In general, we have $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are exponentially distributed with rate λ_{ψ_i}. So for ψ_i we have

$$f_{\psi_i}(\psi_i = y) = \lambda_{\psi_i} e^{-\lambda_{\psi_i}(Q_i - y)} \tag{A.59}$$

meaning the CDF is given by

$$F_{\psi_i}(\psi_i = y) = e^{-\lambda_{\psi_i}(Q_{i+1} - y)} \tag{A.60}$$

where $y \in (-\infty, Q_{i+1}]$.We know from the previous section that for the convolution of ξ and ψ_1 we have

$$F_{\xi+\psi_1}(z) = \begin{cases} \frac{\lambda_\xi}{\lambda_\xi+\lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2-z)} & , \ z \le Q_2 \\ 1 - \frac{\lambda_{\psi_1}}{\lambda_\xi+\lambda_{\psi_1}} e^{-\lambda_\xi(z-Q_2)} & , \ z > Q_2 \end{cases} \tag{A.61}$$

which means for the PDF we have

$$f_{\xi+\psi_1}(z) = \begin{cases} \frac{\lambda_\xi\lambda_{\psi_1}}{\lambda_\xi+\lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2-z)} & , \ z \le Q_2 \\ \frac{\lambda_\xi\lambda_{\psi_1}}{\lambda_\xi+\lambda_{\psi_1}} e^{-\lambda_\xi(z-Q_2)} & , \ z > Q_2. \end{cases} \tag{A.62}$$

To find the CDF of the convolution of $\xi + \psi_1$ and ψ_2, $F_{\xi+\psi_1+\psi_2}(A)$, we must look at two different regions: $A \le Q_2 + Q_3$ and $A > Q_2 + Q_3$. By definition, the convoluted CDF is given by

$$F_{\xi+\psi_1+\psi_2}(A) = \int_{-\infty}^{\infty} F_{\psi_2}(A - z) f_{\xi+\psi_1}(z) dz. \tag{A.63}$$

Through the derivation in section B.1, we find that the convolution of ξ, ψ_1 and ψ_2 is given by

$$F_{\xi+\psi_1+\psi_2}(A) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} & , \ A \le Q_2 + Q_3 \\ 1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} e^{-\lambda_\xi(A-(Q_2+Q_3))} & , \ A > Q_2 + Q_3 \end{cases} \tag{A.64}$$

assuming that $\lambda_{\psi_1} \neq \lambda_{\psi_2}$. The inverse of this function cannot be expressed explicitly for $x > F_{\xi+\psi_1+\psi_2}(Q_2+Q_3) = 1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}$, due to the different exponents in the expression of the CDF. However, for $x \leq F_{\xi+\psi_1+\psi_2}(Q_2+Q_3) = 1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}$, the inverse of $F_{\xi+\psi_1+\psi_2}(A)$ can be expressed explicitly, and is given by

$$F^{-1}_{\xi+\psi_1+\psi_2}(x) = Q_2 + Q_3 + \frac{1}{\lambda_\xi}\log\left(\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})(1-x)}\right). \tag{A.65}$$

This still allows us to find an analytical expression for the solution described in Proposition 4.4. We see the optimal number of market orders $M^\star$ is given by

$$M^\star = S - \frac{1}{\lambda_\xi}\log\left(\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}\frac{h+2\delta+r+\rho_u+\theta}{-h-f+\rho_u+\theta}\right) + Q_1. \tag{A.66}$$

From section 4.2.3 we know the optimal solution shifts from limit orders at the best (first) price-level to the second price-level at the point $\lambda_\xi = \frac{\lambda_{\psi_1}\delta}{h+r+\rho_u+\theta}$. To see when the optimal solution shifts to the third price-level, we look at equation (4.39) for $L_2^\star$ and find that the following must hold for an optimal order allocation:

$$\begin{aligned}
&\delta + (h+\delta+r+\rho_u+\theta)\Big(1 - \frac{\lambda_{\psi_1}}{\lambda_\xi+\lambda_{\psi_1}}e^{-\lambda_\xi((Q_1+Q_2+L_1+L_2)-Q_2)}\Big) \\
&-(h+2\delta+r+\rho_u+\theta)\Big(1 - \frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}e^{-\lambda_\xi((Q_1+Q_2+Q_3+L_1+L_2)-(Q_2+Q_3))}\Big) = 0,
\end{aligned} \tag{A.67}$$

since $Q_1+Q_2+L_1+L_2 \geq Q_2$ and $Q_1+Q_2+Q_3+L_1+L_2 \geq Q_2+Q_3$. We find that equation (A.67) holds if

$$\begin{aligned}
0 &= -e^{-\lambda_\xi(Q_1+L_1+L_2)}\left((h+\delta+r+\rho_u+\theta)\frac{\lambda_{\psi_1}}{\lambda_\xi+\lambda_{\psi_1}} - (h+2\delta+r+\rho_u+\theta)\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}\right), \\
\frac{\lambda_\xi+\lambda_{\psi_2}}{\lambda_{\psi_2}} &= \frac{h+2\delta+r+\rho_u+\theta}{h+\delta+r+\rho_u+\theta}, \\
\frac{\lambda_\xi}{\lambda_{\psi_2}} &= \frac{\delta}{h+\delta+r+\rho_u+\theta}, \\
\lambda_\xi &= \frac{\lambda_{\psi_2}\delta}{h+\delta+r+\rho_u+\theta}.
\end{aligned} \tag{A.68}$$

We find that if $\lambda_\xi > \frac{\lambda_{\psi_2}\delta}{h+\delta+r+\rho_u+\theta}$, the derivative $\frac{\partial V}{\partial L_2} > 0$, so it is no longer optimal to place L_2 limit orders and the trader should instead place L_3 limit orders. We again find that the point at which the optimal limit order depth switches from one price-level to the next can be expressed as a closed-form critical point.

□

A.7 Lemma 4.3

If $\xi \sim X$, where X is exponentially distributed with rate λ_ξ and $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are independent and exponentially distributed with rate λ_{ψ_i}, the cumulative distribution function of the convolution of the random variables $\xi, \psi_1, ...\psi_i$ for $i = 1, 2, ...n-1$ is given by

$$F_{\xi+\psi_1+...+\psi_i}(A) = \begin{cases} \sum_{k=1}^{i} \frac{\lambda_\xi \prod_{j\neq k}^{i}\lambda_{\psi_j}}{(\lambda_\xi+\lambda_{\psi_k})\prod_{j\neq k}^{i}(\lambda_{\psi_j}-\lambda_{\psi_k})} e^{-\lambda_{\psi_k}((\sum_{k=2}^{i+1}Q_k)-A)} & , \; A \leq \sum_{k=2}^{i+1} Q_k \\ 1 - \frac{\prod_{k=1}^{i}\lambda_{\psi_k}}{\prod_{k=1}^{i}(\lambda_\xi+\lambda_{\psi_k})} e^{-\lambda_\xi(A-(\sum_{k=2}^{i+1}Q_k))} & , \; A > \sum_{k=2}^{i+1} Q_k \end{cases} \tag{A.69}$$

Proof. For the exponential distribution the CDF is given by $F_X(x) = 1 - e^{-\lambda_X x}$. We set $\xi \sim X$, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$ where X is exponentially distributed with rate λ_ξ. In general, we have $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are exponentially distributed with rate λ_{ψ_i}. So for ψ_i we have that the PDF is given by

$$f_{\psi_i}(\psi_i = y) = \lambda_{\psi_i} e^{-\lambda_{\psi_i}(Q_i - y)} \tag{A.70}$$

meaning the CDF is given by

$$F_{\psi_i}(\psi_i = y) = e^{-\lambda_{\psi_i}(Q_{i+1} - y)}, \text{ with } y \in (-\infty, Q_{i+1}]. \tag{A.71}$$

From section 4.3.1, we know that the convolution of ξ, ψ_1 and ψ_2 is given by

$$F_{\xi+\psi_1+\psi_2}(A) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi + \lambda_{\psi_2})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} & , \ A \leq Q_2 + Q_3 \\ 1 - \frac{\lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_\xi + \lambda_{\psi_2})} e^{-\lambda_\xi(A - (Q_2+Q_3))} & , \ A > Q_2 + Q_3 \end{cases} \tag{A.72}$$

assuming that $\lambda_{\psi_1} \neq \lambda_{\psi_2}$. This means the PDF is given by

$$f_{\xi+\psi_1+\psi_2}(A) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_2})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-z)} - \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-z)} & , \ z \leq Q_2 + Q_3 \\ \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_\xi + \lambda_{\psi_2})} e^{-\lambda_\xi(z - (Q_2+Q_3))} & , \ z > Q_2 + Q_3. \end{cases} \tag{A.73}$$

To find the cumulative distribution function of the convolution of $\xi + \psi_1 + \psi_2$ and ψ_3, $F_{\xi+\psi_1+\psi_2+\psi_3}(A)$, we must look at two different regions: $A \leq Q_2 + Q_3 + Q_4$ and $A > Q_2 + Q_3 + Q_4$. By definition, the convoluted CDF is given by

$$F_{\xi+\psi_1+\psi_2+\psi_3}(A) = \int_{-\infty}^{\infty} F_{\psi_3}(A - z) f_{\xi+\psi_1+\psi_2}(z) dz. \tag{A.74}$$

We first look at the region $A \leq Q_2 + Q_3 + Q_4$, which means $A - Q_4 \leq Q_2 + Q_3$, so we must split the integral at these points. Therefore, the convoluted CDF is given by

$$\begin{aligned}
F_{\xi+\psi_1+\psi_2+\psi_3}(A) =& \int_{-\infty}^{A-Q_4} F_{\psi_3}(A-z) f_{\xi+\psi_1+\psi_2}(z)dz + \int_{A-Q_4}^{Q_2+Q_3} F_{\psi_3}(A-z) f_{\xi+\psi_1+\psi_2}(z)dz \\
&+ \int_{Q_2+Q_3}^{\infty} F_{\psi_3}(A-z) f_{\xi+\psi_1+\psi_2}(z)dz \\
=& F_{\xi+\psi_1+\psi_2}(A-Q_4) + \int_{A-Q_4}^{Q_2+Q_3} e^{-\lambda_{\psi_3}(Q_4-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-z)} dz \\
&- \int_{A-Q_4}^{Q_2+Q_3} e^{-\lambda_{\psi_3}(Q_4-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-z)} dz \\
&+ \int_{Q_2+Q_3}^{\infty} e^{-\lambda_{\psi_3}(Q_4-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} e^{-\lambda_\xi(z-(Q_2+Q_3))} dz \\
=& \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3+Q_4-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3+Q_4-A)} \\
&+ \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} \frac{1}{\lambda_{\psi_2}-\lambda_{\psi_3}} \left[e^{-\lambda_{\psi_3}(Q_4-(A-z))} e^{-\lambda_{\psi_2}(Q_2+Q_3-z)} \right]_{z=A-Q_4}^{z=Q_2+Q_3} \\
&- \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} \frac{1}{\lambda_{\psi_1}-\lambda_{\psi_3}} \left[e^{-\lambda_{\psi_3}(Q_4-(A-z))} e^{-\lambda_{\psi_1}(Q_2+Q_3-z)} \right]_{z=A-Q_4}^{z=Q_2+Q_3} \\
&+ \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} \frac{-1}{\lambda_\xi+\lambda_{\psi_3}} \left[e^{-\lambda_{\psi_3}(Q_4-(A-z))} e^{-\lambda_\xi(z-(Q_2+Q_3))} \right]_{z=Q_2+Q_3}^{z=\infty} \\
=& \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3+Q_4-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3+Q_4-A)} \\
&+ \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} \frac{1}{\lambda_{\psi_2}-\lambda_{\psi_3}} \left[e^{-\lambda_{\psi_3}(Q_2+Q_3+Q_4-A)} - e^{-\lambda_{\psi_2}(Q_2+Q_3+Q_4-A)} \right] \\
&- \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} \frac{1}{\lambda_{\psi_1}-\lambda_{\psi_3}} \left[e^{-\lambda_{\psi_3}(Q_2+Q_3+Q_4-A)} - e^{-\lambda_{\psi_1}(Q_2+Q_3+Q_4-A)} \right] \\
&+ \frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})} \frac{-1}{\lambda_\xi+\lambda_{\psi_3}} \left[0 - e^{-\lambda_{\psi_3}(Q_2+Q_3+Q_4-A)} \right].
\end{aligned} \tag{A.75}$$

So for $A \leq Q_2 + Q_3 + Q_4$, the CDF of the convolution of ξ, ψ_1, ψ_2 and ψ_3 is given by

$$\begin{aligned}
F_{\xi+\psi_1+\psi_2+\psi_3}(A) =& \left(\frac{\lambda_\xi \lambda_{\psi_2} \lambda_{\psi_3}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_2}-\lambda_{\psi_1})(\lambda_{\psi_3}-\lambda_{\psi_1})} \right) e^{-\lambda_{\psi_1}(Q_2+Q_3+Q_4-A)} \\
&+ \left(\frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_3}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})(\lambda_{\psi_3}-\lambda_{\psi_2})} \right) e^{-\lambda_{\psi_2}(Q_2+Q_3+Q_4-A)} \\
&+ \left(\frac{\lambda_\xi \lambda_{\psi_1} \lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_3})(\lambda_{\psi_1}-\lambda_{\psi_3})(\lambda_{\psi_2}-\lambda_{\psi_3})} \right) e^{-\lambda_{\psi_3}(Q_2+Q_3+Q_4-A)}.
\end{aligned} \tag{A.76}$$

We can now clearly see the pattern start to emerge. For the general convolution of $\xi + \psi_1 + ... + \psi_i$, the CDF is given by

$$F_{\xi+\psi_1+...+\psi_i}(A) = \begin{cases} \sum_{k=1}^{i} \frac{\lambda_\xi \prod_{j\neq k}^{i} \lambda_{\psi_j}}{(\lambda_\xi+\lambda_{\psi_k}) \prod_{j\neq k}^{i}(\lambda_{\psi_j}-\lambda_{\psi_k})} e^{-\lambda_{\psi_k}((\sum_{k=2}^{i+1} Q_k)-A)} & , \ A \leq \sum_{k=2}^{i+1} Q_k \\ 1 - \frac{\prod_{k=1}^{i} \lambda_{\psi_k}}{\prod_{k=1}^{i}(\lambda_\xi+\lambda_{\psi_k})} e^{-\lambda_\xi(A-(\sum_{k=2}^{i+1} Q_k))} & , \ A > \sum_{k=2}^{i+1} Q_k. \end{cases} \tag{A.77}$$

□

A.8 Theorem 3.1

Dynamic Programming Principle for Optimal Stopping and Control. The value function $H(t, \boldsymbol{x})$ satisfies the DPP

$$H(t,\boldsymbol{x}) = \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}\Big[\Big(H(\theta,\boldsymbol{X}^u_\theta) + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}\Big] \tag{A.78}$$

for all $(t,\boldsymbol{x}) \in [0,T]\times\mathbb{R}^m$ and all stopping times $\theta \leq T$, where $G(X^u_\tau)$ is the reward function and $F(s,\boldsymbol{X}^u_s)$ is the running penalty.

Proof. As we know, for the value function is given by

$$H^{\tau,u}(t,\boldsymbol{x}) = \mathbb{E}_{t,\boldsymbol{x}}[G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds] \tag{A.79}$$

where $G(\boldsymbol{X}^u_\tau)$ is the reward upon exercise and $F(s,\boldsymbol{X}^u_s)$ is the running penalty. Choosing another arbitrary stopping time θ, we can rewrite the performance criterion as

$$\begin{aligned}
H^{\tau,u}(t,\boldsymbol{x}) =&\mathbb{E}_{t,\boldsymbol{x}}[G(\boldsymbol{X}^u_\tau) + \Big(\int_\theta^\tau F(s,\boldsymbol{X}^u_s)ds + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta} + \Big(\int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}]\\
=&\mathbb{E}_{t,\boldsymbol{x}}[\Big(G(\boldsymbol{X}^u_\tau) + \int_\theta^\tau F(s,\boldsymbol{X}^u_s)ds + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta}\\
&+\Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}]\\
=&\mathbb{E}_{t,\boldsymbol{x}}[\Big(\mathbb{E}_{\theta,\boldsymbol{x}}[G(\boldsymbol{X}^u_\tau) + \int_\theta^\tau F(s,\boldsymbol{X}^u_s)ds] + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta}\\
&+\Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}]\\
=&\mathbb{E}_{t,\boldsymbol{x}}[\Big(H^{\tau,u}(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}]
\end{aligned} \tag{A.80}$$

Now, for any admissible control and stopping time, $H(t,\boldsymbol{x}) \geq H^{\tau,u}(t,\boldsymbol{x})$, which means we can put the following bound on the equation above

$$H^{\tau,u}(t,\boldsymbol{x}) \leq\mathbb{E}_{t,\boldsymbol{x}}[\Big(H(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}] \tag{A.81}$$

Taking the supremum over all stopping times and controls left and right gives us the first inequality

$$H(t,\boldsymbol{x}) \leq \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}[\Big(H(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}] \tag{A.82}$$

To obtain the reverse inequality, we take an ϵ-optimal control (depth), denoted by $v^\epsilon \in \mathcal{A}$, such that

$$H(t,\boldsymbol{x}) \geq H^{v^\epsilon}(t,\boldsymbol{x}) \geq H(t,\boldsymbol{x}) - \epsilon \tag{A.83}$$

where $\epsilon > 0$. Such a control exists of the value function is continuous in the space of controls. Consider its modification up to time τ:

$$\tilde{v}^\epsilon = u_t 1_{t\leq\theta} + v^\epsilon 1_{t>\theta} \tag{A.84}$$

where $u \in \mathcal{A}$ is an arbitrary admissible control. We then have that

$$\begin{aligned} H(t,\boldsymbol{x}) \geq & H^{\tau,\tilde{v}^\epsilon}(t,\boldsymbol{x}) \\ = & \mathbb{E}_{t,\boldsymbol{x}}[\Big(H^{\tau,\tilde{v}^\epsilon}(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}_s^{\tilde{v}^\epsilon})ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}_\tau^{\tilde{v}^\epsilon}) + \int_t^\tau F(s,\boldsymbol{X}_s^{\tilde{v}^\epsilon})ds\Big)1_{\tau<\theta}] \\ = & \mathbb{E}_{t,\boldsymbol{x}}[\Big(H^{\tau,v^\epsilon}(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}_s^{u})ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}_\tau^{u^\epsilon}) + \int_t^\tau F(s,\boldsymbol{X}_s^{u^\epsilon})ds\Big)1_{\tau<\theta}] \\ \geq & \mathbb{E}_{t,\boldsymbol{x}}[\Big(H(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}_s^{u})ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}_\tau^{u^\epsilon}) + \int_t^\tau F(s,\boldsymbol{X}_s^{u^\epsilon})ds\Big)1_{\tau<\theta}] - \epsilon \end{aligned} \tag{A.85}$$

where the first equality follows from (A.80), the second using (A.84) and the final inequality follows from (A.83). Since the above holds for every δ and ϵ, it holds for the supremum and we can take the limit as $\epsilon \to 0$ and end up with the reverse inequality

$$H(t,\boldsymbol{x}) \geq \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}[\Big(H(\theta,\boldsymbol{X}_\theta) + \int_t^\theta F(s,\boldsymbol{X}_s^{u})ds\Big)1_{\tau\geq\theta} + \Big(G(\boldsymbol{X}_\tau^{u}) + \int_t^\tau F(s,\boldsymbol{X}_s^{u})ds\Big)1_{\tau<\theta}]. \tag{A.86}$$

Combining inequality (A.82) and (A.86) gives us the equation in the theorem. □

A.9 Theorem 3.2

Dynamic Programming Equation for Stopping and Control Problems. Assume that the value function $H(t,\boldsymbol{x})$ is once differentiable in t and all seconder-order derivatives in $\boldsymbol{x}$ exist, i.e. $H \in C^{1,2}([0,T],\mathbb{R}^m)$ and that and $F : \mathbb{R}^m \to \mathbb{R}_+$ are continuous. Then H solves the **quasi-variational inequality (QVI)**,

$$\max\Big(\partial_t H + \sup_{u\in\mathcal{A}_t}\big(\mathcal{L}_t^u H + F\big)\ ;\ G - H\Big) = 0, \quad \text{on } \mathcal{D}, \tag{A.87}$$

where $\mathcal{D} = [0,T]\times\mathbb{R}^m$ and the generator of the process $\mathcal{L}_t$ acts on twice differentiable functions as follows:

$$\mathcal{L}_t h(t,x) = \boldsymbol{\mu}(t,\boldsymbol{x})\cdot\boldsymbol{D}_x h(t,x) + \frac{1}{2}\mathbf{Tr}(\boldsymbol{\sigma}(t,\boldsymbol{x})\boldsymbol{\sigma}(t,\boldsymbol{x})'\boldsymbol{D}_x^2 h(t,\boldsymbol{x})) + \sum_{j=1}^{p}\lambda_j(t,\boldsymbol{x})[h(t,\boldsymbol{x}+\boldsymbol{\gamma}_{\cdot j}(t,\boldsymbol{x})) - h(t,\boldsymbol{x})] \tag{A.88}$$

where $\boldsymbol{\gamma}_{\cdot j}$ represents the j-th column of $\boldsymbol{\gamma}$, $\boldsymbol{D}_x h$ represents the vector of partial derivatives w.r.t. x, and $\boldsymbol{D}_x^2 h$ represents the matrix of second-order partial derivatives w.r.t. x.

Proof. The proof is analogous to that described in [8] and is broken up into two steps by showing (on $\mathcal{D}$) that (i) the left-hand side is first smaller than or equal to zero, and that (ii) by contradiction the left-hand side is also greater than or equal to zero. Therefore, equality must hold.

(i) First, we show that

$$\max\Big(\partial_t H + \sup_{u\in\mathcal{A}_t}\big(\mathcal{L}_t^u H + F\big)\ ;\ G - H\Big) \leq 0, \quad \text{on } \mathcal{D}. \tag{A.89}$$

To show this, we first note that a constant stopping time $\tau = t$ (i.e. stopping immediately) is an admissible strategy. Therefore, by the DPP, the performance criteria with this constant stopping rule equals G and we require $H \geq G$ so that $G - H \leq 0$. Now take any point $(t_0, x_0) \in \mathcal{D}$ and we will show that the inequality holds at this point. Consider the sequence of stopping times equal to the minimum of $t_0 + h$ and the time it takes $\boldsymbol{X}_t$ to exit a ball of size 1 from its current position $\boldsymbol{x}_0$, so

$$\theta_h = \inf\left(t > t_0 : (t, ||\boldsymbol{X}_t - \boldsymbol{x}_0||) \notin [t_0, t_0 + h] \times 1\right), \quad h > 0 \tag{A.90}$$

where $||\cdot||$ denotes the Euclidean norm. As long as $h < T - t_0$ this is an admissible stopping time. Therefore, by taking $\tau = t_0$, the DPP (that tells us H is the maximum value over all stopping times) implies that

$$H(t_0, \boldsymbol{x}_0) \geq \mathbb{E}_{t_0,\boldsymbol{x}_0}[H(\theta_h, \boldsymbol{X}_{\theta_h}) + \int_{t_0}^{\theta_h} F(s, \boldsymbol{X}_s)ds]. \tag{A.91}$$

We then expand $H(\theta_h, \boldsymbol{X}_{\theta_h})$ using Ito's lemma for jump-diffusions to write

$$\begin{aligned} H(\theta_h, \boldsymbol{X}_{\theta_h}) = H(t_0, \boldsymbol{x}_0) & \\ + \int_{t_0}^{\theta_h} & \{\partial_t H(t, \boldsymbol{X}_t) + \mathcal{L}_t H(t, \boldsymbol{X}_t)\}\, dt \\ + \int_{t_0}^{\theta_h} & \left(\sigma(t, \boldsymbol{X}_t) D_{\boldsymbol{x}} H(t, \boldsymbol{X}_t)'\right) dW_t \\ + \sum_{j=1}^{p} \int_{t_0}^{\theta_h} & \left[H\left(t, \boldsymbol{X}_t + \boldsymbol{\gamma}_{\cdot j}(t, \boldsymbol{X}_t)\right) - H(t, \boldsymbol{X}_t)\right] d\widehat{\boldsymbol{N}}_t^j, \end{aligned}$$

where

$$\widehat{\boldsymbol{N}}_t = \boldsymbol{N}_t - \int_0^t \lambda(s, \boldsymbol{X}_s)ds$$

are the compensated versions of the counting process, and $\boldsymbol{\gamma}_{\cdot j}$ denotes the j^{th} column of γ.

Since the stopping time θ_h is chosen so that the process $\boldsymbol{X}$ remains bounded by the ball of size 1 around $\boldsymbol{x}_0$ plus the potential of a jump (which we assume bounded), it follows that the stochastic integrals with respect to both the Brownian motions and the compensated counting processes vanish under the expectation. Hence, we have

$$0 \geq \mathbb{E}_{t_0,\boldsymbol{x}_0}\left[\int_{t_0}^{\theta_h} \{\partial_t H(t, \boldsymbol{X}_t) + \mathcal{L}_t H(t, \boldsymbol{X}_t) + F(t, \boldsymbol{X}_t)\}\, dt\right].$$

Dividing by h and taking $h \searrow 0$, in which case $\theta_h \searrow t_0 + h$ a.s. (since $\boldsymbol{X}_t$ will a.s. not hit the edge of the ball), the Fundamental Theorem of Calculus implies that

$$\lim_{h\to 0} \frac{1}{h} \int_t^{t+h} f(s)ds = f(t) \tag{A.92}$$

which means that we get

$$\begin{aligned} 0 \geq & \lim_{h\searrow 0} \mathbb{E}_{t_0,\boldsymbol{x}_0}\left[\frac{1}{h}\int_{t_0}^{t_0+h} \{\partial_t H(t, \boldsymbol{X}_t) + \mathcal{L}_t H(t, \boldsymbol{X}_t) + F(t, \boldsymbol{X}_t)\}\, dt\right] \\ = & \ \partial_t H(t_0, \boldsymbol{x}_0) + \mathcal{L}_t H(t_0, \boldsymbol{x}_0) + F(t_0, \boldsymbol{x}_0). \end{aligned} \tag{A.93}$$

This completes the first part of the proof.

(ii) Next, we establish by contradiciton that

$$\max\left(\partial_t H + \sup_{u\in\mathcal{A}_t}\left(\mathcal{L}_t^u H + F\right)\ ;\ G - H\right) \geq 0, \quad \text{on } \mathcal{D}. \tag{A.94}$$

If this inequality doesn't hold on $\mathcal{D}$, then there exists a point $(t_0, \boldsymbol{x}_0) \in \mathcal{D}$ such that

$$G(t_0, \boldsymbol{x}_0) - H(t_0, \boldsymbol{x}_0) < 0 \quad \textbf{and} \quad (\partial_t + \mathcal{L}_t)H(t_0, \boldsymbol{x}_0) < 0 \tag{A.95}$$

since $F(t, \boldsymbol{x}) \geq 0$ on $\mathcal{D}$. We also used simplified notation that differentiates between the two cases, first for **stopping immediately** at $\tau = t_0$, meaning the traders receives the reward, so

$$H(t_0, \boldsymbol{x}_0) = \sup_{u\in\mathcal{A}} \mathbb{E}_{t_0,\boldsymbol{x}_0}[G(\boldsymbol{X}_\tau^u) + \int_{t_0}^{\tau} F(s, \boldsymbol{X}_s^u)ds] = G(\boldsymbol{X}_\tau) \tag{A.96}$$

and **continuing** until at least $t_0 + h$, then deciding optimally thereafter, so

$$H(t_0, \boldsymbol{x}_0) = \sup_{u\in\mathcal{A}} \mathbb{E}_{t_0,\boldsymbol{x}_0}[H(t_0 + h, \boldsymbol{X}_{t_0+h}^u) + \int_{t_0}^{t+h} F(s, \boldsymbol{X}_s^u)ds] = H(t_0 + h, \boldsymbol{X}_{t_0+h}). \tag{A.97}$$

We will show that (A.95) contradicts the DPP. We introduce a new function φ_ϵ which approximates the value function near $(t_0, \boldsymbol{x}_0)$ but locally dominates it:

$$\varphi_\epsilon(t, \boldsymbol{x}) := H(t, \boldsymbol{x}) + \epsilon(||\boldsymbol{x} - \boldsymbol{x}_0||^4 + |t - t_0|^2), \quad \forall(t, \boldsymbol{x}) \in \mathcal{D}, \quad \epsilon > 0. \tag{A.98}$$

Under assumption (A.95), for $\epsilon > 0$ but sufficiently small, there is a small neighbourhood around $(t_0, \boldsymbol{x}_0)$ in which the value function is at least δ larger than the reward G (since by (A.95) $G(t_0, \boldsymbol{x}_0) < H(t_0, \boldsymbol{x}_0)$) and for which the operator $(\partial_t + \mathcal{L}_t)$ renders the approximation φ_ϵ negative. Mathematically put, there exists $h > 0$ and $\delta > 0$ such that

$$H \geq G + \delta \quad \text{and} \quad (\partial_t + \mathcal{L}_t)\varphi_\epsilon \leq 0 \quad \text{on} \quad \mathcal{D}_h := [t_0, t_0 + h] \times \mathcal{B}_h \tag{A.99}$$

where $\mathcal{B}_h$ is a ball of size h around $\boldsymbol{x}_0$. Also, near $(t_0, \boldsymbol{x}_0)$ (on the boundary $\partial\mathcal{D}_h$ of $\mathcal{D}_h$), φ_ϵ is locally larger than H, hence

$$-\zeta := \max_{\partial\mathcal{D}_h}(H - \varphi_\epsilon) < 0. \tag{A.100}$$

We now take a stopping time equal to the first time the process exits this ball, namely

$$\theta := \inf\left(t > t_0 : (t, \boldsymbol{X}_t) \notin \mathcal{D}_h\right). \tag{A.101}$$

Take a second stopping rule, this time arbitrary, $\tau \in \mathcal{T}_{[t,T]}$, and let $\psi = \tau \wedge \theta$. Then we have

$$H(\psi, \boldsymbol{X}_\psi) - H(t_0, \boldsymbol{x}_0) = (H - \varphi_\varepsilon)(\psi, \boldsymbol{X}_\psi) + (\varphi_\varepsilon(\psi, \boldsymbol{X}_\psi) - \varphi_\varepsilon(t_0, \boldsymbol{x}_0)), \tag{A.102}$$

since φ_ε and H coincide at $(t_0, \boldsymbol{x}_0)$. From Ito's lemma, and the fact that $\boldsymbol{X}_\psi$ is bounded due to stopping the first time $\boldsymbol{X}$ exits the ball $\mathcal{B}_h$, we have

$$\mathbb{E}_{t_0,\boldsymbol{x}_0}\left[\varphi_\varepsilon(\psi, \boldsymbol{X}_\psi) - \varphi_\varepsilon(t_0, \boldsymbol{x}_0)\right] = \mathbb{E}_{t_0,\boldsymbol{x}_0}\left[\int_{t_0}^{\psi} (\partial_t + \mathcal{L}_t)\varphi_\varepsilon(t, \boldsymbol{X}_t)dt\right] \leq 0. \tag{A.103}$$

The diffusive and jump terms vanish because they are martingales, and the inequality follows from the second inequality in (A.99). Hence, putting this together with (A.102), we have

$$\mathbb{E}_{t_0,\boldsymbol{x}_0}[H(\psi, X_\psi) - H(t_0, \boldsymbol{x}_0)] \leq \mathbb{E}_{t_0,\boldsymbol{x}_0}[(H - \varphi_\varepsilon)(\psi, \boldsymbol{X}_\psi)] \leq -\zeta\mathbb{P}(\tau \geq \theta), \tag{A.104}$$

where the second inequality follows from (A.100). By rearranging to isolate $H(t_0, \boldsymbol{x}_0)$, we have

$$\begin{aligned} H(t_0, \boldsymbol{x}_0) \geq & \zeta \mathbb{P}(\tau \geq \theta) + \mathbb{E}_{t_0,\boldsymbol{x}_0}[H(\psi, \boldsymbol{X}_\psi)] \\ = & \zeta \mathbb{P}(\tau \geq \theta) + \mathbb{E}_{t_0,\boldsymbol{x}_0}[H(\tau, \boldsymbol{X}_\tau)\mathbb{1}_{\tau<\theta} + H(\theta, \boldsymbol{X}_\theta)\mathbb{1}_{\tau\geq\theta}]. \end{aligned} \tag{A.105}$$

By the first inequality in (A.99), $H \geq G + \delta$ on $\mathcal{D}_h$, so therefore

$$H(t_0, \boldsymbol{x}_0) \geq \zeta \mathbb{P}(\tau \geq \theta) + \mathbb{E}_{t_0,\boldsymbol{x}_0}[(G(\boldsymbol{X}_\tau) + \delta)1_{\tau<\theta} + H(\theta, \boldsymbol{X}_\theta)1_{\tau\geq\theta}] \tag{A.106}$$

where we have replaced $H(\tau, \boldsymbol{X}_\tau)$ by its lower bound $(G(\boldsymbol{X}_\tau) + \delta)$ on $\{\tau < \theta\}$, since in that event we are still in $\mathcal{D}_h$. Finally, since $\mathbb{E}_{t_0,\boldsymbol{x}_0}[\mathbb{1}_{\tau<\theta}] = \mathbb{P}(\tau < \theta)$, we have

$$H(t_0, \boldsymbol{x}_0) \geq \zeta \mathbb{P}(\tau \geq \theta) + \delta \mathbb{P}(\tau < \theta) + \mathbb{E}_{t_0,\boldsymbol{x}_0}[G(\boldsymbol{X}_\tau)1_{\tau<\theta} + H(\theta, \boldsymbol{X}_\theta)1_{\tau\geq\theta}]. \tag{A.107}$$

Setting $\theta = t_0 + h$ and inserting (A.96) and (A.97), we find

$$\begin{aligned} H(t_0, \boldsymbol{x}_0) \geq & \zeta \mathbb{P}(\tau \geq \theta) + \delta \mathbb{P}(\tau < \theta) \\ & + \sup_{u \in \mathcal{A}} \mathbb{E}_{t_0,\boldsymbol{x}_0}\Big[\left(G(\boldsymbol{X}^u_\tau) + \int_{t_0}^{\tau} F(s, \boldsymbol{X}^u_s) ds\right)1_{\tau<\theta} + \left(H(\theta, \boldsymbol{X}^u_\theta) + \int_{t_0}^{\theta} F(s, \boldsymbol{X}^u_s) ds\right)1_{\tau\geq\theta}\Big]. \end{aligned} \tag{A.108}$$

By the arbitrariness of $\tau \in \mathcal{T}_{[t,T]}$ (so it is also true for the optimal stopping time), and the fact that the constants added to the expectation above are positive (and both probabilities can't be zero), we arrive at a contradiction to the DPP and the proof is complete.

□

B | Appendix

B.1 Distribution function of the convoluted order flows

For the exponential distribution the CDF is given by $F_X(x) = 1 - e^{-\lambda_X x}$. We set $\xi \sim X$, $\psi_1 \sim Q_2 - Y_1$ and $\psi_2 \sim Q_3 - Y_2$ where X is exponentially distributed with rate λ_ξ. In general, we have $\psi_i \sim Q_{i+1} - Y_i$, where Q_i is the length of our i-th price-level queue and Y_i are exponentially distributed with rate λ_{ψ_i}. Based on the definition, the probability density function (PDF) of ψ_i is given by

$$f_{\psi_i}(\psi_i = y) = \lambda_{\psi_i} e^{-\lambda_{\psi_i}(Q_i - y)} \tag{B.1}$$

meaning the CDF is given by

$$F_{\psi_i}(\psi_i = y) = e^{-\lambda_{\psi_i}(Q_{i+1} - y)}, \tag{B.2}$$

where $y \in (-\infty, Q_{i+1}]$. We know from the previous section that for the convolution of ξ and ψ_1 we have

$$F_{\xi+\psi_1}(z) = \begin{cases} \frac{\lambda_\xi}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2 - z)} & , \quad z \leq Q_2 \\ 1 - \frac{\lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(z - Q_2)} & , \quad z > Q_2 \end{cases} \tag{B.3}$$

which means the PDF of $\xi + \psi_1$ is given by

$$f_{\xi+\psi_1}(z) = \begin{cases} \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2 - z)} & , \quad z \leq Q_2 \\ \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(z - Q_2)} & , \quad z > Q_2. \end{cases} \tag{B.4}$$

To find the CDF of the convolution of $\xi + \psi_1$ and ψ_2, $F_{\xi+\psi_1+\psi_2}(A)$, we must look at two different regions: $A \leq Q_2 + Q_3$ and $A > Q_2 + Q_3$. The convoluted CDF is given by

$$F_{\xi+\psi_1+\psi_2}(A) = \int_{-\infty}^{\infty} F_{\psi_2}(A - z) f_{\xi+\psi_1}(z) dz. \tag{B.5}$$

First we analyse $A \leq Q_2 + Q_3$. By the bounds on the argument of F_{ψ_2}, when $z \leq A - Q_3$ we have $F_{\psi_2} = 1$. For $f_{\xi+\psi_1}$, we split the integral at $z = Q_2$, and since $A \leq Q_2 + Q_3$, $A - Q_3 \leq Q_2$, so all-in-all we split at the integral at $z = A - Q_3$ and $z = Q_2$. We obtain

$$\begin{aligned}
F_{\xi+\psi_1+\psi_2}(A) =& \int_{z=-\infty}^{z=A-Q_3} 1 \cdot f_{\xi+\psi_1}(z)dz + \int_{z=A-Q_3}^{z=Q_2} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz + \int_{z=Q_2}^{z=\infty} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz \\
=& F_{\xi+\psi_1}(A-Q_3) + \int_{z=A-Q_3}^{z=Q_2} e^{-\lambda_{\psi_2}(Q_3-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2-z)} dz \\
&+ \int_{z=Q_2}^{z=\infty} e^{-\lambda_{\psi_2}(Q_3-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(z-Q_2)} dz \\
=& \frac{\lambda_\xi}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_3+Q_2-A)} + \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \left[\frac{1}{\lambda_{\psi_1} - \lambda_{\psi_2}} e^{-\lambda_{\psi_2}(Q_3-(A-z))} e^{-\lambda_{\psi_1}(Q_2-z)} \right]_{z=A-Q_3}^{z=Q_2} \\
&+ \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \left[\frac{-1}{\lambda_\xi + \lambda_{\psi_2}} e^{-\lambda_{\psi_2}(Q_3-(A-z))} e^{-\lambda_\xi(z-Q_2)} \right]_{z=Q_2}^{z=\infty} \\
=& \frac{\lambda_\xi}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} + \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \frac{1}{\lambda_{\psi_1} - \lambda_{\psi_2}} \left((e^{-\lambda_{\psi_2}(Q_2+Q_3-A)}) - (e^{-\lambda_{\psi_1}(Q_2+Q_3-A)}) \right) \\
&+ \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \frac{-1}{\lambda_\xi + \lambda_{\psi_2}} \left(0 - (e^{-\lambda_{\psi_2}(Q_2+Q_3-A)}) \right).
\end{aligned} \tag{B.6}$$

So for $A \leq Q_2 + Q_3$, the CDF of the convolution of ξ, ψ_1 and ψ_2 is given by

$$\begin{aligned}
F_{\xi+\psi_1+\psi_2}(A) =& \frac{\lambda_\xi}{\lambda_\xi + \lambda_{\psi_1}} \left(1 - \frac{\lambda_{\psi_1}}{\lambda_{\psi_1} - \lambda_{\psi_2}} \right) e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} \\
&+ \frac{\lambda_\xi}{\lambda_\xi + \lambda_{\psi_1}} \left(\frac{\lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_2}} + \frac{\lambda_{\psi_1}}{\lambda_{\psi_1} - \lambda_{\psi_2}} \right) e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} \\
=& \frac{\lambda_\xi \lambda_{\psi_1}}{(\lambda_\xi + \lambda_{\psi_2})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_2}(Q_2+Q_3-A)} - \frac{\lambda_\xi \lambda_{\psi_2}}{(\lambda_\xi + \lambda_{\psi_1})(\lambda_{\psi_1} - \lambda_{\psi_2})} e^{-\lambda_{\psi_1}(Q_2+Q_3-A)}.
\end{aligned} \tag{B.7}$$

For the region $A \geq Q_2 + Q_3$ we find

$$\begin{aligned}
F_{\xi+\psi_1+\psi_2}(A) =& \int_{z=-\infty}^{z=Q_2} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz + \int_{z=Q_2}^{z=A-Q_3} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz \\
&+ \int_{z=A-Q_3}^{z=\infty} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz \\
=& \int_{z=-\infty}^{z=Q_2} 1 \cdot f_{\xi+\psi_1}(z)dz + \int_{z=Q_2}^{z=A-Q_3} 1 \cdot f_{\xi+\psi_1}(z)dz + \int_{z=A-Q_3}^{z=\infty} F_{\psi_2}(A-z) f_{\xi+\psi_1}(z)dz \\
=& F_{\xi+\psi_1}(A-Q_3) + \int_{z=A-Q_3}^{z=\infty} e^{-\lambda_{\psi_2}(Q_3-(A-z))} \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(z-Q_2)} dz \\
=& 1 - \frac{\lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(A-(Q_2+Q_3))} + \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \left[\frac{-1}{\lambda_\xi + \lambda_{\psi_2}} e^{-\lambda_{\psi_2}(Q_3-(A-z))} e^{-\lambda_\xi(z-Q_2)} \right]_{z=A-Q_3}^{z=\infty} \\
=& 1 - \frac{\lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} e^{-\lambda_\xi(A-(Q_2+Q_3))} + \frac{\lambda_\xi \lambda_{\psi_1}}{\lambda_\xi + \lambda_{\psi_1}} \frac{-1}{\lambda_\xi + \lambda_{\psi_2}} \left(0 - e^{-\lambda_\xi(A-(Q_2+Q_3))} \right)
\end{aligned} \tag{B.8}$$

where for the CDF $F_{\xi+\psi_1}(A-Q_3)$, we have $A-Q_3>Q_2$. We find that in the region $A>Q_2+Q_3$, the CDF of the convolution of ξ,ψ_1 and ψ_2 is given by

$$\begin{aligned} F_{\xi+\psi_1+\psi_2}(A) =&1+\left(\frac{\lambda_\xi\lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}-\frac{\lambda_{\psi_1}(\lambda_\xi+\lambda_{\psi_2})}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}\right)e^{-\lambda_\xi(A-(Q_2+Q_3))} \\ =&1-\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}e^{-\lambda_\xi(A-(Q_2+Q_3))}. \end{aligned} \tag{B.9}$$

We see that these two cases indeed coincide at the point $A=Q_2+Q_3$, since the coefficent preceding the exponent in equation (B.9) can be rewritten to

$$\begin{aligned} &=\frac{\lambda_\xi\lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})}-\frac{\lambda_\xi\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})} \\ &=\frac{\lambda_\xi\lambda_{\psi_1}(\lambda_\xi+\lambda_{\psi_1})-\lambda_\xi\lambda_{\psi_2}(\lambda_\xi+\lambda_{\psi_2})}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} \\ &=\frac{\lambda_\xi^2\lambda_{\psi_1}+\lambda_\xi\lambda_{\psi_1}^2-\lambda_\xi^2\lambda_{\psi_2}-\lambda_\xi\lambda_{\psi_2}^2}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} \\ &=\frac{(\lambda_\xi^2+\lambda_\xi\lambda_{\psi_1}+\lambda_\xi\lambda_{\psi_2}+\lambda_{\psi_1}\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})-\lambda_{\psi_1}\lambda_{\psi_2}(\lambda_{\psi_1}-\lambda_{\psi_2})}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})} \\ &=1-\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}. \end{aligned} \tag{B.10}$$

So for the convolution of ξ, ψ_1 and ψ_2 we have

$$F_{\xi+\psi_1+\psi_2}(A)=\begin{cases}\frac{\lambda_\xi\lambda_{\psi_1}}{(\lambda_\xi+\lambda_{\psi_2})(\lambda_{\psi_1}-\lambda_{\psi_2})}e^{-\lambda_{\psi_2}(Q_2+Q_3-A)}-\frac{\lambda_\xi\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_{\psi_1}-\lambda_{\psi_2})}e^{-\lambda_{\psi_1}(Q_2+Q_3-A)} & ,\ A\le Q_2+Q_3 \\ 1-\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}e^{-\lambda_\xi(A-(Q_2+Q_3))} & ,\ A>Q_2+Q_3.\end{cases} \tag{B.11}$$

The inverse of this function cannot be expressed explicitly for $x>F_{\xi+\psi_1+\psi_2}(Q_2+Q_3)=1-\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}$, due to the different exponents in the expression of the CDF. For $x\le F_{\xi+\psi_1+\psi_2}(Q_2+Q_3)=1-\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})}$, however, the inverse is given by

$$F^{-1}_{\xi+\psi_1+\psi_2}(x)=Q_2+Q_3+\frac{1}{\lambda_\xi}\log\left(\frac{\lambda_{\psi_1}\lambda_{\psi_2}}{(\lambda_\xi+\lambda_{\psi_1})(\lambda_\xi+\lambda_{\psi_2})(1-x)}\right). \tag{B.12}$$

B.2 Static model: four price-levels

We work in a similar fashion as the previous sections, with the cost function defined as

$$\begin{aligned} v(\boldsymbol{X},\xi,\psi_1,\psi_2,\psi_3):=&(h+f)M-(h+r)\mathrm{OF}_1-(h+r+\delta)\mathrm{OF}_2-(h+r+2\delta)\mathrm{OF}_3-(h+r+3\delta)\mathrm{OF}_4 \\ &+\rho_u(S-A(X,\xi,\psi_1,\psi_2,\psi_3))_++\rho_o(A(X,\xi,\psi_1,\psi_2,\psi_3)-S)_+ \\ &+\theta(M+L_1+L_2+L_3+L_4+(S-A(X,\xi,\psi_1,\psi_2,\psi_3))_+) \\ =&(h+f)M-(h+r)\mathrm{OF}_1-(h+r+\delta)\mathrm{OF}_2-(h+r+2\delta)\mathrm{OF}_3-(h+r+3\delta)\mathrm{OF}_4 \\ &+(\rho_u+\theta)(S-\mathrm{OF}_1-\mathrm{OF}_2-\mathrm{OF}_3-\mathrm{OF}_4)_++\rho_o(\mathrm{OF}_1+\mathrm{OF}_2+\mathrm{OF}_3+\mathrm{OF}_4-S)_+ \\ &+\theta(M+L_1+L_2+L_3+L_4) \end{aligned} \tag{B.13}$$

with the number of orders filled at price-level i defined as

$$\begin{aligned}\mathrm{OF}_i =&(\xi + \sum_{k=1}^{i-1}\psi_k - \sum_{k=1}^{i-1}(Q_k+L_k) - Q_i)\cdot 1_{\{\sum_{k=1}^{i-1}(Q_k+L_k)+Q_i<\xi+\sum_{k=1}^{i}\psi_k<\sum_{k=1}^{i}(Q_k+L_k)\}} \\ &+ L_i \cdot 1_{\{\xi+\sum_{k=1}^{i}\psi_k>\sum_{k=1}^{i}(Q_k+L_k)\}}.\end{aligned} \tag{B.14}$$

We find the expected cost function given by

$$\begin{aligned}V(M,L_1,L_2,L_3) :=&\mathbb{E}[(h+f)M - (h+r)(\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}} - (h+r)L_1\cdot 1_{\{\xi\geq Q_1+L_1\}} \\ &- (h+\delta+r)(\xi+\psi_1-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\ &- (h+\delta+r)L_2\cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} \\ &- (h+2\delta+r)(\xi+\psi_1+\psi_2-Q_1-L_1-Q_2-L_2-Q_3) \\ &\quad\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\ &- (h+2\delta+r)L_3\cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\ &- (h+3\delta+r)(\xi+\psi_1+\psi_2+\psi_3-Q_1-L_1-Q_2-L_2-Q_3-L_3-Q_4) \\ &\quad\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\ &- (h+3\delta+r)(S-M-L_1-L_2-L_3)\cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3>Q_1+Q_2+Q_3+Q_4+S-M\}} \\ &+ (\rho_u+\theta)(\ S-(M+(\xi-Q_1)1_{\{Q_1<\xi<Q_1+L_1\}}+L_1\cdot 1_{\{\xi\geq Q_1+L_1\}} \\ &\quad + (\xi+\psi_1-Q_1-L_1-Q_2)\cdot 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\ &\quad + L_2\cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} \\ &\quad + (\xi+\psi_1+\psi_2-Q_1-L_1-Q_2-L_2-Q_3) \\ &\qquad\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\ &\quad + L_3\cdot 1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+L_3\}}) \\ &\quad + (\xi+\psi_1+\psi_2+\psi_3-Q_1-L_1-Q_2-L_2-Q_3-L_3-Q_4) \\ &\qquad\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\ &\quad + (S-M-L_1-L_2-L_3)\cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3>Q_1+Q_2+Q_3+Q_4+S-M\}}\ ) \\ &\cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} + \theta S].\end{aligned} \tag{B.15}$$

We then take the derivative of V w.r.t. M, to find

$$M^\star = S - F^{-1}_{\xi+\psi_1+\psi_2+\psi_3}\left(\frac{2h+f+r+3\delta}{h+r+3\delta+\rho_u+\theta}\right) + Q_1+Q_2+Q_3+Q_4. \tag{B.16}$$

Next, we take the derivative of V w.r.t. L_1, to find

$$\begin{aligned}
\frac{\partial V}{\partial L_1} =& -(h+r)(1-F_\xi(Q_1+L_1)) + (h+\delta+r)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (h+2\delta+r)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+ (h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+ (\rho_u+\theta)(-1_{\{\xi\geq Q_1+L_1\}} + 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} \\
&\quad + 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\
&\quad + 1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
&\quad + 1_{\{\xi+\psi_1+\psi_2+\psi_3>Q_1+Q_2+Q_3+Q_4+S-M\}}) \\
&\quad \cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
=& -(h+r)(1-F_\xi(Q_1+L_1)) + (h+\delta+r)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (h+2\delta+r)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+ (h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+ (\rho_u+\theta)(-1_{\{\xi\geq Q_1+L_1\}} \cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
&\quad + 1_{\{Q_1+L_1+Q_2<\xi+\psi_1<Q_1+L_1+Q_2+L_2\}} + 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\
&\quad + 1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}})
\end{aligned} \tag{B.17}$$

where in the second equality we used the fact that $S = M + L_1 + L_2 + L_3 + L_4$ and $\psi_i \leq Q_i$, which means that for example $Q_1 + L_1 + Q_2 + L_2 + Q_3 + L_3 < Q_1 + Q_2 + Q_3 + Q_4 + S - M$. Using a method similar to the previous sections to rewrite the integral we are left with, we obtain

$$\begin{aligned}
\frac{\partial V}{\partial L_1} =& -(h+r)(1-F_\xi(Q_1+L_1)) + (h+\delta+r)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (h+2\delta+r)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+ (h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+ (\rho_u+\theta)((F_\xi(Q_1+L_1) - F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M)) \\
&\quad + (F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - (F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&\quad + (F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&\quad + (F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M) \\
&\quad - F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4))) \\
=& 3\delta + (h+r+\rho_u+\theta)F_\xi(Q_1+L_1) \\
&+ (h+\delta+r+\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) - F_{\xi+\psi_1}(Q_1+L_1+Q_2)) \\
&+ (h+2\delta+r+\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) \\
&\quad - F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&- (h+3\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4).
\end{aligned} \tag{B.18}$$

Next, we take the derivative of V w.r.t. L_2, to find

$$
\begin{aligned}
\frac{\partial V}{\partial L_2} =& -(h+\delta+r)\cdot 1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}} \\
&+(h+2\delta+r)\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\
&+(h+3\delta+r)(1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
&\quad+1_{\{\xi+\psi_1+\psi_2+\psi_3>Q_1+Q_2+Q_3+Q_4+S-M\}}) \\
&+(\rho_u+\theta)(-1_{\{\xi+\psi_1>Q_1+L_1+Q_2+L_2\}}+1_{\{Q_1+L_1+Q_2+L_2+Q_3<\xi+\psi_1+\psi_2<Q_1+L_1+Q_2+L_2+Q_3+L_3\}} \\
&\quad+1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}})\cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
=& -(h+\delta+r)(1-F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2)) \\
&+(h+2\delta+r)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3)-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+(h+3\delta+r)(F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M) \\
&\quad-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+(h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M)) \\
&+(\rho_u+\theta)(F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2)-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M)) \\
&+(\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3)-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+(\rho_u+\theta)(F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M) \\
&\quad-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4))
\end{aligned}
\tag{B.19}
$$

and obtain

$$
\begin{aligned}
\frac{\partial V}{\partial L_2} =& 2\delta+(h+\delta+r+\rho_u+\theta)F_{\xi+\psi_1}(Q_1+L_1+Q_2+L_2) \\
&+(h+2\delta+r+\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3) \\
&\quad-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&-(h+3\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4).
\end{aligned}
\tag{B.20}
$$

Lastly, we take the derivative of V w.r.t. L_3, to find

$$
\begin{aligned}
\frac{\partial V}{\partial L_3} =& -(h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+(h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+(\rho_u+\theta)(-1_{\{\xi+\psi_1+\psi_2>Q_1+L_1+Q_2+L_2+Q_3+L_3\}}\cdot 1_{\{\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}}) \\
&+(\rho_u+\theta)\cdot 1_{\{Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4<\xi+\psi_1+\psi_2+\psi_3<Q_1+Q_2+Q_3+Q_4+S-M\}} \\
=& -(h+2\delta+r)(1-F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3)) \\
&+(h+3\delta+r)(1-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4)) \\
&+(\rho_u+\theta)(F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3+L_3)-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M)) \\
&+(\rho_u+\theta)(F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S-M) \\
&\quad-F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4))
\end{aligned}
\tag{B.21}
$$

and we find

$$
\begin{aligned}
\frac{\partial V}{\partial L_3} =& \delta+(h+2\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2}(Q_1+L_1+Q_2+L_2+Q_3) \\
&-(h+3\delta+r+\rho_u+\theta)F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+L_1+Q_2+L_2+Q_3+L_3+Q_4).
\end{aligned}
\tag{B.22}
$$

Substituting (B.22) into (B.20) and (B.20) into (B.18) we find the optimal allocation to be described in the following proposition.

Proposition B.1. *Assume that ξ, ψ_1, ψ_2, and ψ_3 have continuous distributions and Assumptions (A1-A3) hold. Then the optimal allocations for the Order Placement Problem with 4 price-levels satisfy the following:*

(i) If $\rho_u \geq \overline{\rho_u} = \frac{2h+3\delta+f+r}{F\xi+\psi_1+\psi_2+\psi_3(Q_1+Q_2+Q_3+Q_4)} - (h+3\delta+r+\theta)$, *then* $(M^*, L_1^*, L_2^*, L_3^*, L_4^*) = (S, 0, 0, 0, 0)$ *is an optimal allocation.*

(ii) If $\rho_u \leq \underline{\rho_u} = \frac{2h+3\delta+f+r}{F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1+Q_2+Q_3+Q_4+S)} - (h+3\delta+r+\theta)$, *then* $M^* = 0$ *and,* L_1^* *satisfies*

$$\delta + (h + r + \rho_u + \theta)F_\xi(Q_1 + L_1^\star) - (h + \delta + r + \rho_u + \theta)F_{\xi+\psi_1}(Q_1 + L_1^\star + Q_2) = 0, \tag{B.23}$$

L_2^* *satisfies*

$$\begin{aligned} &\delta + (h + \delta + r + \rho_u + \theta)F_{\xi+\psi_1}(Q_1 + L_1^\star + Q_2 + L_2^\star) \\ &\quad - (h + 2\delta + r + \rho_u + \theta)F_{\xi+\psi_1+\psi_2}(Q_1 + L_1^\star + Q_2 + L_2^\star + Q_3) = 0, \end{aligned} \tag{B.24}$$

L_3^* *satisfies*

$$\begin{aligned} &\delta + (h + 2\delta + r + \rho_u + \theta)F_{\xi+\psi_1+\psi_2}(Q_1 + L_1^\star + Q_2 + L_2^\star + Q_3 + L_3^\star) \\ &\quad - (h + 3\delta + r + \rho_u + \theta)F_{\xi+\psi_1+\psi_2+\psi_3}(Q_1 + L_1^\star + Q_2 + L_2^\star + Q_3 + L_3^\star + Q_4) = 0, \end{aligned} \tag{B.25}$$

and L_4^* *satisfies*

$$L_4^\star = S - M - L_1^\star - L_2^\star - L_3^\star. \tag{B.26}$$

(iii) If $\rho_u \in (\underline{\rho_u}, \overline{\rho_u})$, *then the optimal allocation is a mix of limit* $(L_1^*, L_2^*, L_3^*, L_4^*)$ *and market* $(M^\star)$ *orders, where* $L_1^\star$ *satisfies (B.23),* $L_2^\star$ *satisfies (B.24),* $L_3^\star$ *satisfies (B.25),* $L_4^\star$ *satisfies (B.26) and* $M^\star$ *satisfies*

$$M^\star = S - F^{-1}_{\xi+\psi_1+\psi_2+\psi_3}\left(\frac{2h + 3\delta + f + r}{h + 3\delta + r + \rho_u + \theta}\right) + Q_1 + Q_2 + Q_3 + Q_4, \tag{B.27}$$

where $F_\xi(x) = \mathbb{P}(\xi \leq x)$ *is the distribution of the outflow at the first price-level* ξ, $F_{\xi+\psi_1}(x) = \mathbb{P}(\xi+\psi_1 \leq x)$ *is the distribution of the outflow at the first and second price-level* ξ, ψ_1, $F_{\xi+\psi_1+\psi_2}(x) = \mathbb{P}(\xi + \psi_1 + \psi_2 \leq x)$ *is the distribution of the outflow at the first, second and third price-level* ξ, ψ_1,ψ_2, $F_{\xi+\psi_1+\psi_2+\psi_3}(x) = \mathbb{P}(\xi + \psi_1 + \psi_2 + \psi_3 \leq x)$ *is the distribution of the outflow at the first, second, third and fourth price-level* ξ, ψ_1,ψ_2,ψ_3 *and* $F_\xi^{-1}, F^{-1}_{\xi+\psi_1}, F^{-1}_{\xi+\psi_1+\psi_2}, F^{-1}_{\xi+\psi_1+\psi_2+\psi_3}$ *their left-inverses.*

The methods for finding the root of (B.23), (B.24) and (B.25) are discussed in section 3.1.

B.3 ODE solving for the complete dynamic model

B.3.1 Using the DPP to derive the HJB equation (DPE)

As we have shown in the Preliminaries section, a vector of controlled processes $\boldsymbol{X}^u$ is controlled using the control process $u = (u_t)_{0\leq t\leq T}$, and satisfies the SDE

$$d\boldsymbol{X}_t^u = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t + \boldsymbol{\gamma}_t^u d\boldsymbol{N}_t^u \tag{B.28}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_t)_{0\leq t\leq T} \in \mathbb{R}^m$ is the drift, $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_t)_{0\leq t\leq T} \in \mathbb{R}^m_+$ is the volatility and $\boldsymbol{W} = (\boldsymbol{W}_t)_{0\leq t\leq T} \in \mathbb{R}^m$ a collection of independent Brownian motions. Furthermore, $\boldsymbol{N}^u = (\boldsymbol{N}_t^u)_{0\leq t\leq T} \in \mathbb{N}^n$ denotes a collection of counting processes with controlled intensities $\boldsymbol{\lambda}^u = (\boldsymbol{\lambda}^u_t)_{0\le t\le T} \in \mathbb{R}^n$, and $\boldsymbol{\gamma}^u_t := \boldsymbol{\gamma}(t, u_t) \in \mathbb{R}^{n\times m}$ denotes the controlled jump size. From the DPP we know the value function H is given by

$$H(t,\boldsymbol{x}) = \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{u\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\boldsymbol{x}}\Big[\Big(H(\theta,\boldsymbol{X}^u_\theta) + \int_t^\theta F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau\ge\theta} + \Big(G(\boldsymbol{X}^u_\tau) + \int_t^\tau F(s,\boldsymbol{X}^u_s)ds\Big)1_{\tau<\theta}\Big]. \tag{B.29}$$

Then our value function H solves the **quasi-variational inequality (QVI)**,

$$\max\Big(\partial_t H + \sup_{u\in\mathcal{A}_t}\big(\mathcal{L}^u_t H + F\big)\ ;\ G - H\Big) = 0, \quad \text{on } \mathcal{D}, \tag{B.30}$$

where $\mathcal{D} = [0,T]\times\mathbb{R}^m$ and the generator of the process $\mathcal{L}_t$ acts on twice differentiable functions as follows:

$$\mathcal{L}_t h(t,x) = \boldsymbol{\mu}(t,\boldsymbol{x})\cdot\boldsymbol{D}_x h(t,x) + \frac{1}{2}\mathbf{Tr}(\boldsymbol{\sigma}(t,\boldsymbol{x})\boldsymbol{\sigma}(t,\boldsymbol{x})'\boldsymbol{D}^2_x h(t,\boldsymbol{x})) + \sum_{j=1}^p \lambda_j(t,\boldsymbol{x})[h(t,\boldsymbol{x}+\boldsymbol{\gamma}_{\cdot j}(t,\boldsymbol{x})) - h(t,\boldsymbol{x})] \tag{B.31}$$

where $\boldsymbol{\gamma}_{\cdot j}$ represents the j-th column of $\boldsymbol{\gamma}$, $\boldsymbol{D}_x h$ represents the vector of partial derivatives w.r.t. x, and $\boldsymbol{D}^2_x h$ represents the matrix of second-order partial derivatives w.r.t. x.

In our model of the Optimal Placement Problem the vector valued process is denoted by $(\boldsymbol{V}^{\boldsymbol{\tau},\delta}_t)_{0\le t\le T}$ and is controlled by the control function δ, which describes the depth the trader places her LOs, and $\boldsymbol{\tau}$ which describes the stopping times at which the trader must place her MOs. The controlled processes are then described by the SDE

$$d\boldsymbol{V}^{\boldsymbol{\tau},\delta}_t = d\begin{pmatrix} X^{\boldsymbol{\tau},\delta}_t \\ p_t \\ I^{\boldsymbol{\tau},\delta}_t \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \\ 0 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma_p \\ 0 \end{pmatrix} dW_t + \begin{pmatrix} p_t + \frac{\Delta}{2} + \delta_t + r & p_t - \frac{\Delta}{2} - f \\ 0 & 0 \\ -1 & -1 \end{pmatrix} d\begin{pmatrix} L^\delta_t \\ M^{\boldsymbol{\tau}}_t \end{pmatrix}. \tag{B.32}$$

We now define the trader's performance criteria, which is given by

$$H^{\boldsymbol{\tau},\delta}(t,x,p,i) = \mathbb{E}_{t,x,p,i}[X^{\boldsymbol{\tau},\delta}_T + I^{\boldsymbol{\tau},\delta}_T(p_T - \frac{\Delta}{2} - f - \rho_u(I^{\boldsymbol{\tau},\delta}_T)) - \int_t^T \theta(I^{\boldsymbol{\tau},\delta}_s)ds] \tag{B.33}$$

where the notation $\mathbb{E}_{t,x,p,q}[\cdot]$ represents the expectation conditional on t, $X^{\boldsymbol{\tau},\delta}_{t^-} = x, p_t = p$ and $I^{\boldsymbol{\tau},\delta}_{t^-} = i$. Our control variables are the stopping times τ and depth δ, which means that the set of admissible strategies $\mathcal{A}$ consists of seeking over all $\mathcal{F}$-stopping times τ and the set of $\mathcal{F}$-predictable, bounded from below, depths δ. In this case the value function is given by

$$H(t,x,p,i) = \sup_{(\boldsymbol{\tau},\delta)\in\mathcal{A}} H^{\boldsymbol{\tau},\delta}(t,x,p,i) \tag{B.34}$$

where $\mathcal{A} = [t,T]^i \times [0,\infty)$.

In this case the generator of the process is given by

$$\begin{aligned}\mathcal{L}_t H(t,x,p,i) = \begin{pmatrix} 0 \\ \mu \\ 0 \end{pmatrix}\cdot\begin{pmatrix} \partial_x \\ \partial_p \\ \partial_i \end{pmatrix} H(t,x,p,i) + \frac{1}{2}\begin{pmatrix} 0 \\ \sigma_p^2 \\ 0 \end{pmatrix}\cdot\begin{pmatrix} \partial_{xx} \\ \partial_{pp} \\ \partial_{ii} \end{pmatrix} H(t,x,p,i) \\ + \lambda_\nu e^{\kappa\delta}\big[H(t, x + p_t + \frac{\Delta_t}{2} + \delta + r, p, i-1) - H(t,x,p,i)\big].\end{aligned} \tag{B.35}$$

The performance criterion, for each $\mathcal{F}$-stopping time τ_i, given by

$$H^{\boldsymbol{\tau},\delta}(t,x,p,i) = \mathbb{E}_{t,x,p,i}[G(\tau_i,x,p,i) + \int_t^{\tau_i} F(s,x,p,i)ds] \tag{B.36}$$

where $G(t,x,p,i)$ is the reward upon exercise, which is, in our case, given by

$$\begin{aligned} G(t,x,p,i) =& x_t + i_t(p_t - \frac{\Delta}{2} - f) = H(t, x + (p - \frac{\Delta}{2} - f), p, i-1), \\ G(T,x,p,i) =& x_T + i_T(p_T - \frac{\Delta}{2} - f - \rho_u(i_T)), \end{aligned} \tag{B.37}$$

and $F(s,x,p,i)$ is the running inventory penalty given by

$$F(s,x,p,i) = -\theta(i_s). \tag{B.38}$$

Now, using the Dynamic Programming Principle, we can find the Hamiltonian-Jacobi-Bellman equation which tells us that H solves the quasi-variational inequality (QVI)

$$\begin{aligned} 0 = \max \Bigg\{ & \partial_t H + \mu \partial_p H + \frac{1}{2}\sigma_p^2 \partial_{pp} H - \theta(i) \\ & + \sup_{\delta} \lambda_\nu e^{-\kappa\delta} \left[H(t, x + (p + \frac{\Delta}{2} + \delta + r), p, i-1) - H(t,x,p,i) \right] \; ; \\ & \left[H(t, x + (p - \frac{\Delta}{2} - f), p, i-1) - H(t,x,p,i) \right] \Bigg\}, \end{aligned} \tag{B.39}$$

with boundary and terminal conditions

$$\begin{aligned} H(t,x,p,0) &= x, \\ H(T,x,p,i) &= x + i(p - \frac{\Delta}{2} - f - \rho_u(i)). \end{aligned} \tag{B.40}$$

B.3.2 Solving the ODE for the $i = 1$ case

Let us assume that fees, rebates and spread are constant, then the problem we are solving becomes

$$\begin{aligned} \max \Big(\partial_t w(t,1) + \kappa(\mu - \theta(1)) w(t,1) + \tilde{\lambda}_\nu \; ; \; 1 - w(t,1) \Big) &= 0 \\ w(T,1) &= e^{-\kappa \rho_u(1)} \end{aligned} \tag{B.41}$$

where $\tilde{\lambda}_\nu = \frac{\lambda_\nu}{e} e^{\kappa(\Delta + r + f)}$. Using the obvious optimal strategy of never stopping at terminal time T, we see that we have to solve the ODE

$$\partial_t g_1(t) + \kappa(\mu - \theta(1)) g_1(t) + \tilde{\lambda}_\nu = 0, \quad g_1(T^-) = 1. \tag{B.42}$$

We rewrite the ODE as

$$\begin{aligned} e^{\kappa(\mu-\theta(1))t} \partial_t g_1(t) + \kappa(\mu - \theta(1)) e^{\kappa(\mu-\theta(1))t} g_1(t) =& - \tilde{\lambda}_\nu e^{\kappa(\mu-\theta(1))t} \\ \partial_t \left[e^{\kappa(\mu-\theta(1))t} g_1(t) \right] =& - \tilde{\lambda}_\nu e^{\kappa(\mu-\theta(1))t} \\ \int_t^T \partial_u \left[e^{\kappa(\mu-\theta(1))u} g_1(u) \right] du =& \int_t^T -\tilde{\lambda}_\nu e^{\kappa(\mu-\theta(1))u} du \\ e^{\kappa(\mu-\theta(1))T} g_1(T) - e^{\kappa(\mu-\theta(1))t} g_1(t) =& \frac{-\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} \left(e^{\kappa(\mu-\theta(1))T} - e^{\kappa(\mu-\theta(1))t} \right) \end{aligned} \tag{B.43}$$

and find the solution

$$g_1(t) = \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} \left(e^{\kappa(\mu-\theta(1))(T-t)} - 1 \right) + e^{\kappa(\mu-\theta(1))(T-t)}. \tag{B.44}$$

We see that $g_1(t) \geq 1$ for all $t \in (0, T)$ when

$$\frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} \left(e^{\kappa(\mu-\theta(1))(T-t)} - 1\right) \geq 1 - e^{\kappa(\mu-\theta(1))(T-t)} \tag{B.45}$$

we must first distinguish for $\left(e^{\kappa(\mu-\theta(1))(T-t)} - 1\right) > 0$ (i.e. for $\mu > \theta(1)$), which is the case when

$$\begin{aligned} \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} &\geq -1 \\ \frac{-\tilde{\lambda}_\nu}{\kappa} &\leq (\mu - \theta(1)) \\ \mu &\geq \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1) \end{aligned} \tag{B.46}$$

and for $\mu < \theta(1)$, which is the case when

$$\begin{aligned} \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} &\leq -1 \\ \frac{-\tilde{\lambda}_\nu}{\kappa} &\leq (\mu - \theta(1)) \\ \mu &\geq \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1). \end{aligned} \tag{B.47}$$

We see both cases yield the same bound on μ. When this bound is satisfied the continuation value is always greater than the execution value, and the solution ot the QVI (B.41) is

$$w(t, 1) = g_1(t) \cdot 1_{\{t<T\}} + e^{-\kappa\rho_u(1)} \cdot 1_{\{t=T\}} \tag{B.48}$$

and it is never optimal to execute a market order expect for an instant before the terminal time. If, on the other hand, $\mu < \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1)$ the continuation value is always less than the execution value, leading to the solution

$$w(t, 1) = 1_{\{t<T\}} + e^{-\kappa\rho_u(1)} \cdot 1_{\{t=T\}}. \tag{B.49}$$

B.3.3 Solving the ODE for the $i = 2$ case

We try to find an explicit expression for τ_2, by solving

$$\begin{aligned} w(\tau_2, 1) = \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} \left(e^{\kappa(\mu-\theta(1))(T-\tau_2)} - 1\right) + e^{\kappa(\mu-\theta(1))(T-\tau_2)} &= \tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1} \\ e^{\kappa(\mu-\theta(1))(T-\tau_2)} \left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} + 1\right) - \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} &= \tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1} \\ e^{\kappa(\mu-\theta(1))(T-\tau_2)} \left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} + 1\right) &= \tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} \\ \tau_2 = T - \log\left[\left(\tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))}\right)\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu - \theta(1))} + 1\right)^{-1}\right] &\kappa(\mu - \theta(1))^{-1}. \end{aligned} \tag{B.50}$$

Cleaning up the expression, we find the optimal stopping time

$$\tau_2 = T - \frac{\log\left[\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu+\theta(1)-\theta(2))+\tilde{\lambda}_\nu} + \frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))}\right)\right] - \log\left[\left(\frac{\tilde{\lambda}_\nu}{\kappa(\mu-\theta(1))} + 1\right)\right]}{\kappa(\mu - \theta(1))}. \tag{B.51}$$

We find that, to avoid impossible values inside the logarithm, the following bound must hold:

$$\mu \geq \frac{-\tilde{\lambda}_\nu}{\kappa} + \theta(1) - \theta(2). \tag{B.52}$$

Now to find the full solution of $w(t,2)$, we solve the ODE

$$\partial_t g_2(t) + \kappa(2\mu - \theta(2))g_2(t) + \tilde{\lambda}_\nu g_1(t), \quad g_2(\tau_2) = \Gamma \tag{B.53}$$

where $\Gamma = 1 \cdot w(\tau_2, 1) = \tilde{\lambda}_\nu \left(\kappa(\mu + \theta(1) - \theta(2)) + \tilde{\lambda}_\nu\right)^{-1}$. The solution is found by

$$\begin{aligned} \partial_t g_2(t) + \kappa(2\mu - \theta(2))g_2(t) &= -\tilde{\lambda}_\nu g_1(t) \\ \partial_t[e^{\kappa(2\mu-\theta(2))t} g_2(t)] &= -e^{\kappa(2\mu-\theta(2))t}\tilde{\lambda}_\nu g_1(t) \\ e^{\kappa(2\mu-\theta(2))\tau_2} g_2(\tau_2) - e^{\kappa(2\mu-\theta(2))t} g_2(t) &= -\tilde{\lambda}_\nu \int_t^{\tau_2} e^{\kappa(2\mu-\theta(2))s} g_1(s) ds. \end{aligned} \tag{B.54}$$

For the integral, knowing $g_1'(t) = -\kappa(\mu - \theta(1))g_1(t) - \tilde{\lambda}_\nu$ we find

$$\begin{aligned} \int e^{\kappa(2\mu-\theta(2))s} g_1(s) ds &= \frac{1}{\kappa(2\mu - \theta(2))} e^{\kappa(2\mu-\theta(2))t} g_1(t) - \frac{1}{\kappa(2\mu - \theta(2))} \int e^{\kappa(2\mu-\theta(2))s} g_1'(s) ds \\ = \frac{1}{\kappa(2\mu - \theta(2))} e^{\kappa(2\mu-\theta(2))t} g_1(t) &- \frac{1}{\kappa(2\mu - \theta(2))} \left(-\kappa(\mu - \theta(1)) \int e^{\kappa(2\mu-\theta(2))s} g_1(s) ds - \tilde{\lambda}_\nu \int e^{\kappa(2\mu-\theta(2))s} ds \right) \\ &= \frac{1}{\kappa(2\mu - \theta(2))} e^{\kappa(2\mu-\theta(2))t} g_1(t) + \frac{\kappa(\mu - \theta(1))}{\kappa(2\mu - \theta(2))} \int e^{\kappa(2\mu-\theta(2))s} g_1(s) ds + \frac{\tilde{\lambda}_\nu}{\kappa^2(2\mu - \theta(2))^2} e^{\kappa(2\mu-\theta(2))t} \\ \left(1 - \frac{\kappa(\mu - \theta(1))}{\kappa(2\mu - \theta(2))}\right) \int e^{\kappa(2\mu-\theta(2))s} g_1(s) ds &= \frac{1}{\kappa(2\mu - \theta(2))} e^{\kappa(2\mu-\theta(2))t} g_1(t) + \frac{\tilde{\lambda}_\nu}{\kappa^2(2\mu - \theta(2))^2} e^{\kappa(2\mu-\theta(2))t} \\ \int e^{\kappa(2\mu-\theta(2))s} g_1(s) ds &= \frac{1}{\kappa(\mu + \theta(1) - \theta(2))} e^{\kappa(2\mu-\theta(2))t} g_1(t) + \frac{\tilde{\lambda}_\nu}{\kappa(2\mu - \theta(2))\kappa(\mu + \theta(1) - \theta(2))} e^{\kappa(2\mu-\theta(2))t} \end{aligned} \tag{B.55}$$

and using the notation $A_1 = \kappa(\mu - \theta(1))$ and $A_2 = \kappa(2\mu - \theta(2))$, we obtain

$$\begin{aligned} e^{A_2\tau_2} g_2(\tau_2) - e^{A_2 t} g_2(t) &= \frac{-\tilde{\lambda}_\nu}{A_2 - A_1} e^{A_2\tau_2} g_1(\tau_2) - \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)} e^{A_2\tau_2} + \frac{\tilde{\lambda}_\nu}{A_2 - A_1} e^{A_2 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)} e^{A_2 t} \\ -e^{A_2 t} g_2(t) &= e^{A_2\tau_2}\left(\frac{-\tilde{\lambda}_\nu}{A_2 - A_1} g_1(\tau_2) - \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)} - g_2(\tau_2)\right) + e^{A_2 t}\left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) \\ g_2(t) &= e^{A_2(\tau_2 - t)}\left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} g_1(\tau_2) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)} + g_2(\tau_2)\right) - \left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right). \end{aligned} \tag{B.56}$$

From the boundary condition of the ODE we know that $g_2(\tau_2) = g_1(\tau_2) = \frac{\tilde{\lambda}_\nu}{B - A + \tilde{\lambda}_\nu}$. Inserting this into the formula above we find

$$\begin{aligned} g_2(t) =& e^{A_2(\tau_2 - t)}\left(\left(1 + \frac{\tilde{\lambda}_\nu}{A_2 - A_1}\right) g_1(\tau_2) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) - \left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) \\ g_2(t) =& e^{A_2(\tau_2 - t)}\left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) - \left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right). \end{aligned} \tag{B.57}$$

B.3.4 Solving the ODE for the $i = 3$ case

We solve our main ODE

$$\partial_t g_3(t) + \kappa(3\mu - \theta(3))g_3(t) + \tilde{\lambda}_\nu g_2(t) = 0, \quad g_3(\tau_3) = \Gamma_3 \tag{B.58}$$

in the same way as before, using

$$\begin{aligned} \partial_t g_3(t) + \kappa(3\mu - \theta(3))g_3(t) =& -\tilde{\lambda}_\nu g_2(t) \\ \partial_t[e^{\kappa(3\mu-\theta(3))t} g_3(t)] =& -\tilde{\lambda}_\nu e^{\kappa(3\mu-\theta(3))t} g_2(t) \\ e^{\kappa(3\mu-\theta(3))\tau_3} g_3(\tau_3) - e^{\kappa(3\mu-\theta(3))t} g_3(t) =& -\tilde{\lambda}_\nu \int_t^{\tau_3} e^{\kappa(3\mu-\theta(3))s} g_2(s)ds. \end{aligned} \tag{B.59}$$

For the integral, remembering $g_2'(t) = -A_2 g_2(t) - \lambda g_1(t)$ and $g_1'(t) = -A_1 g_1(t) - \lambda$, we find

$$\begin{aligned} =& \int e^{\kappa(3\mu-\theta(3))s} g_2(s)ds \\ =& \frac{1}{\kappa(3\mu - \theta(3))} e^{\kappa(3\mu-\theta(3))t} g_2(t) - \frac{1}{\kappa(3\mu - \theta(3))} \int e^{\kappa(3\mu-\theta(3))s} g_2'(s)ds \\ =& \frac{1}{\kappa(3\mu - \theta(3))} e^{\kappa(3\mu-\theta(3))t} g_2(t) \\ &- \frac{1}{\kappa(3\mu - \theta(3))} \Big(-\kappa(2\mu - \theta(2)) \int e^{\kappa(3\mu-\theta(3))s} g_2(s)ds - \tilde{\lambda}_\nu \int e^{\kappa(3\mu-\theta(3))s} g_1(s)ds \Big) \\ =& \frac{1}{\kappa(3\mu - \theta(3))} e^{\kappa(3\mu-\theta(3))t} g_2(t) + \frac{\kappa(2\mu - \theta(2))}{\kappa(3\mu - \theta(3))} \int e^{\kappa(3\mu-\theta(3))s} g_2(s)ds + \frac{\tilde{\lambda}_\nu}{\kappa(3\mu - \theta(3))} \int e^{\kappa(3\mu-\theta(3))s} g_1(s)ds. \end{aligned} \tag{B.60}$$

Setting the first line equal to the last, and taking all terms containing the initial integral to one side, we obtain

$$\begin{aligned} \Big(1 - \frac{\kappa(2\mu - \theta(2))}{\kappa(3\mu - \theta(3))}\Big) \int e^{\kappa(3\mu-\theta(3))s} g_2(s)ds =& \frac{1}{\kappa(3\mu - \theta(3))} e^{\kappa(3\mu-\theta(3))t} g_2(t) + \frac{\tilde{\lambda}_\nu}{\kappa(3\mu - \theta(3))} \int e^{\kappa(3\mu-\theta(3))s} g_1(s)ds \\ \int e^{\kappa(3\mu-\theta(3))s} g_2(s)ds =& \frac{1}{\kappa(\mu + \theta(2) - \theta(3))} e^{\kappa(3\mu-\theta(3))t} g_2(t) + \frac{\tilde{\lambda}_\nu}{\kappa(\mu + \theta(2) - \theta(3))} \int e^{\kappa(3\mu-\theta(3))s} g_1(s)ds. \end{aligned} \tag{B.61}$$

From the previous subsection we know

$$\int e^{A_3 s} g_1(s)ds = \frac{1}{A_3 - A_1} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu}{A_3(A_3 - A_1)} e^{A_3 t} \tag{B.62}$$

where $A_1 = \kappa(\mu - \theta(1))$ and $A_3 = \kappa(3\mu - \theta(3))$. Remembering that $g_1'(t) = -A_1 g_1(t) - \tilde{\lambda}_\nu$ we find

$$\begin{aligned} \int e^{A_3 s} g_1(s)ds =& \frac{1}{A_3} e^{A_3 t} g_1(t) - \frac{1}{A_3} \int e^{A_3 s} g_1'(s)ds \\ =& \frac{1}{A_3} e^{A_3 t} g_1(t) - \frac{1}{A_3} \Big(-A_1 \int e^{A_3 s} g_1(s)ds - \tilde{\lambda}_\nu \int e^{A_3 s} ds \Big) \\ =& \frac{1}{A_3} e^{A_3 t} g_1(t) + \frac{A_1}{A_3} \int e^{A_3 s} g_1(s)ds + \frac{\tilde{\lambda}_\nu}{A_3^2} e^{A_3 t} \\ \Big(1 - \frac{A_1}{A_3}\Big) \int e^{A_3 s} g_1(s)ds =& \frac{1}{A_3} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu}{A_3^2} e^{A_3 t} \\ \int e^{A_3 s} g_1(s)ds =& \frac{1}{A_3 - A_1} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu}{A_3(A_3 - A_1)} e^{A_3 t}. \end{aligned} \tag{B.63}$$

From the previous result, we obtain

$$\begin{aligned}\int e^{A_3 s} g_2(s) ds =& \frac{1}{A_3 - A_2} e^{A_3 t} g_2(t) + \frac{\tilde{\lambda}_\nu}{A_3 - A_2} \left(\frac{1}{A_3 - A_1} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu}{A_3(A_3 - A_1)} e^{A_3 t} \right) \\ =& \frac{1}{A_3 - A_2} e^{A_3 t} g_2(t) + \frac{\tilde{\lambda}_\nu}{(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3 t}.\end{aligned} \tag{B.64}$$

Inserting (B.64) into our ODE, we obtain

$$\begin{aligned}e^{A_3\tau_3} g_3(\tau_3) - e^{A_3 t} g_3(t) =& -\tilde{\lambda}_\nu \Bigg(\Bigg(\frac{1}{A_3 - A_2} e^{A_3\tau_3} g_2(\tau_3) + \frac{\tilde{\lambda}_\nu}{(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3} g_1(\tau_3) \\ &+ \frac{\tilde{\lambda}_\nu^2}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3} \Bigg) - \Bigg(\frac{1}{A_3 - A_2} e^{A_3 t} g_2(t) \\ &+ \frac{\tilde{\lambda}_\nu}{(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} \Bigg) \Bigg) \\ =& \frac{\tilde{\lambda}_\nu}{A_3 - A_2} e^{A_3 t} g_2(t) + \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} g_1(t) + \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} \\ &- \frac{\tilde{\lambda}_\nu}{A_3 - A_2} e^{A_3\tau_3} g_2(\tau_3) - \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3} g_1(\tau_3) \\ &- \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3}.\end{aligned} \tag{B.65}$$

Continuing and inserting the boundary condition $g_3(\tau_3) = g_2(\tau_3) = \Gamma_3 = \tilde{\lambda}_\nu (A_3 - A_2 + \tilde{\lambda}_\nu)^{-1} g_1(\tau_3)$ into the above, we find

$$\begin{aligned}e^{A_3 t} g_3(t) =& \frac{\tilde{\lambda}_\nu}{A_3 - A_2} e^{A_3\tau_3} g_2(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3} g_1(\tau_3) + \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3\tau_3} \\ &- \frac{\tilde{\lambda}_\nu}{A_3 - A_2} e^{A_3 t} g_2(t) - \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} g_1(t) - \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3 t} + e^{A_3\tau_3} g_3(\tau_3) \\ g_3(t) =& (1 + \frac{\tilde{\lambda}_\nu}{A_3 - A_2}) e^{A_3(\tau_3 - t)} g_2(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} e^{A_3(\tau_3 - t)} g_1(\tau_3) \\ &+ \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} e^{A_3(\tau_3 - t)} - \frac{\tilde{\lambda}_\nu}{A_3 - A_2} g_2(t) - \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} g_1(t) \\ &- \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} \\ =& e^{A_3(\tau_3 - t)} \Bigg((1 + \frac{\tilde{\lambda}_\nu}{A_3 - A_2}) g_2(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} g_1(\tau_3) + \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} \Bigg) \\ &- \Bigg(\frac{\tilde{\lambda}_\nu}{A_3 - A_2} g_2(t) + \frac{\tilde{\lambda}_\nu^2}{(A_3 - A_2)(A_3 - A_1)} g_1(t) + \frac{\tilde{\lambda}_\nu^3}{A_3(A_3 - A_2)(A_3 - A_1)} \Bigg).\end{aligned} \tag{B.66}$$

Let us now try to express the stopping time τ_3 explicitly. We know τ_3 satisfies the equation

$$g_2(\tau_3) = \frac{\tilde{\lambda}_\nu}{A_3 - A_2 + \tilde{\lambda}_\nu} g_1(\tau_3) \tag{B.67}$$

and that the expression for τ_2 is

$$\tau_2 = T - \frac{1}{A_1}\left(\log\left[\frac{\tilde{\lambda}_\nu}{A_2 - A_1 + \tilde{\lambda}_\nu} + \frac{\tilde{\lambda}_\nu}{A_1}\right] - \log\left[\frac{\tilde{\lambda}_\nu}{A_1} + 1\right]\right). \tag{B.68}$$

Let us write out an expression for τ_3, inserting $g_2(t)$ into (B.67), which gives

$$e^{A_2(\tau_2-\tau_3)}\left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) - \left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1}g_1(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) = \frac{\tilde{\lambda}_\nu}{A_3 - A_2 + \tilde{\lambda}_\nu}g_1(\tau_3). \tag{B.69}$$

Dividing both sides by $\tilde{\lambda}_\nu$ this becomes

$$e^{A_2(\tau_2-\tau_3)}\left(\frac{1}{A_2 - A_1} + \frac{\tilde{\lambda}_\nu}{A_2(A_2 - A_1)}\right) = \left(\frac{A_3 - A_1 + \tilde{\lambda}_\nu}{(A_2 - A_1)(A_3 - A_2 + \tilde{\lambda}_\nu)}\right)g_1(\tau_3) + \frac{\tilde{\lambda}_\nu}{A_2(A_2 - A_1)}. \tag{B.70}$$

Inserting the formula for $g_1(\tau_3)$ we find

$$\begin{aligned} &e^{A_2(\tau_2-\tau_3)}\left(\frac{1}{A_2 - A_1} + \frac{\tilde{\lambda}_\nu}{A_2(A_2 - A_1)}\right) = \\ &\left(\frac{A_3 - A_1 + \tilde{\lambda}_\nu}{(A_2 - A_1)(A_3 - A_2 + \tilde{\lambda}_\nu)}\right)\left(e^{A_1(T-\tau_3)}(1 + \frac{\tilde{\lambda}_\nu}{A_1}) - \frac{\tilde{\lambda}_\nu}{A_1}\right) + \frac{\tilde{\lambda}_\nu}{A_2(A_2 - A_1)}. \end{aligned} \tag{B.71}$$

This is known as a transcendental equation (equations of the form $e^{Ax} + e^{Bx} = C$), which cannot be solved in the general sense, except for some special cases like $A = 2B$ or $A = B$. We investigate one such special case in the following subsection, but we must conclude that in the general case, τ_3 cannot be expressed explicitly.

B.3.5 Special case for explicit solution

In the $g_3(t)$ case we saw we could only find an explicit formula for τ_3 under the condition. The stopping time satisfies

$$\begin{aligned} e^{A_2(\tau_2-\tau_3)}\left((1 + \frac{\tilde{\lambda}_\nu}{A_2 - A_1})g_1(\tau_2) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) - \left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1}g_1(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) \\ = \frac{\tilde{\lambda}_\nu}{A_3 - A_2 + \tilde{\lambda}_\nu}g_1(\tau_3) \end{aligned} \tag{B.72}$$

which can be written as

$$e^{A_2(\tau_2-\tau_3)}\left(\frac{\tilde{\lambda}_\nu}{A_2 - A_1} + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)}\right) = \left(\frac{\tilde{\lambda}_\nu}{A_3 - A_2 + \tilde{\lambda}_\nu} + \frac{\tilde{\lambda}_\nu}{A_2 - A_1}\right)g_1(\tau_3) + \frac{\tilde{\lambda}_\nu^2}{A_2(A_2 - A_1)} \tag{B.73}$$

which can be simplified to

$$e^{A_2(\tau_2-\tau_3)}\left(\tilde{\lambda}_\nu + \frac{\tilde{\lambda}_\nu^2}{A_2}\right) = \left(\frac{\tilde{\lambda}_\nu(A_3 - A_2)}{A_3 - A_2 + \tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)\left(e^{A_1(T-\tau_3)}(1 + \frac{\tilde{\lambda}_\nu}{A_1}) - \frac{\tilde{\lambda}_\nu}{A_1}\right)\frac{\tilde{\lambda}_\nu^2}{A_2} \tag{B.74}$$

from which we finally see

$$\begin{aligned} e^{A_2(\tau_2-\tau_3)}\left(\tilde{\lambda}_\nu + \frac{\tilde{\lambda}_\nu^2}{A_2}\right) - e^{A_1(\tau_2-\tau_3)}e^{A_1(T-\tau_2)}\left(\frac{\tilde{\lambda}_\nu(A_3 - A_2)}{A_3 - A_2 + \tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)(1 + \frac{\tilde{\lambda}_\nu}{A_1})\frac{\tilde{\lambda}_\nu^2}{A_2} \\ = -\left(\frac{\tilde{\lambda}_\nu(A_3 - A_2)}{A_3 - A_2 + \tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)\frac{\tilde{\lambda}_\nu^3}{A_2A_1}. \end{aligned} \tag{B.75}$$

We see this leaves us with a transcendental equation, as described in section 3.1, that cannot be solved to arrive at a closed form expression for τ_3 in the general case. However, under the special condition that $\theta(2) = 2\theta(1)$ we have that $A_2 = \kappa(2\mu - \theta(2)) = 2\kappa(\mu - \theta(1)) = 2A_1$. If make the general assumption that $\theta(i) = 2\theta(i-1)$, the equation (B.75) becomes

$$e^{2A_1(\tau_2-\tau_3)}\left(\tilde{\lambda}_\nu + \frac{\tilde{\lambda}_\nu^2}{2A_1}\right) - e^{A_1(\tau_2-\tau_3)}e^{A_1(T-\tau_2)}\left(\frac{\tilde{\lambda}_\nu(2A_1)}{2A_1+\tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)(1+\frac{\tilde{\lambda}_\nu}{A_1})\frac{\tilde{\lambda}_\nu^2}{2A_1} = -\left(\frac{\tilde{\lambda}_\nu(2A_1)}{2A_1+\tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)\frac{\tilde{\lambda}_\nu^3}{2A_1^2}. \tag{B.76}$$

Substituting $x = e^{A_1(\tau_2-\tau_3)}$, leaves us with a quadratic equation of the form

$$x^2\left(\tilde{\lambda}_\nu+\frac{\tilde{\lambda}_\nu^2}{A_2}\right)-xe^{A_1(T-\tau_2)}\left(\frac{\tilde{\lambda}_\nu(A_3-A_2)}{A_3-A_2+\tilde{\lambda}_\nu}+\tilde{\lambda}_\nu\right)(1+\frac{\tilde{\lambda}_\nu}{A_1})\frac{\tilde{\lambda}_\nu^2}{A_2}+\left(\frac{\tilde{\lambda}_\nu(A_3-A_2)}{A_3-A_2+\tilde{\lambda}_\nu}+\tilde{\lambda}_\nu\right)\frac{\tilde{\lambda}_\nu^3}{A_2A_1} = 0. \tag{B.77}$$

This can be solved using the quadratic formula $x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$, where

$$\begin{aligned} a =&\tilde{\lambda}_\nu + \frac{\tilde{\lambda}_\nu^2}{A_2}, \\ b =&e^{A_1(T-\tau_2)}\left(\frac{\tilde{\lambda}_\nu(A_3-A_2)}{A_3-A_2+\tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)(1+\frac{\tilde{\lambda}_\nu}{A_1})\frac{\tilde{\lambda}_\nu^2}{A_2}, \\ c =&\left(\frac{\tilde{\lambda}_\nu(A_3-A_2)}{A_3-A_2+\tilde{\lambda}_\nu} + \tilde{\lambda}_\nu\right)\frac{\tilde{\lambda}_\nu^3}{A_2A_1}. \end{aligned} \tag{B.78}$$

Inserting this into the quadratic formula and rewriting, we end up with

$$x = \frac{(A_3-A_1+\tilde{\lambda}_\nu)}{(A_3-2A_1+\tilde{\lambda}_\nu)}\left(\frac{\tilde{\lambda}_\nu^2(2A_1+\tilde{\lambda}_\nu)}{(A_1+\tilde{\lambda}_\nu)^3}\right) \pm \sqrt{\left(\frac{\tilde{\lambda}_\nu^2}{2A_1^2} + \frac{\tilde{\lambda}_\nu}{A_1}\frac{\tilde{\lambda}_\nu}{A_3-2A_1+\tilde{\lambda}_\nu}\right)\left(\frac{2A_1^2}{(A_1+\tilde{\lambda}_\nu)^4}\right)}. \tag{B.79}$$

We then end up with a solution for x and the optimal stopping time is simply given by $\tau_3 = \tau_2 - \frac{1}{A_1}\log(x)$.

B.3.6 Solving the ODE for the $i = 4$ case

In this case we are solving the system

$$\begin{aligned} \max\left(\partial_t w(t,4) + \kappa(4\mu-\theta(4))w(t,4) + \tilde{\lambda}_\nu w(t,3)\ ;\ w(t,3) - w(t,4)\right) = 0 \\ w(T,4) = e^{-\kappa\rho_u(4)}. \end{aligned} \tag{B.80}$$

We only consider the non-trivial case where $\mu \geq \frac{-\tilde{\lambda}_\nu}{\kappa} - \theta(1) + \theta(2)$ and we use the trivial optimal strategy to execute MOs an instant before the terminal time as our terminal condition face-lifted to $w(T^-,4) = 1\cdot w(T^-,3) = 1$. We must now determine the time τ_4 when the solution to the QVI "peels away" from its immediate execution value, namely $w(t,4) = w(t,3)$. We need to ensure that $w(t,4)$ and its derivative are continuous at this point, so

$$w(\tau_4^-,4) = w(\tau_4,4) \quad \text{and} \quad \partial_t w(\tau_4^-,4) = \partial_t w(\tau_4,4) = \partial_t w(\tau_4,3). \tag{B.81}$$

From the QVI (B.80), we know that in the continuation region, we have

$$\begin{aligned} 0 =&\partial_t w(\tau_4,4) + \kappa(4\mu-\theta(4))w(\tau_4,4) + \tilde{\lambda}_\nu w(\tau_4,3) \\ =&\partial_t w(\tau_4,3) + \kappa(4\mu-\theta(4))w(\tau_4,3) + \tilde{\lambda}_\nu w(\tau_4,3) \\ =&-\kappa(3\mu-\theta(3))w(\tau 4,3) - \tilde{\lambda}_\nu w(\tau_4,2) + \kappa(4\mu-\theta(4))w(\tau_4,3) + \tilde{\lambda}_\nu w(\tau_4,3) \\ =&(\kappa(\mu+\theta(3)-\theta(4)) + \tilde{\lambda}_\nu)w(\tau_4,3) - \tilde{\lambda}_\nu w(\tau_4,2) \end{aligned} \tag{B.82}$$

where the third equality follows from the continuation region of (5.40).

We find that the optimal stopping time, at which the trader should execute a market order when she holds an inventory of size 3, solves

$$w(\tau_4, 3) = \frac{\tilde{\lambda}_\nu}{(\kappa(\mu + \theta(3) - \theta(4)) + \tilde{\lambda}_\nu)} w(\tau 4, 2). \tag{B.83}$$

This can be solved explicitly for τ_4, since $w(t, 3) = g_3(t)$ and $w(t, 2) = g_2(t)$ as given in the previous subsections. We try this in the Appendix, but find an $e^{\kappa(2\mu - \theta(2))(\tau_2 - \tau_3)}$ and an $e^{\kappa(\mu - \theta(1))(T - \tau_3)}$ term in the equation, which poses a problem. We can solve it explicitly for the special case where $\theta(2) = 2\theta(1)$, as it then becomes a quadratic equation.

Now that we know the optimal stopping time, we can find the full solution for $w(t, 3)$ and for that we must solve the continuation equation backwards from τ_3. We do this by solving the ODE

$$\partial_t g_4(t) + \kappa(4\mu - \theta(4)) g_4(t) + \tilde{\lambda}_\nu g_3(t) = 0, \quad g_4(\tau_4) = \Gamma_4 \tag{B.84}$$

where $\Gamma_4 = \frac{\tilde{\lambda}_\nu}{(\kappa(\mu + \theta(3) - \theta(4)) + \tilde{\lambda}_\nu)} w(\tau 4, 2)$. The solution is given found by

$$\begin{aligned} \partial_t g_4(t) + \kappa(4\mu - \theta(4)) g_4(t) &= -\tilde{\lambda}_\nu g_3(t) \\ \partial\left[e^{A_4 t} g_4(t)\right] &= -e^{A_4 t} \tilde{\lambda}_\nu g_3(t) \\ e^{A_4 \tau_4} g_4(\tau_4) - e^{A_4 t} g_4(t) &= -\tilde{\lambda}_\nu \int_t^{\tau_3} e^{A_4 s} g_3(t). \end{aligned} \tag{B.85}$$

For this integral, using the fact that $g_3'(t) = -A_3 g_3(t) - \tilde{\lambda}_\nu g_2(t)$, we have

$$\begin{aligned} \int e^{A_4 s} g_3(s) ds =& \frac{1}{A_4} e^{A_4 t} g_3(t) - \frac{1}{A_4} \int e^{A_4 s} g_3'(s) \\ =& \frac{1}{A_4} e^{A_4 t} g_3(t) + \frac{A_3}{A_4} \int e^{A_4 s} g_3(t) + \frac{\tilde{\lambda}_\nu}{A_4} \int e^{A_4 s} g_2(t). \end{aligned} \tag{B.86}$$

From the previous subsection we know

$$\int e^{A_4 s} g_2(s) ds = \frac{1}{A_4 - A_2} e^{A_4 t} g_2(t) + \frac{\tilde{\lambda}_\nu}{(A_4 - A_2)(A_4 - A_1)} e^{A_4 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_4(A_4 - A_2)(A_3 - A_1)} e^{A_4 t}. \tag{B.87}$$

So (B.86) becomes

$$\begin{aligned} \left(1 - \frac{A_3}{A_4}\right) \int e^{A_4 s} g_3(s) ds =& \frac{1}{A_4} e^{A_4 t} g_3(t) + \frac{\tilde{\lambda}_\nu}{A_4} \left(\frac{1}{A_4 - A_2} e^{A_4 t} g_2(t) \right. \\ & \left. + \frac{\tilde{\lambda}_\nu}{(A_4 - A_2)(A_4 - A_1)} e^{A_4 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_4(A_4 - A_2)(A_3 - A_1)} e^{A_4 t} \right) \\ \int e^{A_4 s} g_3(s) ds =& \frac{1}{A_4 - A_3} e^{A_4 t} g_3(t) + \frac{\tilde{\lambda}_\nu}{A_4 - A_3} \left(\frac{1}{A_4 - A_2} e^{A_4 t} g_2(t) \right. \\ & \left. + \frac{\tilde{\lambda}_\nu}{(A_4 - A_2)(A_4 - A_1)} e^{A_4 t} g_1(t) + \frac{\tilde{\lambda}_\nu^2}{A_4(A_4 - A_2)(A_3 - A_1)} e^{A_4 t} \right). \end{aligned} \tag{B.88}$$

Substituting this back into our solution for the ODE we get

$$\begin{aligned}-e^{A_4\tau_4}g_4(\tau_4)+e^{A_4t}g_4(t)=&\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_3}e^{A_4\tau_4}g_3(\tau_4)+\frac{\tilde\lambda_\nu}{A_4-A_3}\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_2}e^{A_4\tau_4}g_2(\tau_4)\\&+\frac{\tilde\lambda_\nu^2}{(A_4-A_2)(A_4-A_1)}e^{A_4\tau_4}g_1(\tau_4)+\frac{\tilde\lambda_\nu^3}{A_4(A_4-A_2)(A_3-A_1)}e^{A_4\tau_4}\Bigg)\Bigg)\\&-\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_3}e^{A_4t}g_3(t)+\frac{\tilde\lambda_\nu}{A_4-A_3}\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_2}e^{A_4t}g_2(t)\\&+\frac{\tilde\lambda_\nu^2}{(A_4-A_2)(A_4-A_1)}e^{A_4t}g_1(t)+\frac{\tilde\lambda_\nu^3}{A_4(A_4-A_2)(A_3-A_1)}e^{A_4t}\Bigg)\Bigg).\end{aligned}\tag{B.89}$$

We see, using $g_4(\tau_4)=g_3(\tau_4)$, our function $g_4(t)$ can be written as

$$\begin{aligned}g_4(t)=&\Bigg((1+\frac{\tilde\lambda_\nu}{A_4-A_3})e^{A_4(\tau_4-t)}g_3(\tau_4)+\frac{\tilde\lambda_\nu}{A_4-A_3}\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_2}e^{A_4(\tau_4-t)}g_2(\tau_4)\\&+\frac{\tilde\lambda_\nu^2}{(A_4-A_2)(A_4-A_1)}e^{A_4(\tau_4-t)}g_1(\tau_4)+\frac{\tilde\lambda_\nu^3}{A_4(A_4-A_2)(A_3-A_1)}e^{A_4(\tau_4-t)}\Bigg)\Bigg)\\&-\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_3}g_3(t)+\frac{\tilde\lambda_\nu}{A_4-A_3}\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_2}g_2(t)\\&+\frac{\tilde\lambda_\nu^2}{(A_4-A_2)(A_4-A_1)}g_1(t)+\frac{\tilde\lambda_\nu^3}{A_4(A_4-A_2)(A_3-A_1)}\Bigg)\Bigg).\end{aligned}\tag{B.90}$$

We end up with the following form

$$\begin{aligned}g_4(t)=&e^{A_4(\tau_4-t)}\Bigg((1+\frac{\tilde\lambda_\nu}{A_4-A_3})g_3(\tau_4)+\frac{\tilde\lambda_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(\tau_4)\\&+\frac{\tilde\lambda_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(\tau_4)+\frac{\tilde\lambda_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\Bigg)\\&-\Bigg(\frac{\tilde\lambda_\nu}{A_4-A_3}g_3(t)+\frac{\tilde\lambda_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(t)\\&+\frac{\tilde\lambda_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(t)+\frac{\tilde\lambda_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\Bigg)\Bigg)\end{aligned}\tag{B.91}$$

where $g_3(t)$ is expressed in equation (5.45) and τ_4 solves equation (B.83). So the solution of the QVI (B.80) for $w(t,4)$ is

$$w(t,4)=g_4(t)1_{\{t<\tau_4\}}+g_3(t)1_{\{t\geq\tau_4\}}.\tag{B.92}$$

Finally the optimal depth at which the trader should post her LOs is given by

$$\delta^\star(t,4)=\frac{1}{\kappa}+r+f+\frac{1}{\kappa}\log\left(\frac{w(t,4)}{w(t,3)}\right).\tag{B.93}$$

B.3.7 Solving the ODE for the $i = S$ case

For our value formulas $g_1(t), g_2(t)$ and $g_3(t)$ we have found

$$
\begin{aligned}
g_1(t) =& e^{A_1(T-t)}\left(1+\frac{\tilde{\lambda}_\nu}{A_1}\right)-\frac{\tilde{\lambda}_\nu}{A_1},\\
g_2(t) =& e^{A_2(\tau_2-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_2-A_1})g_1(\tau_2)+\frac{\tilde{\lambda}_\nu^2}{A_2(A_2-A_1)}\right)-\left(\frac{\tilde{\lambda}_\nu}{A_2-A_1}g_1(t)+\frac{\tilde{\lambda}_\nu^2}{A_2(A_2-A_1)}\right),\\
g_3(t) =& e^{A_3(\tau_3-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_3-A_2})g_2(\tau_3)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(\tau_3)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right)\\
&-\left(\frac{\tilde{\lambda}_\nu}{A_3-A_2}g_2(t)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(t)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right),\\
g_4(t) =& e^{A_4(\tau_4-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_4-A_3})g_3(\tau_4)+\frac{\tilde{\lambda}_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(\tau_4)\right.\\
&\left.+\frac{\tilde{\lambda}_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(\tau_4)+\frac{\tilde{\lambda}_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\right)\\
&-\left(\frac{\tilde{\lambda}_\nu}{A_4-A_3}g_3(t)+\frac{\tilde{\lambda}_\nu^2}{(A_4-A_3)(A_4-A_2)}g_2(t)\right.\\
&\left.+\frac{\tilde{\lambda}_\nu^3}{(A_4-A_3)(A_4-A_2)(A_4-A_1)}g_1(t)+\frac{\tilde{\lambda}_\nu^4}{A_4(A_4-A_3)(A_4-A_2)(A_3-A_1)}\right)\Bigg)
\end{aligned}
\tag{B.94}
$$

with $A_1 = \kappa(\mu - \theta(1))$, $A_2 = \kappa(2\mu - \theta(2))$ and $A_3 = \kappa(3\mu - \theta(3))$ and where we have the stopping times τ satisfying

$$
\begin{aligned}
g_1(\tau_2) =& \frac{\tilde{\lambda}_\nu}{A_2-A_1+\tilde{\lambda}_\nu},\\
g_2(\tau_3) =& \frac{\tilde{\lambda}_\nu}{A_3-A_2+\tilde{\lambda}_\nu}g_1(\tau_3),\\
g_3(\tau_4) =& \frac{\tilde{\lambda}_\nu}{A_4-A_3+\tilde{\lambda}_\nu}g_2(\tau_4).
\end{aligned}
\tag{B.95}
$$

We see the general formula can be expressed as

$$
g_S(t) = e^{A_s(\tau_S-t)}\left(g_{S-1}(\tau_S)+\sum_{i=1}^{S}\frac{\tilde{\lambda}_\nu^i}{\prod_{j=1}^{i}(A_S-A_{S-j})}g_{S-i}(\tau_S)\right)-\left(\sum_{i=1}^{S}\frac{\tilde{\lambda}_\nu^i}{\prod_{j=1}^{i}(A_S-A_{S-j})}g_{S-i}(t)\right)
\tag{B.96}
$$

with $A_i = \kappa(i\mu - \theta(i))$ for $i = 1, 2, 3...$, $A_0 = 0$ and $g_0(t) = 1$ and where the stopping time τ_S solves the equation

$$
g_{S-1}(\tau_S) = \frac{\tilde{\lambda}_\nu}{A_S - A_{S-1} + \tilde{\lambda}_\nu}g_{S-2}(\tau_S).
\tag{B.97}
$$

To be able to effectively apply Newton-Raphson's method we need the derivative of $g_S(t)$, which is

$$\begin{aligned} g_1'(t) =& -A_1 e^{A_1(T-t)}\left(1+\frac{\tilde{\lambda}_\nu}{A_1}\right),\\ g_2'(t) =& -A_2 e^{A_2(\tau_2-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_2-A_1})g_1(\tau_2)+\frac{\tilde{\lambda}_\nu^2}{A_2(A_2-A_1)}\right)-\frac{\tilde{\lambda}_\nu}{A_2-A_1}g_1'(t),\\ g_3'(t) =& -A_3 e^{A_3(\tau_3-t)}\left((1+\frac{\tilde{\lambda}_\nu}{A_3-A_2})g_2(\tau_3)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(\tau_3)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right)\\ &-\left(\frac{\tilde{\lambda}_\nu}{A_3-A_2}g_2'(t)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1'(t)\right). \end{aligned} \tag{B.98}$$

We evaluate $g_3'(\tau_3)$, and find it is given by

$$\begin{aligned} g_3'(\tau_3) =& -A_3\left((1+\frac{\tilde{\lambda}_\nu}{A_3-A_2})g_2(\tau_3)+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}g_1(\tau_3)+\frac{\tilde{\lambda}_\nu^3}{A_3(A_3-A_2)(A_3-A_1)}\right)\\ &-\left(\frac{\tilde{\lambda}_\nu}{A_3-A_2}(-A_2g_2(\tau_3)-\tilde{\lambda}_\nu g_1(\tau_3))+\frac{\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}(-A_1g_1(\tau_3)-\tilde{\lambda}_\nu)\right)\\ =&\Bigg((-A_3+\frac{-A_3\tilde{\lambda}_\nu}{A_3-A_2}+\frac{A_2\tilde{\lambda}_\nu}{A_3-A_2})g_2(\tau_3)\\ &+(\frac{-A_3\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)}+\frac{\tilde{\lambda}_\nu^2}{A_3-A_2}+\frac{A_1\tilde{\lambda}_\nu^2}{(A_3-A_2)(A_3-A_1)})g_1(\tau_3)\\ &+\frac{-\tilde{\lambda}_\nu^3}{(A_3-A_2)(A_3-A_1)}+\frac{\tilde{\lambda}_\nu^3}{(A_3-A_2)(A_3-A_1)}\Bigg). \end{aligned} \tag{B.99}$$

Using that $g_2'(\tau_3) = -A_2g_2(\tau_3) - \tilde{\lambda}_\nu g_1(\tau_3)$ and $\tilde{\lambda}_\nu g_1(\tau_3) = (A_3 - A_2\tilde{\lambda}_\nu)g_2(\tau_3)$ to find $g_2'(\tau_3) = -(A_3 + \tilde{\lambda}_\nu)g_2(\tau_3)$, we can rewrite (B.99) as

$$g_3'(\tau_3) =(-A_3-\tilde{\lambda}_\nu)g_2(\tau_3)+(\frac{-\tilde{\lambda}_\nu^2}{(A_3-A_2)}+\frac{\tilde{\lambda}_\nu^2}{A_3-A_2})g_1(\tau_3)=g_2'(\tau_3) \tag{B.100}$$

C | Python Codes

The code used to create all the plots and analyses in this thesis can be found in this GitHub repository.