DYNAMIC PROGRAMMING AND TRADE EXECUTION

TIANHUI MICHAEL LI

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

OPERATIONS RESEARCH FINANCIAL ENGINEERING

ADVISER: RENE CARMONA

JUNE, 2013
Abstract

We apply dynamic programming to two different trading problems. We introduce a novel trading model that captures the active-versus-passive order tradeoff faced by a broker when benchmarked to VWAP (Volume Weighted Average Price). We are able to solve both the case where the stock quantity is discrete and continuous. The solution is in terms of a highly intuitive forward and backward boundary, for which we even obtain closed-form solutions in certain cases. The second problem is the dynamic hedging of an option under an Almgren-Chriss model of market impact. We are able to derive a highly intuitive dynamic solution.
To my mother, for her love
and to my grandfather, for all that he taught me.
Acknowledgements

The author would like to thank the Fannie and John Hertz Foundation and the National Science Foundation for their generous support of his graduate studies.

The author would also like to acknowledge his adviser, Rene Carmona, and all those who have supported his intellectual endeavors, including Robert Almgren, Peter Carr, Patrick Cheridito, and Chris Rogers.
# Contents

Abstract iii

Dedication iv

Acknowledgements v

Contents vi

List of Figures ix

List of Tables x

1 Introduction 1
   1.1 Background .................................................. 1
   1.2 Outline .................................................. 1
   1.3 Review of the Literature .................................. 2
   1.4 VWAP Benchmarking ......................................... 3
   1.5 Options Hedging ........................................... 7
   1.6 Outline of Thesis .......................................... 11

2 VWAP Tracking – Discrete Case 12
   2.1 Setup .................................................. 12
   2.2 General Solution .......................................... 20
   2.3 Special Cases and Plots ................................... 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.5 Intraday Trading Patterns</td>
<td>104</td>
</tr>
<tr>
<td>4.4 Discrete-Time Formulation and Solution</td>
<td>106</td>
</tr>
<tr>
<td>4.4.1 Discrete-Time Formulation</td>
<td>107</td>
</tr>
<tr>
<td>4.4.2 Problem Solution</td>
<td>108</td>
</tr>
<tr>
<td>4.5 Simulation Using TAQ Data</td>
<td>108</td>
</tr>
<tr>
<td>A Martingale Principle of Optimal Control</td>
<td>114</td>
</tr>
<tr>
<td>References</td>
<td>116</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Strategies for proportional transaction cost versus market impact . . . . . . . . . . . . 10

2.1 Plot of single-fill case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
2.2 Plot of single-fill case with U-shaped volume . . . . . . . . . . . . . . . . . . . . . . . . . . 41
2.3 Plot of large-fill case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
2.4 Plot of large-fill case with U-shaped volume . . . . . . . . . . . . . . . . . . . . . . . . . . 43

3.1 Diagram of three trading regions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
3.2 Plot of $V$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.3 Plot of $\tau_m$ versus $\tau_l$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
3.4 Diagram of three trading regions and region of convergence . . . . . . . . . . . . . . . . . 68

4.1 Comparative statics of $(KA_2(T - t) - \nu)/\lambda$ . . . . . . . . . . . . . . . . . . . . . . . . 91
4.2 Comparative statics of $A_1(T - t)$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
4.3 Intraperiod Mishedging and Trading Intensity . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100
4.4 Plots of unrestricted and restricted problems . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102
4.5 Intraday trading intensity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
List of Tables

4.1 Units for constants ................................................. 91
4.2 Summary Statistics of Stocks .................................... 109
4.3 Summary Statistics of Simulation ................................. 113
Chapter 1

Introduction

1.1 Background

A defining hallmark of modern financial markets has been the proliferation of algorithmic trading. According to the Financial Times, 73% of all trading in the US is thought to be through algorithmic trading.¹ Many investors lack the specific expertise or resources to implement an algorithmic trading platform and delegate the task of order execution to algorithmic brokers who execute on their behalf. In this paper, we investigate novel problems which arise as a result of electronic trading and which relate to market microstructure and transaction costs.

1.2 Outline

The rest of the chapter is organized as follows. In Section 1.3, we give a review of the existing transaction cost literature and different models that attempt to quantify the ‘frictions’ of trading. A novel trading model that features both temporal and price frictions is introduced in Section 1.4. This model is applied to benchmarking to VWAP (Volume Weighted Average Price) and highlighting its differences with both existing industry practice and the academic literature. In Section 1.5, we introduce options hedging under market impact.

¹SEC runs eye over high-speed trading, Financial Times, July 29, 2009.
and argue that the Almgren Chriss model accurately captures effective market impact. We also compare our model and results with the existing literature on delta-hedging under transaction costs and portfolio selection using Almgren Chriss.

1.3 Review of the Literature

Trading frictions have been modeled via various mechanisms. One strand of the literature is centered around illiquidity [Cetin, Jarrow, and Protter, 2004]. These authors hypothesize a stochastic supply curve whose steepness coincides with illiquidity. Similar work was done by Bank and Baum [2004]. The major drawback of both their theories is that liquidity costs can be completely avoided if trades are finite-variation. Hence, the cost of replicating an option is asymptotically zero as the writer can take finite-variation approximations to the classical Black-Scholes Δ-hedging technique. To get around this problem, Cetin, Jarrow, Protter, and Warachka [2006] impose an artificial constraint on the minimum time between discrete trades.

Another strand of the literature models trading frictions as a cost proportional to trade size, typically interpreted as arising from the bid-ask spread. Mathematically, these problems exhibit a scaling which allows the authors to simplify their formulas. This branch of the literature uses singular control and the optimal solution is typically in the form of a tracking band. As the portfolio exits this band, the trader makes singular corrections to his holdings to keep it strictly within the limits of the band [Cvitanic, 1999, Cvitanic and Karatzas, 1996]. This is analogous to the results of Davis and Norman [1990], Shreve and Soner [1994] for the Merton Problem. Typically, there is no analytic form for these no-trade regions and much work has been done on their numerics [Clewlow and Hodges, 1997, Davis, Panas, and Zariphopoulou, 1993] or asymptotics [Soner and Barles, 1998, Whalley and Wilmott, 1997].

resilient order book where levels gradually replenish after large trades. They generally derive optimal strategies in which the broker makes large block trades at the beginning and end of the trading period with many small trades in between. These models often allow for general order book shape and are mathematically rather challenging to work with.

The final strand of related literature focuses on modeling effective market impact. Rather than dwelling on the details of microstructure, it models ‘effective’ microstructure. The literature began with Bertsimas and Lo [1998] and Almgren and Chriss [1999] who decompose price impact in terms of both temporary and permanent price impact. The model was subsequently refined in Obizhaeva and Wang [2005] by the addition of a transient price impact term. While earlier works in the literature concentrated on the liquidation of a fixed number of shares in a finite period, these impact models were applied to portfolio allocation [Garleanu and Pedersen, 2009] and options hedging [Rogers and Singh, 2007], the liquidation of portfolios [Schied and Schöneborn, 2007], trading in both lit and dark venues [Kratz and Schöneborn, 2010], and multiple-agent competitive liquidation [Schoeneborn and Schied, 2010].

1.4 VWAP Benchmarking

The delegation of order execution to brokers leads to an agency conflict between brokers and investors. To compare broker performance, investors typically benchmark brokers to the Volume Weighted Average Price (VWAP). That is, the broker is asked to execute shares over a particular interval and the order’s average execution price will be compared to the market VWAP during the same interval. Brokers are typically penalized for both underperforming with respect to VWAP and for not tracking VWAP closely. This paper studies how electronic brokerages should place orders when benchmarked to VWAP.

The broker faces a risk-reward trade off in deciding whether to execute with market (active) or limit (passive) orders. Market orders provide execution certainty but require paying half a bid-ask spread (relative mid-price) while limit orders save half a bid-ask
spread at the expense of execution uncertainty. The broker can perfectly replicate interval VWAP by choosing an execution schedule of market orders to match the interval volume profile. However, he would consistently underperform VWAP by half a bid-ask spread. Alternatively, the broker can choose to consistently save half the bid-ask spread by submitting only limit orders. But then the realized execution profile may differ wildly from the market volume profile, thus incurring VWAP mistracking error. This execution timing versus transaction cost decision is the tradeoff at the heart of our limit order versus market order slicing problem and is the novel feature of our paper.

In this paper, we present a solution that optimally balances between these two conflicting objectives. We find that the optimal solution consists of forward and backward boundaries ($\tilde{L}_t$ and $\tilde{M}_t$) that track the ideal VWAP trajectory. The boundaries are time-dependent, satisfy $\tilde{M}_t \geq \tilde{L}_t$, and bound $v_t$, the total remaining shares to execute

$$\tilde{L}_t \leq v_t \leq \tilde{M}_t \quad (1.1)$$

If the number of shares remaining to execute reaches the backward boundary ($v_t \geq \tilde{M}_t$) the broker executes market orders to stay just within the backward boundary (i.e. so that (1.1) is satisfied). At any given point, the broker places $v_t - \tilde{L}_t$ shares on the order book. If he is crossed, at most $v_t - \tilde{L}_t$ limit orders are executed so that he has $v_t \geq \tilde{L}_t$ shares remaining to be executed after the limit orders are lifted. These forward and backward boundaries provide an extremely tractable solution that can be implemented in any order book placement algorithm. The rest of the paper will both prove why such a strategy is optimal and show how to compute the optimal boundaries $\tilde{L}_t$ and $\tilde{M}_t$.

Our result is reminiscent to those of the proportional-transaction-cost literature [Davis and Norman, 1990, Shreve and Soner, 1994] in that we are able to generate a forward and backward boundary for limit and market order execution. However, a key difference is that we allow for both passive and active orders while they assume all orders are active. Hence, while our backward boundary for market order execution is similar to the boundaries in the existing literature, the forward boundary for limit orders is qualitatively different.
Another strand of the literature deals with the optimal depth of limit order placement. This line of research was pioneered by Avellaneda and Stoikov [2008], based closely on the work of Ho and Stoll [1981]. They model a market maker who is able to control posting depth on both the bid and the ask. Subsequent work for the market maker problem has been done by Guéant, Lehalle, and Tapia [2011b] and Cartea, Jaimungal, and Ricci [2011] and a similar model with one-sided posting has been investigated for the liquidation problem [Guéant, Lehalle, and Tapia, 2011a]. These models represent a more tractable take on an order book, representing some features (such as multiple posting levels) while ignoring others (queue priority) and present a more tractable alternative to full order book models [Alfonsi and Schied, 2009, Predoiu, Shaikhet, and Shreve, 2010].

In this strand of the literature, the agent is allowed to post one unit in the book and fills are of unit size as well. It is therefore capable of answering questions about posting depth, but not posting quantity. In contrast, our model only allows posting an arbitrary quantity to the top of the book and allows for both complete and partial fills. We seek to answer questions concerning posting quantity rather than posting depth. Another drawback of Avellaneda and Stoikov [2008] and its descendants is that they do not allow market orders (at least not during the execution period). The latter point is crucial: while in Guéant, Lehalle, and Tapia [2011a], the broker can only use limit orders to liquidate and thus can fall arbitrarily behind his liquidation schedule, in our model he can execute market orders to keep up with the schedule (at a cost).

In the course of this work, we assume that the volume profile (which can be time-varying in our model) is known in advance. While this assumption is not always realistic, it is justified for two reasons. The first is that our results apply directly if the broker were benchmarked to TWAP (Time Weighted Average Price) rather than VWAP. The second is that the VWAP slicing literature has (up to now) been concentrated on stochastic volume at the expense of modeling market impact. Konishi [2002] solves the time-slicing problem with stochastic volatility and stochastic volatility and the result is extended in [McCulloch and Kazakov, 2007]. They show that when trading volume and volatility are positively
correlated (as has been widely documented in real markets), it is optimal to trade behind
the VWAP schedule. That is, the new ‘adjusted’ VWAP schedule is not the expected daily
VWAP, but expected daily VWAP minus a term that accounts for the correlation with the
volume and volatility. Hence, while we consider the VWAP in our model, we could just as
well use the ‘adjusted’ VWAP to account for stochastic volume and volatility. However,
a major drawback of the analysis in Konishi [2002], McCulloch and Kazakov [2007] is its
restriction to static strategies in contrast to the dynamic ones considered in our paper.
This literature also ignores the execution timing versus transaction cost tradeoff that is
the focus of our paper.

Our results bear a close mathematical analog to work in the reinsurance literature
[Goreac, 2008, Mnif and Sulem, 2005] and its Brownian limit is investigated in [Asmussen,
Hjgaard, and Taksar, 1998]. However, their financial interpretation and even some of the
mathematical details are crucially different. We postpone a detailed discussion until after
we have introduced our model (see Remark 2.1.8). At this point, we only note that while
they are only able to prove existence and uniqueness of a viscosity solution, we are able
to prove existence and uniqueness of a classical solution, which in many cases is even
closed-form.

Finally, we compare how our proposed methodology differs from current practice in
industry. Electronic brokerage houses and in-house electronic execution platforms typically
split execution into two rigid steps. A high-level objective-aware algo splices a large trade
into time intervals, scheduling the number of shares that need to be executed in each interval
while a low-level microstructure-aware algo makes trading decisions. For example, the high-
level algo might splice a large full-day VWAP trade into half-hour buckets while the lower-
level algo actually places the limit and market orders on the book. The communication is
exclusively top-down. If the low-level algo encounters adverse market conditions, it has no
way of notifying the top-level algo of the microstructure to modify the trading schedule.
Instead, the low-level algo is forced to cross the spread and execute an expensive “cleanup”
market order to fill his allotted shares before time runs out. On the flip side, if the low-level
algo encounters favorable market conditions, it has no way of requesting a greater allocation to take advantage of the situation. We combine the two algorithms into a single strategy which is able to take advantage of prevailing microstructure conditions while keeping track of the overall objective.

1.5 Options Hedging

The solution to hedging an option in a complete market is well known [Black and Scholes, 1973, Merton, 1973]. However, the assumption of frictionless trading that underlies complete markets is far from reality. Practitioners who trade large positions are familiar with the concept of market impact when trading in real markets with limited liquidity. For instance, a buyer who lifts orders from the offer eats up liquidity and pushes his execution price ever higher. Liquidity of the underlying often dries up during financial crises such as the 1987 Crash, the LTCM’s collapse, the recent subprime debacle, or the London Whale scandal and this has only further heightened investor concerns about liquidity.

We are interested in the optimal hedging of an option by a large risk-averse investor in an illiquid market. There is a large literature on trading under transaction costs. Part of the literature is interested in super-replication [Cetin, Soner, and Touzi, 2010, Soner, Shreve, and Cvitanic, 1995]. Our paper relaxes this requirement by having a finite penalty for being mishedged. Our paper is more closely linked to these using a utility-based framework [Cvitanic and Wang, 2001, Davis and Norman, 1990, Hodges and Neuberger, 1989, Janecek and Shreve, 2004, Leland, 1985, Shreve and Soner, 1994].

In contrast to proportional-transaction-cost models, the source of incompleteness for our model is price impact: the more aggressively one trades, the more impact is incurred proportionally. The cost is no longer linear in the shares purchased and the problem no longer exhibits linear scaling. We readily conceded that an ideal transaction cost model would naturally take into account both bid-ask spread and market impact. However, such a model presents a different set of (very intriguing) mathematical challenges that are beyond
the scope of this paper. Nonetheless, we argue that neglecting an explicit spread term is a more realistic model of the effective market impact than including only the spread. The following cases serve to illustrate this point.

- The spreads-only cost model assumes all orders are active (at the offer for purchases, at the bid for sales). Most executions in US equities occur through electronic brokers who execute both passively and aggressively. If the passive/active fraction is 50%, then the average execution is at the mid price, i.e., the orders incur zero market impact (this assumes that all passive and active orders execute at the top of the book). Most brokers are able to attain near-mid execution for non-urgent orders.

- For more urgent orders, the fraction of aggressive fills typically increases. For example, fills that are 80% aggressive and 20% passive result in an effective market impact of 30% of the bid-ask spread (again, assuming top-of-the-book execution). We can observe that the effective market-impact is actually a smooth function of urgency and not realistically modeled by a fixed, discrete transactions cost as in the spread-only model.

- For super-urgent orders, the top-of-the-book fills assumption is no longer realistic. A spread-only model limits the market impact to the half-spread and is completely unable to account for execution deeper into the book.

In order to realistically account for the last point in a microstructure sensitive way, we would need to directly model the order book and both active and passive orders. This is addressed in another strand of the literature (Alfonsi and Schied [2009], Predoiu et al. [2010]) by explicitly modeling the order book. However, doing so poses a different set of mathematical challenges that are beyond the scope of this article. Our “effective” model is chosen to judiciously capture the effective trading costs while minimizing mathematical fuss. Yet, we are still able to place this model within a tractable framework to obtain a highly intuitive closed-form trading solution.
Despite this mathematical parsimony, our solution more realistically handles the case when the portfolio is far from being hedged, for example due to a sudden movement in underlying price. Under only proportional transaction costs, the optimal solution in this situation is to cross the spread with market orders in order to re-enter the tracking band. In real markets, such an order would likely eat several levels into the order-book, incurring significant market-impact, not modeled by the naive, proportional-transaction-costs model. Under our market-impact model, the optimal solution in this situation is to trade aggressively towards being hedged, in a way that takes into account both available liquidity and the degree of the mishedge. Our trading strategy is much smoother—in fact, we are approximating the impact-free Black-Scholes Delta, an infinite variation process, with trading positions that are differentiable. A comparison is given in Figure 1.1 to illustrate the point.

We separate market impact into temporary and permanent impact terms following the framework developed by Almgren and Chriss [2001]. The temporary impact affects only the execution price $\tilde{P}_t$ but has no effect on the “fair value” or “fundamental price” $P_t$. In contrast, the permanent impact directly affects $P_t$ while having no direct effect on the execution price $\tilde{P}_t$. Thus we can think of the temporary impact as connected to the liquidity cost faced by the agent while the permanent impact as linked to information transmitted to the market by the agent’s trades (see, for example, Back [1992], Kyle [1985]).

Previous work involving the Almgren-Chriss model has involved trading in the presence of Dark Pools [Kratz and Schöneborn, 2010], portfolio liquidation [Schied and Schöneborn, 2007], and competitive liquidation [Schoeneborn and Schied, 2010]. Hernandez-Del-Valle and Pacheco-Gonzalez [2009] extends the work to Geometric Brownian Motion and Horst and Naujokat [2008] extends it to derivative valuation. Rogers and Singh [2007] also examine the delta hedging problem but for agents with different risk preferences and were only able to obtain analytic expressions asymptotically. Our work differs from theirs in that we derive closed-form expressions by using simpler risk preferences. The paper most closely related to ours is Garleanu and Pedersen [2009]. They solve the infinite-horizon ‘Merton
Problem under only temporary market-impact assumptions. As in our setup, they use a linear-quadratic objective rather than the traditional expected utility setup of the classical Merton Problem. They find that trading intensity at time $t$ (Proposition 5) is given by

$$
\theta_t = -\kappa h \cdot (X_t - \text{target}_t) \quad \kappa \propto 1/\sqrt{\lambda}
$$

where $X_t$ is the number of shares, $\kappa$ is an urgency parameter with units of inverse time, $h > 0$ is a dimensionless constant of proportionality related to the convexity of the continuation value, and the “target portfolio” $\text{target}_t$ is related to the frictionless Merton-optimal portfolio. That is, with market-impact costs, it is no longer optimal to hold the instantaneous Merton-optimal portfolio but instead, the agent trades towards a new target portfolio (which is a new Merton-optimal portfolio). The intensity of trading $\theta_t$ is proportional to
the distance between the current holdings and target \((X_t - \text{target}_t)\) and is inversely proportional to the square-root of the illiquidity parameter \(\lambda\).

We use an analogous finite-horizon setup on \([0, T]\) with both temporary and permanent impact and obtain that the trading intensity is

\[
\theta_t = -\kappa h(\kappa(T - t)) \cdot (X_t - \text{target}_t)
\]

where the “target portfolio” \(\text{target}_t\) is now related to the frictionless Black-Scholes delta-hedge and \(h(\cdot)\) is a positive dimensionless function, which comes from the finite-horizon nature of the setup (compare to (4.15) with \(\nu = 0\) so \(K = 1\)). Hence, as in Garleanu and Pedersen [2009], an agent facing market illiquidity no longer maintains the zero-liquidity-cost optimal portfolio but instead trades towards a new target to correct this ‘mishedging’. Furthermore, also as in Garleanu and Pedersen [2009], the agent’s trading intensity \(\theta_t\) is proportional to the difference between his current holdings and the optimal no market-impact portfolio \((X_t - \text{target}_t)\) and inversely related to the market-impact cost \(\lambda\). If the option as well-approximated by having a constant gamma, we find that \(\text{target}_t\) is actually the Black-Scholes hedge (Theorem 4.2.2). This similarity to Garleanu and Pedersen [2009] suggests that we can think of delta-hedging in an illiquid market as a Merton optimal investment problem where the Merton portfolio is the Black-Scholes hedge portfolio. For more general options, the target \(\text{target}_t\) has an extra term accounting for the non-zero third derivative with respect to spot of the Black-Scholes option price (Theorem 4.2.4).

### 1.6 Outline of Thesis

The rest of the dissertation is as follows. In Chapter 2, we introduce and solve the VWAP tracking problem for discrete shares. The continuous shares case is handled in Chapter 3. Finally, we examine the Options Hedging problem in Chapter 4.
Chapter 2

VWAP Tracking – Discrete Case

The subsequent sections are as follows. In Section 2.1, we introduce the discrete setup and model for limit orders and market orders. The solution is presented in Section 2.2. Special cases are discussed in Section 2.3 and plots of the optimal trading policy and continuation value are also provided.

2.1 Setup

We follow the notation of Konishi [2002]. Let $v \in \mathbb{N}$ denote the volume of shares that the broker must sell between time 0 and time $T$. We assume, the order size $v$ is an insignificant fraction of interval-volume so that the market interval volume (total volume traded before $T$) can be written $V$ rather than $V + v$. While it is a stochastic quantity in real life, we will assume $V$ is deterministic for modeling simplicity. Similarly, it is assumed that the broker has no impact on the benchmark interval VWAP. Let the (mid) price be given exogenously by a Gaussian process

$$P_t = P_0 + \int_0^t \sigma(T - u) \, dW_u$$

where $\sigma : [0, T] \to (0, \infty)$ is assumed to be a deterministic function and $W_t$ is a standard Brownian motion. (It turns out it will be more convenient to specify time in terms of time-to-maturity $\tau = T - t$.) We assume the bid and ask (prices) are a fixed $s/2$ below
and above the mid (price). The choice of arithmetic Brownian Motion to model the mid is appropriate for the short time scales under consideration. For fixed $V > 0$, let the traded market volume up to time $t$ (denoted $V(t)$) be a deterministic, continuous strictly increasing function such that $V(0) = 0$ and $V(T) = V$. Then the interval market VWAP is therefore

$$\text{VWAP} = \frac{1}{V} \int_0^T P_t dV(t).$$

The number of shares executed by the broker is the sum of the limit and market orders executed,

$$v_t = l_t + m_t \quad \text{and} \quad v_T = v,$$

where $l_t$ and $m_t$ are non-decreasing and $\mathbb{N}$-valued. The agent’s average liquidation price is therefore

$$\text{vwap} = \frac{1}{v} \int_0^T \left( P_t - \frac{s}{2} \right) dm_t + \left( P_t + \frac{s}{2} \right) dl_t.$$

While the agent has complete control over market orders $m_t$ at the bid, his limit orders $l_t$ at the ask are only lifted when crossing market orders arrive.

We assume crossing market orders are modeled as an integer-valued Poisson arrival process. The $\mathbb{N}$-valued increments of this process are denoted $\Delta Z_t \geq 0$ and are independent of $W$ (and hence $P_t$). We model this as a Poisson process where the event $\Delta Z_t = k$ occurs with time-varying (but deterministic) intensity

$$\nu(T - t, k)$$

and we write $\nu(T - t) = \sum_{j \geq 1} \nu(T - t, j)$ for short. We assume the density $\nu(T - t, j)$ has finite total mass on $[0, T] \times \mathbb{N}$. Let $\mathcal{F}_t$ be the filtration generated by the arrivals $\Delta Z_t$ and $W_t$. At time $t$ the agent can control $L_t$, the number of limit orders on the book. An arrival process $\Delta Z_t$ at time $t$ executes $\Delta Z_t \wedge L_t$ shares at the ask. Thus, the filled limit order process is given by

$$l_t = \sum_{s:s \leq t} \Delta Z_s \wedge L_s.$$
Finally, we allow $\Delta Z_s$ to take on the value $\infty$, i.e. the order is big enough to fill all limit orders up to $L_t$ for any $L_t$.

We have abstracted out the details of the order book to form a parsimonious model that highlights the roles of market and limit orders. For example, we ignore the possibility of placing a limit order deeper in the book and concentrate on the best bid and ask (which fluctuate with $P_t$). We have also (implicitly) assumed that the broker is the only one trading at the ask. This can be thought of as a simplified version of a full order book with competing orders where $\Delta Z$ represents only those crossing orders which execute against broker shares. Even to fully represent the dynamic details of time priority of the ask order queue, we must keep track of all competing orders and the priority of the broker’s orders with respect to those. This leads to a high-dimensional state space that is intractable both analytically and even difficult to model numerically.

One may object that by ignoring time priority, we have abstracted out an important feature of the order book. Our abstraction would be of concern only if our resulting optimal strategies were highly unrealistic, for example if they involved constantly placing and removing orders from the queue, which would be an unrealistic strategy under the presence of time priority. It turns out that the resulting optimal strategy will be to place more limit orders on the queue as we fall behind schedule and limit orders are only removed when they are crossed (see Theorem 2.2.1).

One could also object that fixed deterministic spread and a bid and ask that fluctuate with $P_t$ is an idealization that ignores too much microstructure detail: after all, the queues do not change prices levels, rather new queues form at the new resulting bid and ask. We argue that this factor is not salient. For example, if the price moves down away from the ask, the selling broker would simply move his limit orders from the old ask to the new ask. The broker may lose time priority, but that is a detail we have decided to abstract out of our model. The more interesting direction is if the price moves towards him, then crossing orders execute against his passive orders. By assuming independence of the price process $P_t$ and limit order arrivals $\Delta Z$, we have implicitly ignored the adverse selection between
price movements and passive orders being filled. However, our model is able to cover the adverse selection case without explicit modeling (see Remark 2.1.5).

Let \( X(T - t) \) and \( x_t \) represent the fractions of shares left to be executed in the market and by the broker, respectively,

\[
X(T - t) = \frac{V - V(t)}{V} \quad x_t = \frac{v - v_t}{v}
\]

so that \( X(T) = x_0 = 1 \) and \( X(0) = x_T = 0 \). We sometimes write \( x^{L,m} \) for the trading process \( x \) that results from applying the trading strategy \( (L, m) \). At this point, we observe that we can rescale time so that \( X(T - t) \) is the linear function

\[
X(T - t) = \frac{T - t}{T}.
\]

We make this assumption for the rest of the paper (unless otherwise noted) and recognize that \( \sigma(T - t) \) and \( \nu(T - t, k) \) are rescaled to new quantities \( \tilde{\sigma}(T - t) \) and \( \tilde{\nu}(T - t, k) \), although we shall omit the tilde ornamentation for the rest of the paper. We make the following two assumptions:

**Assumption 2.1.1.** We assume that \( \nu \) and \( \sigma^2 \) are bounded in the sense

\[
\sup_{\tau \in [0, T]} \nu(\tau) < \infty \quad \text{and} \quad \int_0^T \sigma^2(\tau) \, d\tau < \infty.
\]

**Assumption 2.1.2.** We assume that for all \( x \in [0, 1] \), the function

\[
\tau \mapsto \frac{\sigma^2(\tau)}{\nu(\tau)} (x - X(\tau))
\]

is (strictly) decreasing (here, we write \( \tau = T - t \)). The probability-measure-valued function

\[
\tau \mapsto \frac{\nu(\tau, \cdot)}{\nu(\tau)}
\]

is non-decreasing in the sense of first-order stochastic dominance. That is, for \( \tau \leq \tau' \), we have

\[
\sum_{j=1}^{\infty} \frac{\nu(\tau, j)}{\nu(\tau)} f(j) \leq \sum_{j=1}^{\infty} \frac{\nu(\tau', j)}{\nu(\tau')} f(j)
\]

for all bounded, non-decreasing functions \( f : \mathbb{N} \to \mathbb{R} \).
Remark 2.1.3. Assumption 2.1.1 is necessary for the problem to be well-defined and this technical assumption holds true for real applications. Assumption 2.1.2 is sufficient to ensure regularity of the solution. It is not a necessary condition, but is the ‘cleanest’ sufficient condition. The HJB equations, while describing the optimal control problem do not lend themselves to numerical techniques. We will show that under this assumption, a system of ODEs for \( J_n \) which are readily numerically computable can be shown to solve the HJB equation. See Remark 2.2.6 for details.

Remark 2.1.4. The second condition of Assumption 2.1.2 is straightforward to verify. Observe that \( \frac{\nu(T-t, \cdot)}{\nu(T-t)} \) is the (probability) distribution of fill-sizes. The fact that the mapping \( t \mapsto \frac{\nu(T-t)}{\nu(T-t)} \) is non-increasing (in the first-order stochastic dominating sense) implies that the fill sizes do not get “larger” for larger \( t \). The condition is satisfied if e.g. \( \frac{\nu(T-t, \cdot)}{\nu(T-t)} \) is a constant distribution.

The first condition is the analog to Assumption 3 of Konishi [2002]. It is easily satisfied for a large class of parameters. For example, it clearly holds if \( \sigma(\tau) = \sigma \) and \( \nu(\tau) = \nu \) are constant. It is also satisfied if the two quantities vary in such a way so that \( \frac{\sigma^2(\tau)}{\nu(\tau)} \) remains constant. More generally, if \( \sigma(\tau) \) and \( \nu(\tau) \) are differentiable functions of \( \tau \), then the derivative of the above function is given by

\[
\frac{\partial}{\partial \tau} \left[ \frac{\sigma^2(\tau)}{\nu(\tau)} (x - X(\tau)) \right] = \frac{\sigma^2(\tau)}{\nu(\tau)} \left[ \left( \frac{2 \sigma(\tau)}{\sigma(\tau)} - \frac{\nu(\tau)}{\nu(\tau)} \right) (x - X(\tau)) - 1 \right].
\]

(Here and throughout this manuscript, we use \( \dot{\sigma} \) to denote derivative in \( \tau \).) Since \( x - X(\tau) \) remains within the interval \([-1, 1]\) (see (2.1)), the derivative is non-positive for all \( x \) if

\[
\left| 2 \frac{\dot{\sigma}(\tau)}{\sigma(\tau)} - \frac{\dot{\nu}(\tau)}{\nu(\tau)} \right| \leq \frac{1}{T}.
\]

This says that a sufficient condition for Assumption 2.1.2 to be satisfied is when the log-derivative of \( \sigma \) and \( \nu \) remain close to one another. This is the important case since we know from the microstructure literature that \( \sigma^2(\tau) \) and \( \nu(\tau) \) tend to co-vary throughout the day [Ané and Geman, 2000, Cont, 2001, Jain and Joh, 1988].
Remark 2.1.5. We can easily incorporate adverse selection into the model by decreasing the spread $s$. Intuitively, this decreases the cost (relative to limit orders) of a market order because the limit order price is not “as good as it appeared.” Equivalently, we can view this as lowering the ask at which sale limit orders execute so they are less than $s/2$ above the mid.

Definition 2.1.6. We define $\mathcal{A}_v$ to be the set of all pairs $(L,m)$ such that

- $L$ is a prévisible $\mathbb{N}$-valued process and
- $m$ is an adapted, non-decreasing $\mathbb{N}$-valued process (here, we take $\mathbb{N}$ to include 0 by convention) with $m_0 = 0$

and such that the trading process $x = x^{L,m}$ is always within

$$0 \leq x_t \leq 1 \quad \mathbb{P} \times \text{Leb} \text{ a.e.}$$

(2.1)

where Leb is the Lebesgue measure on $[0,T]$ and where all the shares have been executed by time $T$,

$$x_T = 0.$$  

We assume the agent has the linear-quadratic objective,

$$\inf_{(L,m) \in \mathcal{A}_v} \mathbb{E} \left[ \int_0^T \frac{\gamma \sigma^2(T - u)}{2} (X(T-u) - x_u)^2 \, du + \frac{s}{v} dm_u \right].$$

(2.2)

First, the objective (2.2) is thus always finite. Secondly, we observe that (2.2) captures the essential aspects of the problem. There is a quadratic mistracking penalty proportional to the risk aversion $\gamma$ and riskiness of the stock $\sigma(\tau)^2$ and the square of the VWAP tracking error, $(X(T - t) - x_t)^2$. We can view this as a ‘risk’ term. There is also a ‘cost’ term proportional to the number of market orders executed $dm_t$. The utility function can be motivated more directly by considering the CARA utility Remark 2.1.7

Remark 2.1.7. This remark is meant to motivate the choice of the objective (2.2) and the statements are not meant to be rigorous. We assume an exponential utility $u_\gamma(x) =$
\[- \exp(-\gamma x) \text{ with absolute risk aversion } \gamma > 0, \]

\[
\sup_{L,m} \mathbb{E}[u_{\gamma}(\text{vwap} - \text{VWAP})]. \tag{2.3}
\]

As in a traditional utility maximization problem, the broker is penalized for a higher VWAP cost. However, the agent is penalized for fluctuations from the benchmarked VWAP, rather than fluctuations from a constant return, as in a traditional utility maximization problem. Elementary manipulations yield that this is

\[
\inf_{L,m} \mathbb{E}\left[ \exp\left(-\gamma \left( \frac{s}{2} + \int_0^T (x_t - X(T - t)) \, dp_t - \frac{s}{v} dm_t \right) \right) \right]. \tag{2.4}
\]

Observe that since the leading $\frac{s}{2}$ in (2.4) is a constant shift that does not affect the CARA utility, we may simply drop it. This corresponds to benchmarking the cost to the ask (i.e. the price at which limit orders execute) rather than the mid, which we may freely do without changing the optimal policy. We now observe that the crucial term that determines our ‘risk’ exposure is

\[
\frac{\sigma^2(T-t)}{2} (x_t - X(T - t))^2
\]

which partially motivates our original objective (2.2). It turns out that $x_t$ and $m_t$ will be independent of $P_t$ (see Theorem 2.2.1) and so we can write (2.4) as

\[
\inf_{L,m} \mathbb{E}\left[ \exp\left(-\left( \int_0^T \frac{\gamma^2 \sigma^2(T-t)}{2} (x_t - X(T - t))^2 \, dt - \frac{s}{v} dm_t \right) \right) \right]
\]

which bears a close resemblance to (2.2). The CARA utility is useful for understanding the objective and manipulating it into a form that is suitable for dynamic programming (2.4) but is ultimately less tractable than the linear-quadratic objective (2.2). This tradeoff is well-known in the portfolio theory literature Fabozzi [2007, Chapter 2].

This linear quadratic objective bears similarities to the Markowitz investment problem’s risk-return objective. In Markowitz, the investor chooses a mix between risky stocks and a risk-free bond balancing the higher returns from holding stocks with their higher risk. In our model, the risk is not price uncertainty but execution uncertainty. The “risky assets” are the limit orders which offer price improvement at the risk of execution uncertainty.
The “safe asset” are market orders which guarantee execution but cost a bid-ask spread to execute.

While our objective shares similar features with the objective of the standard asset allocation problem, the controls do not. In the standard investment problem, the controls have a linear effect on shares. The challenge of our problem is the non-linear and stochastic effect of $L$ on the number of shares $v_t$. The standard functional-analytic techniques of optimization rely on the objective (2.2) being convex in the control $L$. However, since

$$L \mapsto \Delta Z \land L$$

is concave and non-linear, we see that the objective is not convex in the control process.

Remark 2.1.8. We pause to remark on the relationship with a strand of the insurance literature which is mathematically related to our problem [Asmussen et al., 1998, Goreac, 2008]. Assume insurance claims of size $Y \in dy$ arrive as a Poisson process with a Poisson random measure $\mu(dt, dy)$. Given a claim of size $Y$, the agent then pays out

$$Y \land \alpha_t$$

and a reinsurer pays the excess over the retention level $\alpha_t$, $(Y - \alpha_t)^+$. The insurer’s net premium (after paying the reinsurer) is $p(\alpha_t)$ and the company pays out a dividend $dD_t$, so that its wealth process evolves as

$$X_t = x + \int_0^t p(\alpha_u) du - \int_0^t \int_{\mathbb{R}^+} (y \land \alpha_u) \mu(du, dy) - \int_0^t dD_u$$

and his objective is to maximize dividend payments until bankruptcy $\tau = \inf \{t \geq 0 : X \leq 0\}$,

$$\sup_{D, \alpha} \mathbb{E} \int_0^\tau e^{-ru} dD_u.$$  

The problem is analogous to ours. The claims play the roll of crossing orders and the retention level $\alpha_t$ plays the role of limit orders $L_t$. The dividend $dD_t$ singular control in the insurance problem is analogous to market orders $dm_t$ and they both directly affect the objective. The major differences are that our objective (2.2) is finite horizon and contains a
risk-aversion term which is absent in that of the insurance problem. Furthermore, Asmussen et al. [1998], Goreac [2008] are only able to prove existence and uniqueness of a viscosity solution. In our problem, we are able to prove existence and uniqueness of a classical solution, which in many cases is even closed-form. Finally, while Mnif and Sulem [2005] solves their steady-state setup numerically using generic policy iteration, we are able to derive results specific to our model that yield different numerical techniques.

2.2 General Solution

The continuation value with \( n \) shares left to be executed at time \( t \) is defined as

\[
J_n(T - t) = \inf_{L,m} E_t \left[ \int_t^T \frac{\gamma \sigma (T - u)^2}{2} (X(T - u) - x_u)^2 \, du + \frac{s}{v} \, d\mu \bigg| x_t = \frac{n}{v} \right].
\]

We can write the HJB equation as

\[
0 = \min \left\{ \frac{s}{v} + J_{n-1}(T - t) - J_n(T - t), \min_{L \leq n} \sum_{k=1}^{\infty} \nu(t, k) \left[ J_{n-k\wedge L}(T - t) - J_n(T - t) \right] \right. \\
+ \left. \frac{\gamma \sigma^2 (T - t)}{2} (X(T - t) - x(n))^2 \, dt - \dot{J}_n(T - t) \right\} \quad \text{for } n \geq 1 \tag{2.5}
\]

where \( x(n) = n/v \). The first term represents the effect of executing a market order. The second represents the probability the broker’s limit orders being crossed at the spread. The quadratic term represents the running cost due to slipping away from the desired schedule and the final term is the time value of \( J \). For notational convenience, we now denote \( T - t \) by \( \tau \). The next proposition presents a representation of the solution which is directly numerically computable.
Theorem 2.2.1. The continuation values $J_n$ are given inductively for $n = 0, \ldots, v$ by

$$J_0(\tau) = \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du$$

$$\dot{J}_n(\tau) = \min \left\{ \dot{J}_{n-1}(\tau), \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x(n))^2 - \sum_{k=1}^\infty \nu(\tau, k) \left[ J_n(\tau) - J_{n-k} \right] \right\}$$

$$J_n(0) = \frac{S_n}{v}$$

$$\tilde{L}(\tau, n) = \arg \min_{0 \leq k \leq n} J_k(\tau).$$

These $J$’s are uniformly bounded in $n$, $\tau$, and $v$ and we have

$$\tilde{L}(\tau, n) = n \wedge \tilde{L}(\tau),$$

where $\tilde{L}(\tau) = \tilde{L}(\tau, v)$. We furthermore define the market and limit order times $\tau_m$ and $\tau_l$ as

$$\tau_m^n = \inf \left\{ \tau \in [0, T] : J_n(\tau) - J_{n-1}(\tau) < \frac{S}{v} \right\}$$

$$\tau_l^n = \inf \left\{ \tau \in [0, T] : J_n(\tau) - J_{n-1}(\tau) < 0 \right\}$$

for $n = 1, \ldots, v$ and set $\tau_l^0 = \tau_m^0 = 0$. We have the estimate

$$\tau_m^n \leq T x \left( n - \frac{1}{2} \right) < \tau_l^n. \tag{2.6}$$

The optimal policy is as follows. If at time $t$ there are $n$ shares remaining, we place $L_t = n - \tilde{L}(n, T - t) \geq 0$ limit orders on the book if $\tau \leq \tau_m^n$ and place no limit orders $L_t = 0$ otherwise. A market order is executed (dm fires) as soon as $\tau \leq \tau_l^n$ is satisfied.

The resulting optimal processes $x_t$, $l_t$, and $m_t$ are independent of $P_t$.

We prove the result by demonstrating that the $J$ defined by the integro-differential equations in Theorem 2.2.1 satisfy (2.5) from which we can easily verify the optimality of the controls given in Theorem 2.2.1. From henceforth, we assume that $J$ is as given in Theorem 2.2.1 and its existence follows from Lemma 2.2.2.

Before beginning the proof, we make a few remarks. First, it is clear from the definitions of $\tau_m^n$ and $\tau_l^n$ and the continuity of $J_n(\cdot)$ that

$$\tau_m^n < \tau_l^n.$$
Second, we define
\[ G_n(\tau) = \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x(n))^2 \] (2.7)
\[ F_n(\tau) = \sum_{j=1}^{\infty} \nu(\tau, j) \left[ J_n(\tau) - J_{(n-j)\land L(\tau,n)}(\tau) \right] \] (2.8)
for \( n = 0, \ldots, v \) and by induction, we may rewrite \( \dot{J}_n \) as
\[ \dot{J}_n(\tau) = \min \left\{ \dot{J}_{n-1}(\tau), G_n(\tau) - F_n(\tau) \right\} \]
\[ = \min_{0 \leq k \leq n} \left[ G_k(\tau) - F_k(\tau) \right] . \] (2.9)

We first observe that because \( x(k) = \frac{k}{v} \), we have
\[ G_{n+1}(\tau) - G_n(\tau) = \frac{\gamma \sigma^2(\tau)}{2} \left[ \left( X(\tau) - x(n) - \frac{1}{v} \right)^2 - (X(\tau) - x(n))^2 \right] \]
\[ = \frac{\gamma \sigma^2(\tau)}{2} \left[ \frac{1}{v^2} - \frac{2}{v} (X(\tau) - x(n)) \right] \]
\[ = \frac{\gamma \sigma^2(\tau)}{v} \left( x \left( n + \frac{1}{2} \right) - X(\tau) \right) . \] (2.10)

Finally, we define the function \( H_n \) as
\[ H_n = J_n - J_{n-1} \quad n = 1, \ldots, v . \]

We can use this to rewrite \( F \) as
\[ F_n(\tau) = \sum_{j \geq 1} \nu(\tau, j) \left[ J_n(\tau) - J_{(n-j)\land L(\tau,n)}(\tau) \right] \]
\[ = \sum_{j \geq 1} \nu(\tau, j) \left[ H_n^+(\tau) + H_{n-1}^+(\tau) + \cdots + H_{n+1-j}^+(\tau) \right] \]
and their difference as
\[ F_{n+1}(\tau) - F_n(\tau) = \sum_{j \geq 0} \nu(\tau, j) \left[ H_{n+1}^+(\tau) - H_{n+1-j}^+(\tau) \right] . \] (2.11)

**Lemma 2.2.2.** The function \( J_n(\tau) \) defined in Theorem 2.2.1 is well-defined.
Proof. The proof follows by induction. Observe that $J_0$ is completely defined by the bound on $\sigma$ in Assumption 2.1.1. Now assume $J_0, \ldots, J_{n-1}$ are well-defined. Then a simple inductive argument shows

$$\left| \dot{J}_n(\tau) \right| \leq \left| \dot{J}_{n-1}(\tau) \right| + \gamma \sigma^2(\tau) \left( X(\tau) - x(n) \right)^2 - \sum_{k=1}^{\infty} \nu(\tau, k) \left[ J_n(\tau) - J_{(n-k)\wedge \tilde{L}(\tau,n)}(\tau) \right]$$

Notice that $J_n(\tau)$ only appears in two places, directly as $J_n(\tau)$ and indirectly through $\tilde{L}(\tau,n)$. Define

$$A = \{ \tau : \tilde{L}(\tau,n) < n \} \quad A^c = \{ \tau : \tilde{L}(\tau,n) = n \}$$

and note that since $\tilde{L}(\tau,n) \leq n$, $[0,T]$ is indeed the disjoint union of $A$ and $A^c$. Then for $\tau \in A$, we have $J_{(n-k)\wedge \tilde{L}(\tau,n)}(\tau)$ does not depend on $J_n(\tau)$. For $\tau \in A^c$, we have $J_n(\tau) - J_{(n-k)\wedge \tilde{L}(\tau,n)}(\tau) = 0$. Either way, we have the bound

$$\left| \dot{J}_n(\tau) \right| \leq \alpha(\tau) + \nu(\tau) J_n(\tau)$$

where $\alpha$ is a finite, integrable function on $\tau \in [0,T]$. Hence, we have by an elementary application of Gronwall’s Lemma that

$$\left| J_n(\tau) \right| \leq \alpha(\tau) + \left[ \int_0^T \alpha(u)\nu(u) \exp \left( \left| \int_u^T \nu(r) dr \right| \right) du \right]$$

Furthermore, we have from Assumption 2.1.1 that $\nu(\tau)$ is uniformly bounded in $\tau$ and hence $J_n$ is well-defined.

Based on our definitions, we have the following lemmas whose proofs are straightforward.

**Lemma 2.2.3.** For $n = 1, \ldots, v$, we have

$$J_n(\tau) = J_{n-1}(\tau) + \frac{s}{v} \tau \leq \tau^n$$

and

$$J_n(\tau) < J_{n-1}(\tau) + \frac{s}{v} \tau > \tau^n$$

(2.12)

(2.13)
Using $H$, we can also write these equations as

$$H_n(\tau) = \frac{s}{v} \quad \tau \leq \tau^n_m$$

$$H_n(\tau) < \frac{s}{v} \quad \tau > \tau^n_m$$

(2.14)

and

$$H_n(\tau) \geq 0 \quad \tau \leq \tau^n_l$$

$$H_n(\tau) < 0 \quad \tau > \tau^n_l.$$  

(2.15)

Proof. The proof of (2.12) follows from the observation that $\dot{J}_n(\tau) \leq \dot{J}_{n-1}(\tau)$ from (2.9) and the initial condition $J_n(0) = \frac{sa}{v}$. The proof for (2.13) follows similarly. The ones for (2.14) and (2.15) follow from the definition of $H$. \hfill \Box

The meat of the argument follows from the following Proposition.

**Proposition 2.2.4.** We have for $n = 0, \ldots, v - 1$ that

$$\tau^n_m \leq \tau^{n+1}_m$$

$$\tau^n_l \leq \tau^{n+1}_l.$$  

Furthermore, we have that $J_*(\tau)$ is “convex”, that is

$$H_1(\tau) \leq H_2(\tau) \leq \cdots H_{L(\tau) - 1}(\tau) \leq 0 \leq H_{L(\tau)}(\tau) \leq \cdots \leq H_v(\tau)$$

(2.16)

and that the mapping

$$\tau \mapsto H_n(\tau) \quad \text{is non-increasing.}$$  

(2.17)

Proof. The proof follows by induction on the following inequalities

$$H_k(\tau) \leq H_{k+1}(\tau) \quad \text{for } 0 \leq \tau \leq T$$

(2.18)

$$G_{k+1}(\tau) - F_{k+1}(\tau) < G_k(\tau) - F_k(\tau) \quad \text{for } \tau^{k+1}_m < \tau \leq T.$$  

(2.19)

We assume (2.18) holds for $k = 1, \ldots, n - 1$ and (2.19) holds for $k = 0, \ldots, n - 1$ and we will prove it holds for $k = n$. (The proof of the base case $n = 1$ follows similarly.)
We first make a few preliminary remarks. Observe that (2.16) follows from (2.18) and the definition of $\tilde{L}$ and (2.17) follows from (2.9). Also, it follows from this inductive hypothesis (2.18) and conditions (2.14) and (2.15) that
\[ \tau^k_m \leq \tau^{k+1}_m \]  
\[ \tau^k_l \leq \tau^{k+1}_l \]  
for $k = 0, \ldots, n - 1$. (They hold for $k = 0$ by definition of $\tau^0_m$ and $\tau^0_l$.) Hence, for $k < n$, let $\tau \geq \tau^{k+1}_m$ so that
\[ \tau \geq \tau^{k+1}_m \geq \cdots \geq \tau^0_m \]  
and by the induction hypothesis (2.19) we have
\[ G_{k+1}(\tau) - F_{k+1}(\tau) \leq G_k(\tau) - F_k(\tau) \leq G_{k-1}(\tau) - F_{k-1}(\tau) \leq \cdots \leq G_0(\tau) - F_0(\tau) \]  
and by the definition of $\dot{J}$ (2.9), we have
\[ G_{k+1}(\tau) - F_{k+1}(\tau) \leq G_k(\tau) - F_k(\tau) = \dot{J}_k(\tau) \leq \dot{J}_{k-1}(\tau) \leq \cdots \leq \dot{J}_0(\tau). \]  
Hence, we may conclude that
\[ \dot{J}_{k+1}(\tau) = G_{k+1}(\tau) - F_{k+1}(\tau) \leq G_k(\tau) - F_k(\tau) = \dot{J}_k(\tau) \]  
for $\tau \geq \tau^{k+1}_m$ \hspace{1cm} (2.22)
for $i = 0, \ldots k$ and $k = 0, \ldots, n - 1$. We now show (2.18) and (2.19) for $k = n$.

We now show that $\tau^n_m \leq \tau^{n+1}_m$ by assuming $\tau^n_m > \tau^{n+1}_m$ to show a contradiction. By the induction hypothesis (2.20), we have
\[ \tau^0_m \leq \tau^1_m \leq \cdots \leq \tau^n_m \]  
and so there exists some $k < n$ such that
\[ \tau^k_m \leq \tau^{n+1}_m < \tau^{k+1}_m. \]  
By the definition of $\tau^{n+1}_m$ (2.12), we have $\dot{J}_n(\tau) = \dot{J}_{n+1}(\tau)$ for $\tau \leq \tau^{n+1}_m$ and $\dot{J}_{n+1}(\tau) < \dot{J}_n(\tau)$ for $\tau > \tau^{n+1}_m$. So by the definition of $\dot{J}_{n+1}$ (2.9) there exists a $\tau$ satisfying
\[ \tau \]  
in a neighborhood of $\tau^{n+1}_m$ and $\tau^{n+1}_m \leq \tau < \tau^{k+1}_m$ \hspace{1cm} (2.24)
such that
\[ G_{n+1}(\tau) - F_{n+1}(\tau) \leq \dot{J}_n(\tau). \]
Indeed, the above holds for all \( \tau \) in some neighborhood of \( \tau^{n+1}_m \) satisfying (2.24). The last equation implies that
\[
H_{n+1}(\tau) = H_{n+1}(\tau^{n+1}_m) - \int_{\tau^{n+1}_m}^{\tau} \left[ G_{n+1}(u) - F_{n+1}(u) - \dot{J}_n(u) \right] \, du
= H_{n+1}(\tau^{n+1}_m) + \int_{\tau^{n+1}_m}^{\tau} G_{n+1}(u) - F_{n+1}(u) - \dot{J}_n(u) \, du.
\]
Since \( H_n(\tau^{n+1}_m) \leq \frac{s}{v} = H_{n+1}(\tau^{n+1}_m) \), we can write
\[
H_n(\tau) - H_{n+1}(\tau) \leq \int_{\tau^{n+1}_m}^{\tau} \left[ G_n(u) - F_n(u) - \dot{J}_n(u) \right] \, du
+ \left[ G_{n+1}(u) - F_{n+1}(u) - \dot{J}_n(u) \right] \, du
\leq \int_{\tau^{n+1}_m}^{\tau} G_n(u) - F_n(u) - \dot{J}_n(u) - \left[ G_{n+1}(u) - F_{n+1}(u) - \dot{J}_n(u) \right] \, du
\leq \int_{\tau^{n+1}_m}^{\tau} G_n(u) - F_n(u) - [G_{n+1}(u) - F_{n+1}(u)] \, du
\]
where we have used \( \dot{J}_n \leq \dot{J}_{n-1} \) from (2.9). From (2.10) and the fact that \( x \left( n + \frac{1}{2} \right) - x \left( k + \frac{1}{2} \right) = \frac{n-k}{v} \), we have
\[
G_{n+1}(\tau) - G_n(\tau) = \frac{\gamma \sigma^2(\tau)}{v} \left( x \left( n + \frac{1}{2} \right) - X(\tau) \right)
= G_{k+1}(\tau) - G_k(\tau) + \frac{\gamma \sigma^2(\tau)}{v^2} (n - k)
> G_{k+1}(\tau) - G_k(\tau),
\] (2.25)
where we have used the fact that \( n > k \) in the last line. However (2.23) and (2.24) imply \( \tau^k_m \leq \tau < \tau^{k+1}_m \) and for \( \tau \) within this range have by (2.9) and (2.22) that
\[
G_{k+1}(\tau) - F_{k+1}(\tau) \geq \dot{J}_{k+1}(\tau) = \dot{J}_k(\tau) = G_k(\tau) - F_k(\tau).
\]
Combining the last two results with (2.11), we obtain

\[
H_n(\tau) - H_{n+1}(\tau) < \int_{\tau_m}^{\tau} G_k(u) - G_{k+1}(u) - [F_n(u) - F_{n+1}(u)] \, du
\]

\[
\leq \int_{\tau_m}^{\tau} F_k(u) - F_{k+1}(u) - [F_n(u) - F_{n+1}(u)] \, du
\]

\[
= \int_{\tau_m}^{\tau} \sum_{j \geq 1} \nu(u, j) \left[ H_{n+1}^+(u) - H_{n+1}^+(u-j) \right]
\]

\[
- \sum_{j \geq 1} \nu(u, j) \left[ H_{k+1}^+(u) - H_{k+1}^+(u-j) \right] \, du
\]

\[
= \int_{\tau_m}^{\tau} \sum_{j \geq 1} \nu(u, j) \left[ H_{n+1}^+(u) - H_n^+(u) \right]
\]

\[
+ \sum_{j \geq 1} \nu(u, j) \left[ H_k^+(u) - H_{k+1}^+(u) - \left( H_{n+1-j}^+(u) - H_{k+1-j}^+(u) \right) \right] \, du.
\]

Now, for \( \tau \) within our range (2.24), we have \( \tau < \tau_{m+1}^k \leq \cdots \leq \tau_m^k \). By (2.14) we have

\[
H_{k+1}(u) = \cdots = H_n(u) = \frac{s}{v} \leq \tau_{m+1}^k.
\]

Furthermore, we have from (2.18) that \( H_1^+ \leq \cdots \leq H_n^+ \) and since \( n > k \), it follows that

\[
H_{n+1-j}^+ \geq H_{k+1-j}^+ \quad \text{for} \quad j = 0, \ldots, k - 1.
\]

Hence, we have

\[
H_n(\tau) - H_{n+1}(\tau) < \int_{\tau_m}^{\tau} \sum_{j \geq 1} \nu(u, j) \left[ H_{n+1}^+(u) - H_n^+(u) \right] \, du
\]

\[
\leq \int_{\tau_m}^{\tau} \nu(u) \left[ H_{n+1}(u) - H_n(u) \right] \, du
\]

where we have used the fact that \( \frac{s}{v} \geq H_{n+1}(\tau) > 0 \) for \( \tau \) (in a possibly smaller neighborhood of \( \tau_{m+1}^n \)) satisfying (2.24). A simple application of Gronwall’s Lemma now yields

\[
H_n(\tau) - H_{n+1}(\tau) < 0
\]

which contradicts (2.23) and so \( \tau_{m+1}^n \geq \tau_m^n \).

Next, we prove (2.19) for \( k = n \). Fix \( \tau_{m+1}^n \leq \tau \leq T \). We use the recently-proved fact that \( \tau_{m+1}^n \geq \tau_m^n \) and (2.22) to obtain \( \dot{J}_n(\tau) = G_n(\tau) - F_n(\tau) \). It then follows from (2.9)
and (2.10) that

\[ \dot{H}_{n+1}(\tau) = - \left[ G_{n+1}(\tau) - G_n(\tau) - (F_{n+1}(\tau) - F_n(\tau)) \right]^- \]

\[ = -\nu(\tau) \left[ \varphi(\tau) - H_{n+1}^+(\tau) \right]^- \]  

(2.26)

where

\[ \varphi(\tau) = \frac{\gamma}{\nu(\tau)} \left( x(n + \frac{1}{2}) - X(\tau) \right) + \sum_{j \geq 1} \frac{\nu(\tau,j)}{\nu(\tau)} H_{n+1-j}^+(\tau). \]

Assumption 2.1.2 gives us that the first term of \( \varphi(\tau) \) is (strictly) decreasing in \( \tau \). For the second term, we first use the induction hypothesis (2.22) to obtain

\[ \dot{H}_{k+1}(\tau) = G_{k+1}(\tau) - F_{k+1}(\tau) - [G_k(\tau) - F_k(\tau)] \leq 0 \quad \text{for} \quad \tau \geq \tau_{m}^{k+1} \]

(2.27)

for \( k = 1, \ldots, n-1 \). Now for \( \tau_{m}^{k+1} \leq \tau \leq \tau' \), we have

\[ \sum_{j \geq 1} \frac{\nu(\tau',j)}{\nu(\tau')} H_{n+1-j}^+(\tau') - \sum_{j \geq 1} \frac{\nu(\tau,j)}{\nu(\tau)} H_{n+1-j}^+(\tau) \]

\[ = \sum_{j \geq 1} \frac{\nu(\tau',j)}{\nu(\tau')} \left( H_{n+1-j}^+(\tau') - H_{n+1-j}^+(\tau) \right) + \left( \sum_{j \geq 1} \frac{\nu(\tau',j)}{\nu(\tau')} - \sum_{j \geq 1} \frac{\nu(\tau,j)}{\nu(\tau)} \right) H_{n+1-j}^+(\tau) \]

The first term on the right-hand side is non-positive by the Fundamental Theorem of Calculus and (2.27). The second term on the right-hand side is non-positive since the distribution \( \frac{\nu(\tau,j)}{\nu(\tau)} \) is first-order-stochastic-dominance non-decreasing by Assumption 2.1.2 and \( j \mapsto H_{n-j}^+(\tau) \) is a non-increasing map by the induction hypothesis (2.18). The above then gives us that the second term of \( \varphi(\tau) \) is non-increasing and hence, \( \varphi(\tau) \) is (strictly) decreasing. The mapping \( \tau \mapsto J_n(\tau) \) is differentiable by its definition in Theorem 2.2.1.

By (2.9) and (2.12), we have that \( J_{n+1}(\tau_{m}^{n+1}) = J_n(\tau_{m}^{n+1}) \) and so

\[ \dot{H}_{n+1}(\tau_{m}^{n+1}) = 0. \]

Furthermore, since \( H(\tau) < \frac{\varepsilon}{\nu} \) for \( \tau > \tau_{m}^{n+1} \) by (2.14), we have that \( \dot{H}(\tau) < 0 \) for \( \tau \) in a neighborhood of \( \tau_{m}^{n+1} \) such that \( \tau > \tau_{m}^{n+1} \). Hence, in this neighborhood, (2.26) becomes

\[ \dot{H}_{n+1}(\tau) = \nu(\tau) \left[ \varphi(\tau) - H_{n+1}^+(\tau) \right] \]
and by continuity, this also holds for $\tau = \tau^{n+1}_m$. Therefore, we have

$$0 = \phi(\tau^{n+1}_m) - H^+_{n+1}(\tau^{n+1}_m).$$

Finally, define $\psi$ to solve the linear ODE,

$$\dot{\psi}(\tau) = \nu(\tau) [\phi(\tau) - \psi(\tau)] \quad \psi(\tau^{n+1}_m) = \frac{s}{v}.$$

Since $\phi(\tau)$ is (strictly) decreasing, the solution satisfies

$$\phi(\tau) - \psi(\tau) \leq 0 \quad \text{and} \quad \dot{\psi}(\tau) < 0 \quad \text{for} \quad \tau \geq \tau^{n+1}_m.$$

But taking $\psi$ to be $H^+_{n+1}$ in (2.26) shows that $\psi$ also satisfies (2.26). We have from the unicity of ODEs that $H^+_{n+1}(\tau) = \psi(\tau)$ for $\tau \geq \tau^{n+1}_m$. Therefore $\dot{H}_{n+1}(\tau) < 0$ for $\tau \geq \tau^{n+1}_m$ and we have from (2.26) that (2.19) holds for $k = n$.

Finally, we will prove the inequality (2.18) for $k = n$. From (2.14) we have

$$H_n(\tau) \leq \frac{s}{v} = H_{n+1}(\tau) \quad \text{for} \quad \tau \leq \tau^{n+1}_m.$$

Hence, it only remains to demonstrate the result for $\tau > \tau^{n+1}_m$ so fix such a $\tau$. We have just shown (2.19) for $k = n$ and so we have

$$G_{n+1}(\tau) - F_{n+1}(\tau) - \dot{J}_n(\tau) \leq 0$$

and this implies that for $\tau > \tau^{n+1}_m$, we again have

$$\dot{H}_{n+1}(\tau) = - \left[ G_{n+1}(\tau) - F_{n+1}(\tau) - \dot{J}_n(\tau) \right] = G_{n+1}(\tau) - F_{n+1}(\tau) - \dot{J}_n(\tau).$$

We see that this implies $\tau \mapsto H_{n+1}(\tau)$ is a non-increasing mapping, and by similar logic,
so is \( \tau \mapsto H_n(\tau) \). We can again write

\[
\dot{H}_n(\tau) - \dot{H}_{n+1}(\tau) \leq \left[ G_{n+1}(\tau) - F_{n+1}(\tau) - J_n(\tau) \right] - \left[ G_n(\tau) - F_n(\tau) - J_{n-1}(\tau) \right] = -[G_{n+1}(\tau) - 2G_n(\tau) + G_{n-1}(\tau)] \\
+ \sum_{j \geq 1} \nu(\tau, j) \left[ H^+_{n+1}(\tau) - H^+_{n}(\tau) - H^+_{n+1-j}(\tau) + H^+_{n-j}(\tau) \right] \\
\leq -\nu(\tau) \left[ H_n(\tau) - H^+_{n+1}(\tau) \right] - [G_{n+1}(\tau) - 2G_n(\tau) + G_{n-1}(\tau)] \\
- \sum_{j \geq 1} \nu(\tau, j) \left[ H^+_{n+1-j}(\tau) + H^+_{n-j}(\tau) \right].
\]

By induction hypothesis on (2.18) for \( k = 1, \ldots, n - 1 \), we have \( H^+_{n+1-j}(\tau) + H^+_{n-j}(\tau) \geq 0 \) for \( j \geq 1 \). Similarly, by (2.25), we have that \( G_{n+1}(\tau) - 2G_n(\tau) + G_{n-1}(\tau) \geq 0 \). Finally, for \( \tau^+_{m+1} \leq \tau \leq \tau^{n+1}_l \) we have \( H^+_{n+1}(\tau) = H_{n+1}(\tau) \) and so

\[
\dot{H}_n(\tau) - \dot{H}_{n+1}(\tau) \leq -\nu(\tau) \left[ H_n(\tau) - H_{n+1}(\tau) \right].
\]

Since \( H_n(\tau^{n+1}_m) \leq H_{n+1}(\tau^{n+1}_m) \), we have by Gronwall’s Lemma, that

\[
H_{n+1}(\tau) \geq H_n(\tau) \quad \text{for} \quad \tau^{n+1}_m < \tau \leq \tau^{n+1}_l.
\]

By definition, we have \( H_{n+1}(\tau^{n+1}_l) = 0 \) and so \( \tau^{n}_l \leq \tau^{n+1}_l \). So by (2.15), we have \( H^+_{n}(\tau) - H^+_{n+1}(\tau) = 0 \) for \( \tau > \tau^{n+1}_l \). Therefore \( H^+_{n}(\tau) - H^+_{n+1}(\tau) = 0 \) for \( \tau > \tau^{n+1}_l \). Since \( H_n(\tau^{n+1}_l) - H_{n+1}(\tau^{n+1}_l) \leq 0 \), we have by integrating that

\[
H_n(\tau) - H_{n+1}(\tau) \leq \int_{\tau^{n+1}_l}^{\tau} \left[ G_{n+1}(u) - 2G_n(u) + G_{n-1}(u) \right] \\
- \sum_{j \geq 1} \nu(\tau, j) \left[ H^+_{n+1-j}(u) + H^+_{n-j}(u) \right] du.
\]

and the right-hand is non-positive as before. Hence, we have

\[
H_n(\tau) \leq H_{n+1}(\tau) \quad \text{for} \quad \tau > \tau^{n+1}_l
\]

and hence (2.18) holds for \( k = n \). \( \square \)
Lemma 2.2.5. The equation (2.6) holds.

Proof. For \( \tau \leq \tau_m^n \), we have

\[
G_n(\tau) - F_n(\tau) \geq G_{n-1}(\tau) - F_{n-1}(\tau)
\]

and so by (2.10) we have

\[
\frac{\gamma \sigma^2(\tau)}{v} [x (n - \frac{1}{2}) - X(\tau)] = G_n(\tau) - G_{n-1}(\tau)
\geq F_n(\tau) - F_{n-1}(\tau).
\]

We also have that \( \tau \leq \tau_m^n < \tau^n_l \), and so \( H_n(\tau) > 0 \) and by (2.16) we have \( H_1(\tau) \leq \cdots \leq H_n(\tau) \) so that by (2.11) we have

\[
F_n(\tau) - F_{n-1}(\tau) = \sum_{j=1}^{\infty} \nu(\tau,j) \left[H_n^+(\tau) - H_{n-j}^+(\tau)\right] > 0.
\]

Hence we have \( x (n - \frac{1}{2}) - X(\tau) > 0 \) and thus

\[
\tau < Tx \left(n - \frac{1}{2}\right)
\]

for all \( \tau \leq \tau_m^n \) and in particular for \( \tau = \tau_m^n \). Similarly, for \( \tau \geq \tau^n_l > \tau_m^n \), we have

\[
\frac{\gamma \sigma^2(\tau)}{v} [x (n - \frac{1}{2}) - X(\tau)] = G_n(\tau) - G_{n-1}(\tau)
\leq F_n(\tau) - F_{n-1}(\tau).
\]

However, from (2.15) and (2.16), we have \( H_1(\tau) \leq \cdots \leq H_n(\tau) \leq 0 \) and so by (2.11), we have

\[
F_n(\tau) = F_{n-1}(\tau) = 0
\]

and hence

\[
\tau \geq Tx \left(n - \frac{1}{2}\right)
\]

for all \( \tau \geq \tau^n_l \), and in particular \( \tau = \tau^n_l \) and the result (2.6) then obtains. \( \square \)
Proof of Theorem 2.2.1. The objective (2.2) is clearly non-negative. Executing only clean-up market orders, i.e. adopting the policy

\[ L_t = 0 \quad \text{for all } t \in [0,T] \]

\[ m_t = 0 \quad \text{for all } t \in [0,T) \quad \text{and} \quad m_T = v \]
yields \( x_t = m_t \) and thus

\[ J_n(\tau) \leq \int_0^T \frac{\gamma \sigma^2(u)}{2} du + s \]

which is finite by Assumption 2.1.1. Therefore, \( J_n(\tau) \) is bounded by a constant independent of \( n, \tau, \) and \( v \). We first show that the \( J \) defined in Theorem 2.2.1 satisfies the HJB equation (2.5). By (2.17), we have \( \tau \mapsto J_n(\tau) - J_{n-1}(\tau) \) is decreasing and from the initial condition \( J_n(0) - J_{n-1}(0) = \frac{s}{v} \), we have

\[ \frac{s}{v} + J_{n-1}(\tau) - J_n(\tau) \geq 0. \tag{2.28} \]

From the definition of \( \dot{J}_n \), we have

\[ \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x(n))^2 - \sum_{k=1}^{\infty} \nu(\tau,k) \left[ J_n(\tau) - J_{(n-k)\vee L(\tau,n)}(\tau) \right] - \dot{J}_n(\tau) \geq 0. \tag{2.29} \]

By the (2.12) it is easy to see that the inequality in (2.28) is equality when \( \tau \leq \tau_m^n \) and by (2.22), the inequality in (2.29) is equality when \( \tau \geq \tau_m^n \). Hence, the function \( J_n(\tau) \) defined in Theorem 2.2.1 satisfies the HJB equation (2.5).

It is easy to see that \( J_n \) is attained by the optimal policy. That is, we must show that

\[ \int_0^t \frac{\gamma \sigma(T-u)^2}{2} (X(T-u) - x_u)^2 du + \frac{s}{v} dm_u + J_{v-v}(T-t) \]
is a martingale. We have by (2.16) that setting \( L_t = n - \tilde{L}(T-t) \) minimizes

\[ \sum_{j=1}^{\infty} \nu(T-t,j) [J_{n-(j\wedge L_t)}(T-t) - J_n(T-t)] \] over all possible choices of \( L_t \). \( dm_t \) only fires when \( \tau = \tau_m^n \) where

\[ J_n(\tau_m^n) - J_{n-1}(\tau_m^n) = \frac{s}{v} \]
and the loss to the continuation value perfectly cancels the gain from $\frac{r}{u} dm_t$. The optimality of the prescribed policy then follows. Also, the independence of $(x_t, l_t, m_t)$ and $P_t$ follows directly from the form of our solution and our assumption of the independence of the limit-order crossing process and $P_t$. Remark 2.1.5. It remains to show (2.6), which is done in Lemma 2.2.5.

The execution policy in Theorem 2.2.1 can be broken into three cases. Assume the broker has $n$ shares remaining to execute. For $\tau \leq \tau^n_m$, it is optimal to cross the spread and immediately execute a market order. In this region, he is too far behind VWAP and pays the spread to catch up. Mathematically, we have

$$J_{n-1}(\tau) + \frac{s}{v} < J_n(\tau)$$

so it is worth paying $s$ to reach state $n - 1$. This may continue to hold true for state $n - 1$, in which case another market order would immediately be executed. For $\tau^n_m < \tau \leq \tau^n_l$, the broker places limit orders on the bid but does not execute a market order. In this region, he is comfortably within the tracking region but is willing to accept free crossing executions. Mathematically, we have

$$J_{\tilde{L}(\tau,n)}(\tau) \leq \cdots \leq J_{n-1}(\tau) < J_n(\tau) \leq J_{n-1}(\tau) + \frac{s}{v}.$$ 

It is worth placing limit orders on the book so that he can enter state $n - j$ if there is a crossing order $\Delta Z \geq j$. It is optimal to place $L_t = n - \tilde{L}(\tau,n)$ shares because $\tilde{L}(\tau,n)$ minimizes $J(\tau)$. Placing fewer shares would not take advantage of larger crossing orders, while placing more risks getting too far ahead of VWAP. For $\tau^n_l < \tau$, the broker places neither a limit order nor executes a market order. In this region, he is so far ahead of VWAP that he does not even accept free-executions. Mathematically, we have

$$J_n(\tau) \leq J_{n-1}(\tau)$$

and the broker would prefer to remain in state $n$ rather than drop down to state $n - 1$. 

33
Remark 2.2.6. We are now in a position to explain the rationale for Assumption 2.1.2. We observe that Assumption 2.1.2 was only used to prove (2.19) and hence (2.22). This gave us the critical fact that (2.29) holds with equality for \( \tau \geq \tau_m \). Without this condition, we could have points \( \tau > \tau_m \) such that

\[
G_n(\tau) - F_n(\tau) > G_k(\tau) - F_k(\tau) \quad \text{for } k = 1, \ldots, n - 1
\]

(and hence (2.29) holds with strict inequality) but where

\[
J_n(\tau) - J_{n-1}(\tau) < \frac{s}{v}
\]

(and hence (2.28) holds with strict inequality). Then the HJB equation (2.5) would fail to hold. Hence, Assumption 2.1.2 is used to show that the (readily numerically computable) solution \( J_n \) defined in Theorem 2.2.1 satisfies the HJB condition (2.5).

We conclude by giving bounds that will be used in the next section

**Proposition 2.2.7.** Fix \( k \) in \( \mathbb{N} \). Then we have the following bounds

\[
|G_{n+k}(\tau) - G_n(\tau)| \leq \gamma \sigma^2(\tau) \frac{k}{v} \quad (2.30)
\]

\[
|J_{n+k}(\tau) - J_n(\tau)| \leq K \frac{k}{v} \quad (2.31)
\]

\[
|F_{n+k}(\tau) - F_n(\tau)| \leq K \frac{k}{v} \quad (2.32)
\]

where the constant \( K \) depends only on \( s, \gamma \int_0^T \sigma^2(u) \, du \), and \( \int_0^T \nu(u) \, du \). Furthermore, we have

\[
0 \leq H_{n+k}(\tau) - H_n(\tau) \leq K' \frac{k}{v^2} \quad (2.33)
\]

where the constant \( K' \) depends only on \( \gamma \int_0^T \sigma^2(u) \, du \) and \( \int_0^T \nu(u) \, du \). In particular, both \( K \) and \( K' \) are independent of \( \tau, n, k \) and \( v \).

**Proof.** We prove all these statements for \( k = 1 \). The result for general \( k \) follows from the Triangle Inequality.

We first prove (2.30), which comes from (2.10)

\[
G_{n+1}(\tau) - G_n(\tau) = \frac{\gamma \sigma^2(\tau)}{v} \left( x \left( n + \frac{1}{2} \right) - X(\tau) \right)
\]
and the fact that the point $x(n + \frac{1}{2})$ and the path $X$ are confined within $[0, 1]$. The proof for general $k$ then follows from the Triangle Inequality.

For (2.31), observe that from (2.12), we have that

$$|J_{n+1}(\tau) - J_n(\tau)| > \frac{s}{v} \implies J_{n+1}(\tau) < J_n(\tau) - \frac{s}{v}$$

and so that by (2.6) we have $\tau > \tau^{n+1}_l \geq \tau^{n+1}_m \geq \tau_n^m$ so that

$$\dot{J}_n(\tau) = G_n(\tau) - F_n(\tau) \quad \text{and} \quad \dot{J}_{n+1}(\tau) = G_{n+1}(\tau) - F_{n+1}(\tau).$$

Hence, we have from our previous two results (2.30) and (2.32) that

$$|J_{n+1}(\tau) - J_n(\tau)| \leq \frac{s}{v} + \int_0^\tau |G_{n+1}(u) - G_n(u)| + |F_{n+1}(u) - F_n(u)| \, du$$

An application of Gronwall’s Lemma yields

$$|J_{n+1}(\tau) - J_n(\tau)| \leq \frac{1}{v} \left( s + \gamma \int_0^\tau \sigma^2(u) \, du \right) \exp \left( \int_0^\tau \nu(u) \, du \right)$$

and the result follows from Assumption 2.1.1.

The proof for (2.32) comes from (2.11) and (2.31),

$$|F_{n+1}(\tau) - F_n(\tau)| \leq \sum_{j \geq 1} \nu(u, j) \left| H_{n+1}^+(\tau) - H_{n+1-j}^+(\tau) \right|$$

Finally, the proof of (2.33) also follows from the $k = 1$ case and the Triangle Inequality so we will concentrate on this case. Then $H_{n+1} - H_n$ can be written as a new operator of $J_n$,

$$\Delta_n J_n = J_{n+1} - 2J_n + J_{n-1}.$$
For $\tau \geq \tau_m n^{+1}$, we have

$$\Delta_n \dot{J}_n(\tau) = \Delta_n G_n(\tau) - \Delta_n F_n(\tau)$$

$$= \frac{\gamma \sigma^2(u)}{v^2} - [(F_{n+1}(\tau) - F_n(\tau)) - (F_n(\tau) - F_{n-1}(\tau))]$$

$$= \frac{\gamma \sigma^2(u)}{v^2} - \sum_{j \geq 1} \nu(\tau,j) \left[ (H_{n+1}^+(\tau) - H_n^+(\tau)) - (H_{n+1-j}^+(\tau) - H_{n-j}^+(\tau)) \right].$$

By (2.16), we have

$$0 \leq H_{n+1}^+(\tau) - H_n^+(\tau) \leq H_{n+1}(\tau) - H_n(\tau) = \Delta_n J_n(\tau) \quad \text{for } n = 1, \ldots, v$$

so that

$$\Delta_n \dot{J}_n(\tau) \leq \frac{\gamma \sigma^2(u)}{v^2} + \sum_{j \geq 1} \nu(\tau,j) \left[ (H_{n+1}^+(\tau) - H_n^+(\tau)) - (H_{n+1-j}^+(\tau) - H_{n-j}^+(\tau)) \right]$$

$$\leq \frac{\gamma \sigma^2(u)}{v^2} + \nu(\tau) \sup_n \Delta_n J_n(\tau). \quad (2.34)$$

Observe that the RHS is always non-negative. For $\tau_m \leq \tau < \tau_m^{n+1}$, we have

$$\Delta_n \dot{J}_n(\tau) = -\dot{J}_n(\tau) + \dot{J}_{n-1}(\tau) \leq 0$$

and for $\tau < \tau_m$, we have $\Delta_n \dot{J}_n(\tau) = 0$. Hence, the inequality (2.34) holds true for all $0 \leq \tau \leq T$. An application of Gronwall’s Lemma then yields the bound

$$\sup_n \Delta_n J_n(\tau) \leq \int_0^\tau \sup_n \Delta_n \dot{J}_n(u) \, du$$

$$\leq \frac{1}{v^2} \exp \left( \int_0^T \nu(u) \, du \right) \int_0^\tau \gamma \sigma^2(u) \, du$$

and the result obtains. \hfill \Box

### 2.3 Special Cases and Plots

#### 2.3.1 Single-lot fills

We first analyze the case when $\Delta Z = 1$ a.s. and so we may assume $L_t$ takes only values in \{0, 1\}. 

36
For the remainder of this subsection, assume the broker has \( n \) shares remaining to execute. For \( \tau \leq \tau_m^n \), the broker executes market orders as in the general case. For \( \tau_m^n < \tau \leq \tau_l^n \), the broker places a limit order on the bid but does not execute a market order. In this region,

\[
J_{n-1}(\tau) < J_n(\tau) \leq J_{n-1}(\tau) + s \frac{v}{v}.
\]

The limit is hit with intensity \( \nu(t,1) \), in which case the number of shares remaining goes from \( n \) to \( n - 1 \). Since limit orders are ‘free’ to execute, it is worth placing one on the book in order to reach state \( n - 1 \) but it is not (yet) worth paying \( s \) to cross the spread in order to reach state \( n - 1 \). This differs from the general case in that we can only place 1 share in the order book and thus simplifies our analysis. For \( \tau_l^n < \tau \), the broker places neither a limit order nor executes a market order as in the general case.

Hence, an agent can only be in state \( x(n) \) for \( \tau \in [\tau_m^n, \tau_l^{n+1}] \): \( \tau_l^{n+1} \) is the first point at which it is optimal to place a limit order with \( n + 1 \) shares left to execute and \( \tau_m^n \) is the point beyond which it is optimal to execute a market order to drop to state \( n - 1 \). These results are plotted in Fig. 2.1 and Fig. 2.2. The following corollary makes our above explanation precise.

**Corollary 2.3.1.** The continuation values are given inductively by

\[
\begin{align*}
J_0(\tau) &= \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du \\
J_n(\tau) &= \min \left\{ J_{n-1}(\tau), \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x(n))^2 - \nu(t,1) \left[ J_{(n-1)+}(\tau) - J_n(\tau) \right] \right\} \\
J_n(0) &= \frac{sn}{v}.
\end{align*}
\]

**Proof.** The proof follows from Theorem 2.2.1 by assuming \( \nu(\tau,k) = 0 \) for \( k > 1 \). \( \square \)

### 2.3.2 Large-Fill Distribution

Here, we assume the crossing quantity is infinite, i.e. \( \Delta Z = \infty \) a.s. and these occur with intensity \( \nu(\tau) \). This assumption would be appropriate in the limit where the number of
Figure 2.1: Plots of $X(\tau)$ along with $x(n)$ on the interval ($\tau_m^n, \tau_l^{n+1}$). The plots of $J_n(\tau)$ are also on the interval ($\tau_m^n, \tau_l^{n+1}$). The limit orders are single-lot fills and we used the parameterization $T = 6.5$, $\sigma = 0.5$, $\gamma = 1000$, $\nu = 1.0$ and $s = 10$ with various values of $v = 4, 10$. We assume constant order flow $X(\tau) = \tau/T$, i.e. VWAP = TWAP.

shares to be executed is small compared to the size of crossing orders. Again, assume the broker has $n$ shares remaining to execute. As in the single-fill case, when $\tau \leq \tau_m^n$ market orders are executed immediately. However, unlike in the single-fill case, the broker can place any number $L$ of limit orders on the book when $\tau > \tau_m^n$. The choice of $L$ follows from minimizing the continuation value

$$\min_{L \geq 0} J_{n-L}(\tau).$$

(2.35)

(However, he is in general minimizing a different continuation value $J$ than in the general case because of our specific assumptions on $\nu$). For $\tau_m^n < \tau \leq \tau_l^n$, the broker places $L \geq 1$ shares on the book where $L$ is determined by (2.35). For $\tau > \tau_l^n$, the $L$ that minimizes (2.35) is $L = 0$. These results are plotted in Fig. 2.3 and Fig. 2.4.
Corollary 2.3.2. The continuation values are given inductively by

\[
J_0(\tau) = \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du \\
J_n(\tau) = \min \left\{ J_{n-1}(\tau), \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x(n))^2 + \nu(\tau) \min_{k \geq 0} \left[ J_{(n-k)}(\tau) - J_n(\tau) \right] \right\} \\
J_n(0) = sn / v.
\]

Proof. The proof follows from Theorem 2.2.1 by assuming \( \nu(\tau, k) = 0 \) for all \( k \geq 1 \) finite.

2.3.3 Arrival-Price Benchmark

It turns out the arrival price benchmark is a sub-case of the VWAP problem. Within the framework of our motivating remark Remark 2.1.7, we see that benchmarking to arrival price is given by

\[
\sup_{(L,m) \in \mathcal{A}_v} \mathbb{E} \left[ u \gamma (vwap - P_0) \right].
\]

Comparing with (2.3), we observe that this corresponds to thinking of \( P_0 \) as the daily VWAP where all market trading occurs instantly at the order arrival time, i.e. \( X(T - t) = 0 \) for all \( t \). In other words, the VWAP benchmark liquidated all its shares instantly and the broker is playing ‘catch-up’. Following the line of reasoning, we obtain that the corresponding linear-quadratic objective is

\[
\inf_{(L,m) \in \mathcal{A}_v} \mathbb{E} \left[ \int_0^T \frac{\gamma \sigma^2(T - u)}{2} x^2 du + \frac{s}{v} dm_u \right].
\]

Hence, we have the following theorem, whose proof follows as a corollary of (2.2.1). Observe that in this case, it is optimal to place \( L_t = n \) limit orders on the book as free execution is always desirable when trying to ‘catch up.’
Proposition 2.3.3.

\[ J_0(\tau) = 0 \]
\[ \dot{J}_n(\tau) = \min \left\{ \dot{J}_{n-1}(\tau), \frac{\gamma \sigma^2(\tau)}{2} x^2(n) + \sum_{k=1}^{\infty} \nu(\tau, k) \left[ J_{(n-k)+}(\tau) - J_n(\tau) \right] \right\} \]
\[ J_n(0) = \frac{sn}{v} \]
Figure 2.2: Plots of $X(\tau)$ along with $x(n)$ on the interval $(\tau^n_m, \tau^n_{l+1})$. The plots of $J_n(\tau)$ are also on the interval $(\tau^n_m, \tau^n_{l+1})$. The limit orders are single-lot fills and we used the parameterization $T = 6.5$, $\sigma_0 = 0.25$, $\gamma = 1000$, $\nu_0 = 0.5$ and $s = 10$ with various values of $v = 4, 10$. We assume $\dot{V}(T - t) = bcosh(b(t - T/2))$ with $b = .5$, $\nu(t) = \nu_0 \dot{V}(t)$, $\sigma(t) = \nu_0 \dot{V}(t)$ which simulates the higher trading volume (and hence volatility and intensity of market order arrivals) at the beginning and ending of the trading day, which is what is typically observed.
Figure 2.3: Plots of $X(\tau)$ along with $x(n)$ on the interval $(\tau^n_m, \tau^{n+1}_m)$. The plots of $J_n(\tau)$ are also on the interval $(\tau^n_m, \tau^{n+1}_l)$. The limit orders have ‘large-fill distributions’ and we used the parameterization $T = 6.5$, $\sigma = 0.5$, $\gamma = 1000$, $\nu = 1.0$ and $s = 10$ with various values of $v = 4, 10$. We assume constant order flow $X(\tau) = \tau/T$, i.e. VWAP = TWAP.
Figure 2.4: Plots of \( X(\tau) \) along with \( x(n) \) on the interval \((\tau^m_n, \tau^{n+1}_l)\). The plots of \( J_n(\tau) \) are also on the interval \((\tau^m_n, \tau^{n+1}_l)\). The limit orders have ‘large-fill distributions’ and we used the parameterization \( T = 6.5, \sigma_0 = 0.25, \gamma = 1000, \nu_0 = 0.5 \) and \( s = 10 \) with various values of \( v = 4, 10 \). We assume \( \dot{V}(T - t) = b \cosh(b(t - T/2)) \) with \( b = 0.5, \nu(t) = \nu_0 \dot{V}(t), \sigma(t) = \nu_0 \dot{V}(t) \) which simulates the higher trading volume (and hence volatility and intensity of market order arrivals) at the beginning and ending of the trading day, which is what is typically observed.
Chapter 3

VWAP Tracking – Continuous Case

The subsequent sections are as follows. In Section 3.1, we introduce the continuous setup and the VWAP tracking model with limit and market order controls. The solution is presented in Section 3.2, as is a proof that the continuous solution is the uniform limit of the discrete solution. Special cases are explored in Section 3.3. We solve the stationary limit of the continuous case in Section 3.4, deriving a closed-form solution for the large-fill case Section 3.4.1 and Fourier-Transform-based solution for the general case Section 3.4.2, complete with examples.

3.1 Setup

We now briefly give the setup for the continuous problem (see the discrete case for motivation and justification of our assumptions). We take the continuous-shares limit of the VWAP tracking problem. That is, we take $v, V \to \infty$ so that we may assume $X(T - t)$ and $x_t$ are continuous quantities. The price of the asset is still given by

$$P_t = P_0 + \int_0^t \sigma(T - u) dW_u$$
for a deterministic function $\sigma$. As before, we assume (without loss of generality) that the fraction of shares remaining to be sold $X(T-t)$ and corresponding market VWAP (VWAP) are given by

$$X(T-t) = \frac{T-t}{t} \quad \text{VWAP} = \int_0^T P_t \, dX(t).$$

For the agent, number of remaining shares to trade ($x_{t}^{L,m}$ or $x_t$ for short) and corresponding VWAP (vwap) are given by

$$1 - x_{t}^{L,m} = l_t + m_t \quad \text{vwap} = \int_0^T P_t \, dx_t.$$

That is, the number of shares already executed ($1 - x$) is the sum of market orders ($m$) and limit orders ($l$) to be executed and these are now in units of fractions of shares executed and not shares as in the discrete case. The limit orders are given in terms of the order boundary ($L$), and the size of the crossing shares ($\Delta Z$),

$$l_t = \sum_{s \leq t} L_s \wedge \Delta Z_s.$$

We assume $\Delta Z$ are the poisson arrivals which arrive with time-varying but deterministic intensity with support on $[0, T] \times \mathbb{R}^+$ denoted

$$\nu(T-t, dy)$$

and write $\nu(T-t) = \int_0^\infty \nu(T-t, dy)$ and $\nu = \int_0^\infty \nu(u) \, du$ for notational brevity. The objective is given by

$$\inf_{(L,m) \in \mathcal{A}} \mathbb{E} \left[ \int_0^T \frac{\gamma \sigma(T-t)^2}{2} (X(T-t) - x_t)^2 \, dt + s \, dm_t \right] \quad (3.1)$$

where $\mathcal{A}$ is to be defined below. We observe that we could rewrite $\nu(\tau, dz)$ to have support on $[0, T] \times [0, 1]$. We choose not to because 1: it is possible for incoming crossing orders to be of size greater than 1. More to the point, this convention will be compatible with the asymptotic case Section 3.4.

**Definition 3.1.1.** The set $\mathcal{A}$ is the set of all pairs of processes $(L, m)$ such that

- $L$ is a prévisible $\mathbb{R}^+$-valued process and
\begin{itemize}
  \item $m$ is an adapted, non-decreasing $\mathbb{R}^+$-valued process with $m_0 = 0$
\end{itemize}

and such that the trading process $x = x^{L,m}$ remains within

$$0 \leq x_t \leq 1 \quad \mathbb{P} \times \text{Leb} \; \text{a.e.} \quad (3.2)$$

and where all the shares have been executed by time $T$,

$$x_T = 0 \quad \text{a.s.} \quad (3.3)$$

The objective (3.1) is thus always finite. Here, we take the convention that $\mathbb{R}^+$ includes 0.

We now present the continuous analogs of Assumption 2.1.1 and Assumption 2.1.2 from the discrete case. They are given in Assumption 3.1.2 and Assumption 3.1.3. The first assumption is standard and technical in nature. The second assumption is necessary and the motivation for the analogous discrete assumption is given in Remark 2.2.6.

**Assumption 3.1.2.** We assume that $\nu$ and $\sigma^2$ are bounded in the sense

$$\sup_{\tau \in [0,T]} \nu(\tau) < \infty \quad \text{and} \quad \int_0^T \sigma^2(\tau) \, d\tau < \infty.$$ 

**Assumption 3.1.3.** We assume that for all $x \in [0,1]$, we have the function

$$\tau \mapsto \frac{\sigma^2(\tau)}{\nu(\tau)} (x - X(\tau))$$

is (strictly) decreasing and the probability distribution mapping

$$\tau \mapsto \frac{\nu(\tau, \cdot)}{\nu(\tau)}$$

is non-decreasing in the sense of first-order stochastic dominance. That is, for $\tau \leq \tau'$, we have

$$\int_0^\infty \frac{\nu(\tau, dy)}{\nu(\tau)} f(y) \leq \int_0^\infty \frac{\nu(\tau', dy)}{\nu(\tau')} f(y)$$

for all bounded, non-decreasing functions $f : \mathbb{R}^+ \to \mathbb{R}$. 

46
Definition 3.1.4. For two non-negative Borel measures \( \mu' \) and \( \mu \) on \( \mathbb{R}^+ \) of finite mass, define the metric function \( d(\mu', \mu) \) to be the Wasserstein distance, which we define by its dual representation
\[
d(\mu', \mu) = \sup \left\{ \int_{\mathbb{R}^+} (\mu'(dz) - \mu(dz)) f(z), \|f\|_{Lip} \leq 1 \right\}.
\]
This supremum is written in terms of the space of all Lipschitz-continuous functions and the associated norm
\[
\|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|},
\]
(see Theorem 1.14 and Theorem 7.3 in Villani [2000]).

Lemma 3.1.5. Let \( \mu \) be a non-negative Borel measure of finite mass on \( \mathbb{R}^+ \) and \( v \in \mathbb{N}^+ \). Define
\[
\mu^v(k) = \int_{(\frac{k-1}{v}, \frac{k}{v}]} \mu(dz)
\]
and by abuse of notation define
\[
\mu^v(dz) = \sum_{k=1}^{\infty} \mu^v(k) \delta_{k/v}(dz).
\]
Then we have that
\[
d(\mu^v, \mu) \leq \frac{1}{v} \int_{\mathbb{R}^+} \mu(dz).
\]

Proof. Let \( f \) be a Lipschitz-continuous function satisfying \( \|f\|_{Lip} \leq 1 \). Then
\[
\int_{\mathbb{R}^+} (\mu^v(dz) - \mu(dz)) f(z) = \sum_{k=1}^{\infty} \mu^v(k) f(k/v) - \int_{(\frac{k-1}{v}, \frac{k}{v}]} \mu(dz) f(z)
\]
\[
\leq \sum_{k=1}^{\infty} \mu^v(k) \sup_{z \in (\frac{k-1}{v}, \frac{k}{v}]} |f(\frac{k}{v}) - f(z)|
\]
and the result then follows. \( \square \)

If \( \nu(\tau, dz) \) satisfies Assumption 3.1.2 and Assumption 3.1.3, it is the limit of discrete distributions \( \nu^v(\tau, k) \) which satisfy Assumption 2.1.1 and Assumption 2.1.2 by taking
\[
\nu^v(\tau, k) = \int_{(\frac{k-1}{v}, \frac{k}{v}]} \nu(\tau, dz).
\]
(3.4)
We abuse notation by taking \( \nu^v(\tau, dz) \) to be the induced singular measure on \( \mathbb{R}^+ \) given by the sum of point masses,

\[
\nu^v(\tau, dz) = \sum_{k=1}^{\infty} \nu^v(\tau, k) \delta_{k/v}(dz).
\] (3.5)

By Lemma 3.1.5, \( d(\nu^v(\tau, \cdot), \nu^v(\tau, \cdot)) \to 0 \) for every \( \tau \) and by Assumption 3.1.2, this convergence is uniform in \( \tau \).

Define

\[
J(T-t, x) = \inf_{L,m} \mathbb{E}_t \left[ \int_t^T \frac{\gamma \sigma(T-u)^2}{2} (X(T-u) - x_u)^2 \, du + sdm_u \mid x_t = x \right].
\]

Then the HJB equation is

\[
0 = \min \left\{ s - \partial J(\tau, x), \frac{\gamma \sigma^2(t)}{2} (X(\tau) - x)^2 
+ \min_{0 \leq L \leq x} \int_0^\infty \nu(\tau, dy) [J(\tau, x - (y \wedge L)) - J(\tau, x)] - \dot{J}(\tau, x) \right\}. \tag{3.6}
\]

(Here and throughout this manuscript, we use the convention that \( \dot{J} \) represents derivative in \( \tau \) and \( \partial J \) the derivative in \( x \).)

Remark 3.1.6. We return to an earlier remark about the non-convexity of \( \mathcal{A} \). Let \( T = 3 \) and fix \( \nu \) so that we only have arrivals \( \Delta Z_t \) at \( t = 1, 2 \). While the point holds true for general \( \nu \), fixing the arrival times simplifies our discussion. Under this assumption, it suffices to only consider the number of limit orders \( L_1 \) and \( L_2 \) at times 1 and 2. We specify two trading strategies \( L^a \) and \( L^b \) with corresponding trading paths \( x^a \) and \( x^b \),

\[
\begin{align*}
L^a_1 &= 1 & L^a_2 &= x^a_1 & m^a_t = x^a_2 \mathbb{1}_{t \geq 3} \\
L^b_1 &= \frac{1}{3} & L^b_2 &= x^b_1 & m^b_t = x^b_2 \mathbb{1}_{t \geq 3}
\end{align*}
\]

In words, we place a fixed number of limit orders at time 1, then place the remaining unexecuted shares on the limit-book at time 2, and clean-up at time 3 using market orders. These strategies are in \( \mathcal{A} \) as they satisfy (3.2) and (3.3) for almost sure \( \omega \). Now fix \( \omega \) such
that \( \Delta Z_1 = \frac{2}{3} \) and \( \Delta Z_2 \geq \frac{2}{3} \). Then the path becomes

\[
x^a_t = \begin{cases} 1 & 0 \leq t \leq 1 \\ \frac{1}{3} & 1 < t \leq 2 \\ 0 & 2 < t \leq 3 \end{cases}
x^b_t = \begin{cases} \frac{2}{3} & 1 < t \leq 2 \\ 0 & 2 < t \leq 3 \end{cases}
\]

and our actions are

\[
L^a_1(\omega) = 1 \quad L^b_1(\omega) = \frac{1}{3} \quad m^a_t(\omega) = 0
\]
\[
L^a_2(\omega) = \frac{1}{3} \quad L^b_2(\omega) = \frac{2}{3} \quad m^b_t(\omega) = 0.
\]

However, \( L^c = \frac{1}{2} (L^a_t + L^b_t) \) and \( m^c = \frac{1}{2} (m^a_t + m^b_t) \) results in a trading strategy

\[
L^c_1(\omega) = \frac{2}{3} \quad L^c_2(\omega) = \frac{1}{2} \quad m^c_t(\omega) = 0.
\]

with

\[
x^c_2 = l^c_2 = \sum_{t=1,2} \Delta Z_t \wedge L^c_t = \frac{2}{3} + \frac{1}{2} > 1.
\]

which violates (3.2). Hence, the set \( \mathcal{A} \) is not convex. We also see that the variable \( X_T \) is not even quasi-convex (and hence not convex) in the processes \( (L, m) \). A moment’s reflection will convince the reader that this problem holds for general \( \nu \).

### 3.2 General Solution

We first give the analogous theorem to Theorem 2.2.1.

**Theorem 3.2.1.** The continuation value \( J \) is the solution of the following partial integro-differential equation,

\[
J(\tau, x) = \min \left\{ \min_{0 \leq y < x} J(\tau, y), \frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x)^2 \\
- \int_0^\infty \nu(\tau, dy) \left[ J(\tau, x) - J(\tau, (x - y) \vee \tilde{L}(\tau, x)) \right] \right\}
\]

\[
J(\tau, 0) = \int_0^{\tau} \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du
\]

\[
J(0, x) = sx
\]

\[
\tilde{L}(\tau, x) = \arg\min_{0 \leq y \leq x} J(\tau, y).
\]
Again, we have the analogous result
\[
\tilde{L}(\tau, x) = x \land \check{L}(\tau)
\]
where \(\check{L}(\tau) = \check{L}(\tau, 1)\). We define the market and limit order times \(\tau_m(x)\) and \(\tau_l(x)\) analogously by
\[
\tau_l(x) = \inf \{ \tau \in [0, T] : \partial J(\tau, x) < 0 \}
\]
\[
\tau_m(x) = \inf \{ \tau \in [0, T] : \partial J(\tau, x) < s \}.
\]
and have the estimate,
\[
0 \leq \tau_m(x) \leq T x \leq \tau_l(x).
\]
If \(T - t < \tau_m(x_t)\), the agent executes a singular market order of size
\[
\Delta m_t = x - \inf \{ y : \tau_m(y) > T - t \}
\]
so that he just exits the region. If \(T - t = \tau_m(x_t)\), the agent executes market orders at a fixed rate
\[
dm_t = \frac{1}{\tau'_m(x)}
\]
so that he just remains on the boundary of the region. The function \(\check{L}(\tau)\) is the inverse of \(\tau_l(x)\) and is the threshold below which the agent does not place limit orders for execution. That is, at time \(t = T - \tau\), the agent places \(L_t = x - \check{L}(\tau)\) limit orders on the book if the agent has \(x = x_t\) fraction of shares remaining to execute.

**Proof.** The proof that \(J\) defined in the theorem exists is given in Lemma 3.2.4. We need to prove that this \(J\) satisfies (3.6). We have from Theorem 3.2.2 that \(\tau \mapsto \partial J(\tau, x)\) is a non-increasing function, hence
\[
s - \partial J(\tau, x) = \partial J(0, x) - \partial J(\tau, x) \geq 0. \tag{3.7}
\]
Secondly, by definition,
\[
\frac{\gamma \sigma^2(\tau)}{2} (X(\tau) - x)^2 - \int_0^\infty \nu(\tau, dy) \left[ J(\tau, x) - J(\tau, (x - y) \lor \check{L}(\tau, x)) \right] - \dot{J}(\tau, x) \geq 0. \tag{3.8}
\]
We have from Theorem 3.2.2 that for $\tau \geq \tau_m(x)$, equation (3.7) holds with equality by definition of $\tau_m(x)$. For $\tau \leq \tau_m(x)$, equation (3.8) holds with equality by the continuous version of the discrete result (2.22) and the convergence result Theorem 3.2.2. The rest of the proof is analogous to the one in Theorem 2.2.1 where all the claims in the continuous case follow from their counterparts in the discrete case and the convergence result Theorem 3.2.2.

Theorem 3.2.2. The function $J(\tau, x)$ is the uniform limit of the discrete solutions $J_n(\tau)$ as $n, v \to \infty$. In particular, $J(\tau, x)$ is $C^{1,1}$-differentiable, convex in $x$ and non-increasing in $\tau$. Furthermore, $\partial J(\tau, x)$ is non-positive and strictly negative for $\tau > \tau_m(x)$.

Proof. Let $J^v$ and $\tilde{L}^v$ be from the solution to the discrete problem from Theorem 2.2.1 with $\nu = \nu^v$ given by (3.4). Define $G^v$ and $F^v$ analogously from (2.7) and (2.8), respectively. Second define $\varphi^v_G$ as the linear interpolation of $G^v$

$$
\varphi^v_G(\tau, x) = \lambda G^v_n(\tau) + (1 - \lambda)G^v_{n+1}(\tau) \quad \text{for} \quad x = \lambda \frac{n}{v} + (1 - \lambda)\frac{n + 1}{v}, \quad \lambda \in [0, 1]
$$

and similarly for $\varphi^v_F$ and $\varphi^v_J$. The existence and uniqueness of $J$ defined in the theorem is given by Lemma 3.2.4. However, we will show that $J^v$ converges uniformly to $J$, which provides an independent proof of existence.

Let $v' = 2k'$ and $v = 2k$ for integers $k' > k > 0$ so that $v' > v$. We first show that $\{\varphi^v_J\}_{v=1}^\infty$ is a uniform (in $\tau$ and $x$) Cauchy sequence. Fix

$$
x = \frac{n}{v} = \frac{n'}{v'}
$$

so that $J^v_n(0) = J^v_{n'}(0)$. Then from (2.9), we have

$$
|J^v_n(\tau) - J^v_{n'}(\tau)| \leq \int_0^\tau \min_{j \leq n'} \left[ G^v_j(u) - F^v_j(u) \right] - \min_{j \leq n} \left[ G^v_j(u) - F^v_j(u) \right] \, du
$$

$$
= \int_0^\tau \min_{y \leq x} \left[ \varphi^v_G(u, y) - \varphi^v_F(u, y) \right] - \min_{y \leq x} \left[ \varphi^v_G(u, y) - \varphi^v_F(u, y) \right] \, du \quad (3.10)
$$

$$
\leq \int_0^\tau \max_{y \leq x} \left| \varphi^v_G(u, y) - \varphi^v_F(u, y) \right| + \max_{y \leq x} \left| \varphi^v_G(u, y) - \varphi^v_F(u, y) \right| \, du. \quad (3.11)
$$
The inequality (3.11) comes from Lemma 3.2.3. The equality (3.10) follows because

\[ \varphi^v_G(u, \cdot) - \varphi^v_F(u, \cdot) \]

is a piecewise-linear function whose minimum is attained at a breakpoint \( y = n/v \) for some \( n = 1, \ldots, v \). Furthermore, the breakpoints of the piecewise-linear function

\[ \varphi^v_G(u, \cdot) - \varphi^v_G(u, \cdot) \]  

occur at the the points \( y = n/v' \) for \( n = 0, \ldots, v' \). Hence, there exists \( n'' \) such that the maximum of (3.12) occurs at some point

\[ y = \frac{n''}{v'} \in [0, 1] . \]  

Furthermore, we can find \( \underline{n} \) and \( \overline{n} \) in \( 0, \ldots, v' \) satisfying \( \underline{n} \leq n'' \leq \overline{n} \) and

\[ \frac{\underline{n}}{v'} = \frac{n}{v} \quad \text{and} \quad \frac{\overline{n}}{v'} = \frac{n + 1}{v} . \]  

Then by definition of \( G \), we have

\[ G^{v'}_{\underline{n}}(u) = G^v_{\underline{n}}(u) \quad \text{and} \quad G^{v'}_{\overline{n}}(u) = G^v_{\overline{n} + 1}(u) \]

and there exists some \( \lambda \in [0, 1] \) such that

\[ n'' = \lambda \underline{n} + (1 - \lambda) \overline{n} \]  

and so

\[
\max_{y \leq x} \left| \varphi^{v'}_{\underline{n}}(u, y') - \varphi^{v'}_{\overline{n}}(u, y') \right| = \left| \varphi^{v'}_{\underline{n}}(u, y) - \varphi^{v'}_{\overline{n}}(u, y) \right| \\
\leq \left| G^{v'}_{\underline{n}}(u) - \lambda G^v_{\underline{n}}(u) - (1 - \lambda) G^v_{\overline{n} + 1}(u) \right| \\
\leq \lambda \left| G^{v'}_{\underline{n}}(u) - G^{v'}_{\overline{n}}(u) \right| + (1 - \lambda) \left| G^{v'}_{\underline{n}}(u) - G^{v'}_{\overline{n}}(u) \right| \\
\leq \lambda \gamma \sigma^2(u) \frac{n'' - \underline{n}}{v'} + (1 - \lambda) \gamma \sigma^2(u) \frac{\overline{n} - n''}{v'} \\
\leq \frac{\gamma \sigma^2(u)}{v},
\]
where we have used the bound on $G$ from (2.30) and the fact that $\frac{n-n}{v} = \frac{1}{v}$ in the last two inequalities. By similar logic, we can find a new $y$, $n''$, $\pi$, $\eta$, and $\lambda \in [0, 1]$ satisfying (3.13), (3.14), (3.15) and the analogous conditions for $F$ hold such that

$$\max_{y' \leq y} |\varphi^{uv}_F(u, y') - \varphi^{uv}_F(u, y)| = |\varphi^{uv}_F(u, y) - \varphi^{uv}_F(u, y)|$$

$$= |F^{uv}_n(u) - \lambda F^{uv}_{n-\pi}(u) - (1 - \lambda) F^{uv}_{n+\eta}(u)|$$

$$\leq \lambda |F^{uv}_n(u) - F^{uv}_{n/\eta}(u)| + (1 - \lambda) |F^{uv}_{n-\pi}(u) - F^{uv}_{n+\eta}(u)|$$

$$+ \lambda |F^{uv}_{n/\pi}(u) - F^{uv}_{n+\eta}(u)| + (1 - \lambda) |F^{uv}_{n/\pi}(u) - F^{uv}_{n+\eta}(u)|.$$

The bound for the first two terms follows from the bound on $F$ and $J$ from (2.32)

$$\lambda |F^{uv}_n(u) - F^{uv}_{n/\eta}(u)| + (1 - \lambda) |F^{uv}_{n-\pi}(u) - F^{uv}_{n+\eta}(u)|$$

$$\leq K \nu(u) \left( \frac{n'' - n}{v'} + (1 - \lambda) \frac{n - n''}{v'} \right)$$

$$\leq K \nu(u) \frac{1}{v'}.$$

To show the bound for the second two terms, we first define $\underline{y}$ and $\overline{y}$ such that

$$\underline{y} = \frac{n}{v'} = \frac{n}{v} \quad \text{and} \quad \overline{y} = \frac{\pi}{v'} = \frac{n+1}{v}.$$

If we take $\nu^v(u, dz)$ as in (3.5), then we have

$$F^v_n(u) = \sum_{j \geq 1} \nu^v(u, j) \left[ J^v_n(u) - J^v_{n-j}(L^v(u,v_j)) \right]$$

$$= \int_0^{\infty} \nu^v(u, dz) \left[ \varphi^v_j(u, y) - \varphi^v_j(u, (y - z) \lor (L^v(u, v_j))/v)) \right]$$

$$= \int_0^{\infty} \nu^v(u, dz) \psi^v(u, \underline{y}, z)$$

where

$$\psi^v(u, y, z) = \varphi^v_j(u, y) - \varphi^v_j(u, (y - z) \lor (L^v(u, v_j))/v))$$

and by similar logic

$$F^{uv}_n(u) = \int_0^{\infty} \nu^{uv}(u, dz) \left[ \varphi^{uv}_j(u, y) - \varphi^{uv}_j(u, (y - z) \lor (L^{uv}(u, v'j))/v')) \right]$$

$$= \int_0^{\infty} \nu^{uv}(u, dz) \psi^{uv}(u, \underline{y}, z).$$
Then we have by the Triangle Inequality

\[
|F_n^{v'}(u) - F_n^v(u)| \leq \int_0^\infty |\nu^{v'}(u, dz)\psi^{v'}(u, y, z) - \nu^v(u, dz)\psi^v(u, y, z)| \\
\leq \int_0^\infty \left| (\nu^{v'} - \nu^v) (u, dz)\psi^{v'}(u, y, z) \right| \\
+ \int_0^\infty \nu^v(u, dz) \left( \psi^{v'} - \psi^v \right) (u, y, z) \right|
\]

Observe that since \( J^v \) is uniformly bounded in \( n, \tau, \) and \( v \) from Theorem 2.2.1, we have that the first term is bounded by

\[
\int_0^\infty \left| (\nu^{v'} - \nu^v) (u, dz)\psi^{v'}(u, y, z) \right| \leq Kd \left( \nu^{v'}(u, \cdot), \nu^v(u, \cdot) \right)
\]

where the metric \( d \) is as defined in Definition 3.1.4 and the coefficient \( K \) is the Lipschitz coefficient for \( \psi^v(u, y, z) \) in \( z \) which we showed in (2.31) is independent of \( v \). The second term is given by

\[
\int_0^\infty \nu^v(u, dz) \left( \psi^{v'} - \psi^v \right) (u, y, z) \right| \\
\leq \nu^v(u) \left\| \varphi_j^{v'}(u, \cdot) - \varphi_j^v(u, \cdot) \right\|_\infty
\]

where we have used Lemma 3.2.3. Hence, we have

\[
|F_n^{v'}(u) - F_n^v(u)| \leq Kd \left( \nu^{v'}(u, \cdot), \nu^v(u, \cdot) \right) + 2\nu^v(u) \left\| \varphi_j^{v'}(u, \cdot) - \varphi_j^v(u, \cdot) \right\|_\infty.
\]

and we have the identical for \( |F_{n+1}^{v'}(u) - F_{n+1}^v(u)| \). Observe that the bound is independent of \( n \) and \( n \). Hence, we have

\[
|J_n^{v'}(\tau) - J_n^v(\tau)| \leq \int_0^\tau \frac{1}{v} \left( \gamma\sigma^2(u) + K\nu(u) \right) + Kd \left( \nu^{v'}(u, \cdot), \nu^v(u, \cdot) \right) \\
+ 2\nu^v(u) \left\| \varphi_j^{v'}(u, \cdot) - \varphi_j^v(u, \cdot) \right\|_\infty du.
\]

Finally observe that for any \( x' \in [0, 1] \), we can find an \( x \) such that \( |x - x'| \leq \frac{1}{v} \) with \( x, n, \)
and \( n' \) satisfying (3.9). Then (2.31) and the Triangle Inequality imply

\[
\begin{align*}
\left| \varphi_{j}^{v'}(\tau, x') - \varphi_{j}^{v}(\tau, x') \right| & \leq \left| \varphi_{j}^{v'}(\tau, x') - J_{n}^{v'}(\tau) \right| + \left| J_{n}^{v'}(\tau) - J_{n}^{v}(\tau) \right| + \left| J_{n}^{v}(\tau) - \varphi_{j}^{v}(\tau, x') \right| \\
& \leq J_{n}^{v'}(\tau) - J_{n}^{v}(\tau) + \frac{2K}{v}.
\end{align*}
\]

Finally, an application of Grownwall’s Lemma yields

\[
\left\| \varphi_{j}^{v'}(\tau, \cdot) - \varphi_{j}^{v}(\tau, \cdot) \right\|_{\infty} \leq \exp \left( \int_{0}^{T} 2\nu^{v}(u) \, du \right) \left[ \frac{2K}{v} + \int_{0}^{T} \frac{1}{v} \left( \gamma \sigma^{2}(u) + K \nu(u) \right) + Kd \left( \nu^{v}(u, \cdot), \nu^{v}(u, \cdot) \right) \, du \right].
\]

Then from Assumption 2.1.1 and Lemma 3.1.5, we have that \( \varphi_{j}^{v}(\tau, x) \) where \( v = 2^k \) is uniformly Cauchy as \( v \to \infty \). Hence, the sequence converges to a function, which we dub \( J(\tau, x) \).

To show that \( J \) satisfies the differential equation in the theorem, fix

\[ x = \frac{n}{v} \quad \text{for} \quad v = 2^k, \quad n \leq v, \quad k = 1, 2, \ldots \quad (3.16) \]

We have from (2.9) that

\[ J_{n}^{v}(\tau) = \min_{j \leq n} \left[ G_{j}^{v}(\tau) - F_{j}^{v}(\tau) \right] \]

Taking the limit as \( n, v \to \infty \) with \( x = \frac{n}{v} \) fixed and \( v = 2^k \), we see that

\[
\lim_{n, v \to \infty} G_{n}^{v}(\tau) = (X(\tau) - x)^{2} \frac{\gamma \sigma^{2}(\tau)}{2}
\]

\[
\lim_{n, v \to \infty} F_{j}^{v}(\tau) = \lim_{n, v \to \infty} \int_{0}^{\infty} \nu^{v}(\tau, dy) \left[ \varphi_{j}^{v}(\tau, x) - \varphi_{j}^{v'} \left( u, (y - z) \lor (\tilde{L}^{v}(u, v))/(v) \right) \right]
\]

\[
= \int_{0}^{\infty} \nu(\tau, dy) \left[ J(\tau, x) - J \left( \tau, (x - y) \lor \tilde{L}(\tau, x) \right) \right]
\]

converges (for \( F_{j}^{v} \), we again used the fact that \( J_{n}(\tau) \) is bounded independently of \( n, v, \) and \( \tau \) by Theorem 2.2.1). Hence for \( x \) satisfying (3.16), we have \( \tilde{J}(\tau, x) \) exists and is equal to

\[
\tilde{J}(\tau, x) = \min_{z \leq x} \left\{ (X(\tau) - z)^{2} \frac{\gamma \sigma^{2}(\tau)}{2} - \int_{0}^{\infty} \nu(\tau, dy) \left[ J(\tau, z) - J \left( \tau, (z - y) \lor \tilde{L}(\tau, z) \right) \right] \right\}
\]

\[
= \min \left\{ \min_{0 \leq y < x} \mathcal{J}(\tau, y), \ (X(\tau) - x)^{2} \frac{\gamma \sigma^{2}(\tau)}{2} - \int_{0}^{\infty} \nu(\tau, dy) \left[ J(\tau, x) - J \left( \tau, (x - y) \lor \tilde{L}(\tau, x) \right) \right] \right\}.
\]
This holds for general $x$ by continuity.

That $J(\tau, \cdot)$ is convex follows from the limit of (2.16) and that it is continuous follows from (2.31). We now show that $J(\tau, \cdot)$ is differentiable. We know that $\phi_j^v(\tau, \cdot)$ is differentiable for all $x \neq n/v$ for any $n$. Taking $x$ and $y$ so that

\[
\frac{n - 1}{v} < x < \frac{n}{v} \quad \text{and} \quad \frac{n - 1 + k}{v} < y < \frac{n + k}{v}
\]

we obtain from (2.33) that

\[
\partial \phi_j^v(\tau, y) - \partial \phi_j^v(\tau, x) = v \left( H_n^v(\tau) - H_{n+k}^v(\tau) \right) \leq \frac{k}{v} \leq y - x + \frac{1}{v}.
\]

Hence, the sequence $\partial \phi_j^v(\tau, y)$ is equicontinuous and since it is uniformly bounded by (2.31), we may apply Arzelà-Ascoli to obtain a uniformly bounded subsequence which converges to the continuous function $\partial J$. That is, as $v \to \infty$ such that $n/v \to x$, we have

\[
vH_n^v(\tau) \to \partial J(\tau, x)
\]

In the limit, we have that $\tau_m^n$ and $\tau_l^n$ approach $\tau_m(x)$ and $\tau_l(x)$. We also have from (2.9), (2.10), and (2.11) that

\[
\dot{H}_{n+1}^v(\tau) = - \left[ G_{n+1}(\tau) - G_n(\tau) - (F_{n+1}(\tau) - F_n(\tau)) \right] - \\
= - \left[ G_{n+1}(\tau) - G_n(\tau) - \sum_{j \geq 1} \nu(\tau, j) \left( H_{n+1}^v(\tau) - H_{n+1-j}^v(\tau) \right) \right]
\]

and in the limit as $v \to \infty$, we have

\[
\partial J(\tau, x) = - \left[ \gamma \sigma^2(\tau)(x - X(\tau)) - \nu(\tau)\partial J(\tau, x) + \int_{0}^{L(\tau, x) - x} \nu(\tau, dy)\partial J(\tau, x - y) \right].
\]

Hence the function is $C^{1,1}$ differentiable and non-positive. In fact, it is strictly negative for $\tau > \tau_m(x)$ by taking the limit of (2.19). By taking the limit as $v \to \infty$ in (2.6), we obtain that the thresholds $\tau_m(x)$ and $\tau_l(x)$ satisfy

\[
\tau_m(x) \leq T x \leq \tau_l(x).
\]
It follows from Proposition 2.2.4 that \( \tau_m(\cdot) \) and \( \tau_l(\cdot) \) are non-decreasing functions. It follows from (2.17) that the mapping \( \tau \mapsto \partial J(\tau, x) \) is non-increasing. It is easy to verify that \( \tilde{L}(\tau) \) is the right inverse of \( \tau_l(x) \). The rest of the theorem then follows. \( \square \)

**Lemma 3.2.3.** Let \( f, g \) be two continuous functions on \([0, 1]\). Then we have

\[
\left| \min_x f(x) - \min_x g(x) \right| \leq \| f - g \|_\infty.
\]

If in addition they are convex and attain their minima on \( L_f \) and \( L_g \), respectively, then we have

\[
|f(x \wedge L_f) - g(x \wedge L_g)| \leq \| f - g \|_\infty \quad \text{for all } x \in [0, 1].
\]

**Proof.** First, observe that \( L_f \) and \( L_g \) are guaranteed to exist by Weierstrass’s Theorem. For the first result, observe that

\[
f(L_f) - g(L_g) \leq f(L_g) - g(L_g) \leq \| f - g \|_\infty.
\]

Reversing \( f \) and \( g \) gives us the first result.

For the second result, assume without loss of generality that \( L_f \leq L_g \). The result is clear for \( x \leq L_f \). We now prove the result for \( L_f \leq x \leq L_g \). The function \( g(x) \) is non-increasing for \( x \leq L_g \) so that

\[
f(L_f) - g(x) = f(L_f) - g(L_f) + g(L_f) - g(x) \geq - \| f - g \|_\infty.
\]

Similarly, \( f(x) \) is non-decreasing for \( x \geq L_f \) so that

\[
f(L_f) - g(x) \leq (f(L_f) - f(x)) + (f(x) - g(x)) \leq \| f - g \|_\infty.
\]

Hence, we have the result for \( x \leq L_g \):

\[
|f(x \wedge L_f) - g(x)| \leq \| f - g \|_\infty.
\]

The result for \( x \geq L_g \) follows from taking the limit of the above as \( x \uparrow L_g \) and observing that \( g(x \wedge L_g) = g(L_g) \) for \( x \geq L_g \). The second result thus attains. \( \square \)
Lemma 3.2.4. The equation for $J(\tau, x)$ given in Theorem 3.2.1 is well-defined.

Proof. The first step is to realize that the integro-differential equation can be written

$$
\dot{J}(\tau, x) = \min_{z \leq x} \left\{ (X(\tau) - z)^2 \frac{\gamma \sigma^2(\tau)}{2} - \int_0^\infty \nu(\tau, dy) \left[ J(\tau, z) - J(\tau, (z - y) \lor \tilde{L}(\tau, z)) \right] \right\}
$$

The proof follows from taking the Picard Iterates

$$
\Gamma J(\tau, x) = sx + \int_0^\tau \min_{z \leq x} \left\{ (X(u) - z)^2 \frac{\gamma \sigma^2(u)}{2} \right. \\
\left. - \int_0^\infty \nu(u, dy) \left[ J(u, z) - J(u, (z - y) \lor \tilde{L}(u, z)) \right] \right\} du.
$$

First, note that $\Gamma J(0, x) = sx$ and since $\tilde{L}(0, \tau) = 0$, we have

$$
\Gamma J(\tau, 0) = \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) du.
$$

Hence, regardless of $J$, we know that $\Gamma J$ for will always satisfy the boundary conditions of Theorem 3.2.1. Furthermore, for fixed $\tilde{L} \in \mathbb{R}$, we have

$$
\left| \int_0^\infty \nu(\tau, dy) \left[ J_a(\tau, z) - J_a(\tau, (z - y) \lor \tilde{L}) \right] \\
- \int_0^\infty \nu(\tau, dy) \left[ J_b(\tau, z) - J_b(\tau, (z - y) \lor \tilde{L}) \right] \right| \leq \nu(\tau) \| J_a(\tau, \cdot) - J_b(\tau, \cdot) \|_{\infty}
$$

and so by the first half of Lemma 3.2.3 the Picard Iterates are uniformly Lipschitz continuous,

$$
|\Gamma J_a(\tau, x) - \Gamma J_b(\tau, x)| \leq \nu(\tau) \| J_a(\tau, \cdot) - J_b(\tau, \cdot) \|_{\infty}.
$$

By Assumption 3.1.2, we have after integrating in $\tau$ that

$$
|\Gamma J_a - \Gamma J_b|_{\infty} \leq \nu \| J_a - J_b \|_{\infty}
$$

and applying the usual of Picard argument to the complete Banach Metric Space under the norm $\| \cdot \|_{\infty}$, we have that $J(\tau, x)$ both exists and is unique. \qed

We now draw a picture of the $(\tau, x)$ plane, which is separated into three regions. The first region (B) is the ‘behind’ region defined when the pair $(\tau, x)$ satisfies

$$
\tau < \tau_m(x).
$$
In this region, we have $\partial J(\tau, x) = s$. As soon as trading reaches the boundary of this region $\tau = \tau_m(x)$, the agent executes market orders so as to avoid entering it. The second region is the trade region, $(T)$, and is defined by

$$\tau_m(x) < \tau < \tau_l(x).$$

In this region the agent places $x - \tilde{L}(\tau)$ limit orders on the order book but is not so far behind that he executes market orders. This region contains and tracks the perfect-hedging path, $X(\tau) = \tau/T$. The final region is the ahead region, $(A)$, and is defined by

$$\tau_l(x) < \tau.$$

The agent never enters this region because he would never execute market orders to enter and his limit orders are far enough away to keep him from this region. These regions are diagramed in Fig. 3.1 and labeled (B), (T), and (A) respectively.

![Diagram of three trading regions](image)

Figure 3.1: Diagram of three trading regions: a schematic of three ‘Ahead’ (A), ‘Behind’ (B), and ‘Trade’ (T) regions. Observe that the shaded (T) region tightly tracks an interval around $\tau/T$. 
3.3 Special Cases

3.3.1 Large-fill Distribution

We give the continuous-time case solution for the Large-fill distribution where \( \nu(\tau) \) is the intensity of the (infinitely-sized) crossing orders.

**Corollary 3.3.1.** The continuation value is the solution of the following partial integro-differential equation

\[
\dot{J}(\tau, x) = \min \left\{ \min_{0 \leq y < x} \dot{J}(\tau, y), \frac{\gamma \sigma^2(\tau)}{2} \left( X(\tau) - x \right)^2 - \nu(\tau) \left[ J(\tau, x) - J(\tau, \bar{L}(\tau, x)) \right] \right\}
\]

\[
J(\tau, 0) = \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du
\]

\[
J(0, x) = sx
\]

\[
\bar{L}(\tau, x) = \arg \min_{0 \leq y \leq x} J(\tau, y).
\]

3.3.2 Arrival-Price Benchmark

As in the discrete case, the arrival price benchmark is given by taking \( X(\tau) = 0 \).

**Corollary 3.3.2.** The continuation value is the solution of the following partial integro-differential equation

\[
\dot{J}(\tau, x) = \min \left\{ \min_{0 \leq y < x} \dot{J}(\tau, y), \frac{\gamma \sigma^2(\tau)}{2} x^2 - \int_0^\infty \nu(\tau, dy) \left[ J(\tau, x) - J(\tau, x-y^+) \right] \right\}
\]

\[
J(\tau, 0) = \int_0^\tau \frac{\gamma \sigma^2(u)}{2} X^2(u) \, du
\]

\[
J(0, x) = sx.
\]

3.4 Stationary Limit

To motivate our stationary limit, observe from Fig. 2.3 that the value function \( J_{n+1} \) appears to be the function \( J_n \) shifted up and right. This suggests that for some value function \( V \) and constants \( a \) and \( k \), the function \( J \) (with continuous fill quantities) has the form

\[
J(\tau, x) \approx a + kx + V(\tau - X^{-1}(x)) = a + kx + V(\tau - Tx) \quad \text{for large } (\tau, x).
\]
We expect this to hold for large \( \tau \) and \( x \) as we move away from the boundary condition at \( \tau = 0 \) and \( x = 0 \). We plot a sample function “\( V \)” in Fig. 3.2.

We explore the setup of this new stationary problem and eventually reduce it to a new simplified HJB equation (3.4.1). In this setup, we are able to obtain semi-explicit solutions for a large class of processes and closed-form solutions for a rich family of examples. The rest of the section is as follows. We first solve in Theorem 3.4.2 for \( V \) explicitly for the large-fill case to gain intuition. In Theorem 3.4.3, we prove that for the large-fill case \( V \) precisely equals \( J \) (that is, (3.18) holds with exact equality) for sufficiently large \( \tau \). For general fill distributions \( \nu(dy) \), we are able to obtain closed-form solutions for \( V \) using Fourier methods when \( \nu(dy) \) is exponential (Example 3.4.7), or the “sum of Gamma distributions” (Example 3.4.8).

![Figure 3.2: A sample of \( V \) for the single-lot distribution with \( v = 20 \).](image)

To make this more precise, we must first consider some necessary conditions for this to hold. In particular, it must be the case that \( V \) ‘senses’ the same dynamics parameters regardless of whether they are (jointly) small or large. Therefore, we assume \( \nu(\tau, dy) \) and \( \sigma(\tau) \) are constant in time and the trading volume is still flat throughout the day, that is

\[
\nu(\tau, dy) = \nu(dy) \quad \sigma(\tau) = \sigma \quad X(\tau) = \frac{\tau}{T}.
\]

Hence, the crossing arrivals now occur as a homogenous Poisson process. Under these cir-
cumstances, we are able to simplify the problem by reducing the number of state-variables from two \((\tau, x)\) to one \((\tau - Tx)\). Intuitively, the value function \(V\) given in (3.18) depends only on the distance from tracking VWAP, \(\tau - Tx\). One traditionally measures this distance in units of shares but it is equivalent to measure it in units of time as the daily trading rate \(X(\tau)\) provides the conversion between the two. The agent is perfectly tracking VWAP if \(\tau = Tx\), behind if \(\tau < Tx\), and ahead if \(\tau > Tx\).

Assume a differentiable solution \(V\) exists. Then under our ansatz (3.18), the HJB equation (3.6) for \(V\) would reduce to

\[
V'(\tau - Tx) = \min \left\{ \min_{y \geq 0} V'(\tau - T(x - y)), \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 + \min_{L \geq 0} \int_{0}^{\infty} \nu(dy) \left[ V(\tau - T(x - (y \land L))) - V(\tau - Tx) - k(y \land L) \right] \right\}. \tag{3.19}
\]

We rewrite our variable

\[
\tau \mapsto \tau + Tx
\]

so that \(\tau = 0\) corresponds to being perfectly hedged \((x_t = X(T - t))\). We expect that the market and limit order boundaries are given by the new terms \(\tau_m\) and \(\tau_l\),

\[
\tau_m(x) - Tx \mapsto \tau_m \quad \text{and} \quad \tau_l(x) - Tx \mapsto \tau_l.
\]

Observe that \(\nu(dy)\) induces a measure \(\nu(ds)\) via

\[
\nu(ds) = T \nu(dy)
\]

where we call both \(\nu\) by an abuse of notation. Then the HJB equation simplifies to

\[
V'(\tau) = \min \left\{ \min_{t \geq 0} V'(\tau + t), \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 + \min_{t \geq 0} \int_{0}^{\infty} \nu(ds) \left[ V(\tau + (s \land t)) - V(\tau) - k(s \land t) \right] \right\}. \tag{3.20}
\]

We first write the right-hand case of the outer minimum of (3.20) as

\[
f(\tau) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 + \min_{t \geq 0} \int_{0}^{\infty} \nu(ds) \left[ V(\tau + (s \land t)) - V(\tau) - k(s \land t) \right].
\]
Since \( \min_{t \geq 0} V'(\tau + t) \geq V'(\tau) \) for all \( \tau \) we have that \( V \) must be convex. We have that

\[
V'(\tau) = \begin{cases} 
\min_{t \geq 0} V'(\tau + t) & \text{for } \tau \leq \tau_m \\
 f(\tau) & \text{for } \tau > \tau_m
\end{cases}
\]

Convexity implies \( V'(\tau) \) is nondecreasing so that

\[
V'(\tau) = V'(\tau_m) \quad \text{for } \tau < \tau_m.
\]

Observe that by assumption, \( V \) is continuously differentiable so that

\[
V'(\tau_m) = \lim_{\tau \downarrow \tau_m} V'(\tau) = f(\tau_m).
\]

Furthermore, differentiating both sides of (3.20), we obtain

\[
V''(\tau) = \begin{cases} 
0 & \tau \leq \tau_m \\
f'(\tau) & \tau > \tau_m
\end{cases}
\]

\( f'(\tau) \) is continuous because \( V'(\tau) \) is continuous. Hence, we have that \( V'' \) is continuous if we can show it is continuous at \( \tau = \tau_m \). By convexity of \( V \), we know that \( f'(\tau_m) \geq 0 \). If the equality were strict, then we would have for some \( \varepsilon > 0 \) that \( f(\tau_m - \varepsilon) < f(\tau_m) \) and thus

\[
V'(\tau_m - \varepsilon) = V'(\tau_m) = f(\tau_m) > f(\tau_m - \varepsilon)
\]

which contradicts (3.20). Therefore, we have \( f'(\tau_m) = 0 \) and so \( V'' \) is continuous everywhere.

We have \( \partial J(\tau, x) = s \) for \( \tau \leq \tau_m(x) \) and so

\[
V'(\tau) = \frac{k - s}{T} \quad \text{for } \tau \leq \tau_m.
\] (3.21)

Similarly, since \( \partial J(\tau_l(x), x) = 0 \) and so we have

\[
V'(\tau_l) = \frac{k}{T}.
\]

By convexity of \( V \), we have \( V'(\tau) \leq \frac{k}{T} \) for \( \tau \leq \tau_l \) and vice-versa. In particular, we have that \( \tau_m < \tau_l \). Then for \( \tau \leq \tau_l \), we have

\[
\min_{t \geq 0} \int_0^\infty \nu(ds) \left[ V(\tau + (s \wedge t)) - \frac{k s \wedge \tau_l}{T} \right] = \int_0^\infty \nu(ds) \left[ V((\tau + s) \wedge \tau_l) - \frac{k s \wedge (\tau_l - \tau)}{T} \right].
\]
When \( \tau > \tau_l \), we have \( V'(\tau + t) \geq \frac{k}{T} \) and so

\[
\min_{t \geq 0} \int_0^\infty \nu(ds) \left[ V(\tau + (s \wedge t)) - \frac{k}{T} s \wedge t \right] = V(\tau).
\]

We now summarize our results in the following proposition

**Proposition 3.4.1.** There exists a unique convex \( C^2 \) solution \( V \) to the HJB equation (3.20) with \( \tau_l \) and \( \tau_m \) satisfying the free-boundary conditions

\[
V'(\tau_m) = \frac{k - s}{T} \quad V''(\tau_m) = 0 \quad V'(\tau_l) = \frac{k}{T}.
\]

These three equations allow us to solve for the 3 unknowns \( \tau_l, \tau_m, k \). The differential equation for \( V \) is given by

\[
V'(\tau) = \begin{cases} 
\frac{k - s}{T} & \tau < \tau_m \\
V_0'(\tau) & \tau_m \leq \tau < \tau_l \\
\frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 & \tau \leq \tau_l
\end{cases}
\]

where \( V_0 \) satisfies the equation

\[
V_0'(\tau) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 + \int_0^\infty \nu(ds) \left[ V_0((\tau + s) \wedge \tau_l) - kT \int_0^\infty \nu(ds) \left[ V_0((\tau + s) \wedge \tau_l) - kT \right] \right]
\]

for \( \tau \leq \tau_l \).

### 3.4.1 Large-fill Case

We prove a result for the large-fill case. That is, we assume that all crossing orders are infinitely-sized and we denote by \( \nu \) the intensity of these orders. Then the HJB equation (3.20) simplifies to

\[
V'(\tau) = \min \left\{ \min_{t \geq 0} V'(\tau + t), \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} \right)^2 + \nu \min_{t \geq 0} \left[ V(\tau + t) - V(\tau) - kT \right] \right\}.
\]

When \( \tau > \tau_l \), we have \( V'(\tau + t) \geq \frac{k}{T} \) and so

\[
\min_{t \geq 0} \left[ V(\tau + t) - \frac{k}{T} t \right] = V(\tau).
\]
Since $V'(\tau) = \frac{k}{T} \geq V'(\tau_m)$, we have that

$$V'(\tau_l) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau_l}{T} \right)^2 + \nu \min_{t \geq 0} \left[ V(\tau_l + t) - V(\tau_l) - \frac{k t}{T} \right]$$

$$= \frac{\gamma \sigma^2}{2} \left( \frac{\tau_l}{T} \right)^2.$$

We solve the HJB equation (3.22) in the next theorem (Theorem 3.4.2) and compute the location of the free boundary conditions $\tau_l$ and $\tau_m$. It turns out that the two are related by the Lambert W function. To illustrate the relationship for readers unfamiliar with this transcendental function, we plot the relationship between $\tau_l$ and $\tau_m$ for different values of $\nu$ in Fig. 3.3. The other parameters $\gamma, \sigma, s, T$ only directly affect the optimal policy through $\tau_m$. We observe that the smaller (more negative) the value of $\tau_m$, the larger the value of $\tau_l$, as one would expect.

![Figure 3.3: A plot of the relationship between $\tau_m$ and $\tau_l$. The smaller (more negative) the value of $\tau_m$, the larger the value of $\tau_l$.](image)

**Theorem 3.4.2.** For large-fill crossing orders, the limit-order and market-order boundaries are given by

$$\tau_m = -\frac{\nu s T}{\gamma \sigma^2}$$

$$\tau_l = \frac{1}{\nu} \left( 1 + W \left( -e^{\nu \tau_m - 1} \right) \right)$$
with \( \tau_l > 0 \) and \( \tau_m < 0 \). Here \( W : [-e^{-1}, \infty) \to [-1, \infty) \) is the Lambert W function.

**Proof.** The solution \( V \) is divided into three parts,

\[
V(\tau) = \begin{cases} 
V_0(\tau_m) + \frac{s - k}{T}(\tau_m - \tau) & \tau < \tau_m \\
V_0(\tau) & \tau_m \leq \tau < \tau_l \\
V_0(\tau_l) + \int_{\tau_l}^{\tau} \frac{\gamma \sigma^2}{2} \left( \frac{t}{T} \right)^2 dt & \tau \leq \tau_l 
\end{cases}
\]

where \( V_0 \) is defined by

\[
V_0(\tau) = e^{-\nu \tau} \left[ V_0(\tau_l) e^{\nu \tau_l} + \int_{\tau_l}^{\tau} e^{\nu t} \left[ \frac{\gamma \sigma^2}{2} \left( \frac{t}{T} \right)^2 + \nu \left( V_0(\tau_l) - \frac{k}{T} \tau_l - t \right) \right] dt \right] \\
= V_0(\tau_l) + e^{-\nu \tau} \int_{\tau_l}^{\tau} e^{\nu t} \left[ \frac{\gamma \sigma^2}{2} \left( \frac{t}{T} \right)^2 - \nu \frac{k}{T} (\tau_l - t) \right] dt.
\]

Hence, we see that the constant of integration \( V_0(\tau_l) \) is a ‘gauge symmetry’ (to use the language of theoretical physics): it shifts the value function by a constant but does not affect the optimal behavior. This result was guaranteed by the fact that an arbitrary additive constant to \( V_0(\tau_l) \) can be added to the constant \( a \) in (3.18) without consequence.

It is easy to verify that on \( \tau \geq \tau_m \), we have that the right case of the min in (3.22) is taken and on \( \tau < \tau_m \), we have that \( V \) takes on the left case of the min in (3.22).

We have 3 unknowns \((\tau_m, \tau_l, k)\) and 3 transcendental equations given by the boundary conditions of Proposition 3.4.1. The boundary conditions for \( V''_0(\tau_m) \) and \( V'_0(\tau_l) \) are actually algebraic and give (respectively)

\[
0 = 2\eta \tau_m + \nu s \\
k = \eta \tau_l^2
\]

where \( \eta = \frac{\gamma \sigma^2}{2T} \). To obtain (3.23a), observe that differentiating (3.22) for \( \tau \geq \tau_m \) yields

\[
V''(\tau) = \frac{\gamma \sigma^2}{T^2} \tau + \nu \left[ V'(\tau_l) - V'(\tau) \right]
\]

and the result follows from the expressions for \( V'(\tau_l) \) and \( V'(\tau_m) \). Hence, the optimal boundary comes from solving for the final variable in the boundary condition for \( V'_0(\tau_m) \),

\[
k - s = \frac{d}{dT} \bigg|_{\tau=\tau_m} e^{-\nu \tau} \int_{\tau_l}^{\tau} e^{\nu t} (\eta t^2 - \nu k (\tau_l - t)) dt.
\]
Using (3.23a) and (3.23b), the $t^2$ terms in the above derivative miraculously cancel and we obtain the equation simplifies to

$$1 + e^{\nu(\tau_l - \tau_m)}(\nu \tau_l - 1) = 0.$$ 

The formulas for $\tau_m$ and $\tau_l$ then follow.

We first investigate the relationship between $\tau_m$ and $\tau_l$. Writing $\xi = -e^{\nu \tau_m - 1}$, we note that for $\xi \in (-e^{-1}, 0)$, we have $W(\xi) \in (-1, 0)$ and so $\tau_m \leq 0$ and $\tau_l \geq 0$.

Clearly, if $\tau_m = 0$ then $\tau_l = 0$. A simple calculation yields

$$\frac{d\tau_l}{d\tau_m} = \frac{1}{\nu \xi (1 + W(\xi))} \cdot \frac{d\xi}{d\tau_m}$$

$$= -\frac{W(\xi)}{1 + W(\xi)}$$

$$\geq -1.$$ 

Hence for $\tau_m > 0$.

$$-\tau_m < \tau_l.$$ 

When facing a large-fill distribution, this says that we are more willing to be ahead of the trading schedule ($\tau_l$) than behind ($\tau_m$).

We now give an interpretation for $k$ based on (3.23b). Writing $\xi = -\exp\left(-\frac{\nu^2 T}{2\eta}s - 1\right)$, we note that for $\xi \in (-e^{-1}, 0)$, we have $W(\xi) \in (-1, 0)$ and so $\tau_m < 0$ implies $\tau_l > 0$. We also have that

$$\frac{dk}{ds} = 2\eta \frac{d\tau_l}{ds}$$

$$= 2\eta \frac{1}{\nu^2} (1 + W(\xi)) \cdot W'(\xi) \cdot \frac{d\xi}{ds}$$

$$= 2\eta \frac{1}{\nu^2} (1 + W(\xi)) \cdot \frac{W(\xi)}{\xi(1 + W(\xi))} \cdot -\xi \frac{\nu^2}{2\eta}$$

$$= -W(\xi)$$
and for $\xi \in (-e^{-1}, 0)$, we have $\frac{d k}{d s} = -W(\xi) \in (0, 1)$. Observe that $k = 0$ if $s = 0$ and so we have

$$0 \leq k < s.$$ 

Recall from (3.18) that $k$ is the additional cost of liquidating extra shares $dx$ while keeping the tracking error $\tau - Tx$ fixed. We expect this quantity to be nonnegative (there is a non-zero chance we will have to liquidate these shares via market orders at a marginal cost $s$) but for it to be less than $s$ (there is some chance our limit orders may be filled). The relative size of $k$ gives the likelihood of each of these two cases. The next theorem gives the conditions under which $V$ and $J$ (and hence the optimal policies) coincide. The results are summarized in Fig. 3.4.

Figure 3.4: A schematic of three ‘Ahead’ (A), ‘Behind’ (B), and ‘Trade’ (T) regions with the critical points $x_s$ and $x^*$. By Theorem 3.4.3, we have the derivatives of the value functions $V$ and $J$ agree for $x \geq x_s$ in the trade region (T) and that the the values agree for $x \geq x^*$. The geometry is much more rigid than in Fig. 3.1. Observe that the shaded (T) region tightly tracks an interval around $\tau/T$. 
Theorem 3.4.3. Assume large-fill crossing orders. We first define 

\[ x_* = \frac{\nu s}{\gamma \sigma^2}. \]

Then for \( x \geq x_* \), we have 

\[ \tau_m(x) = T(x - x_*) \quad \tau_l(x) = \tau_l(x_*) + T(x - x_*) \]

where 

\[ \tau_l(x_*) = \frac{\nu s T}{\gamma \sigma^2} + \frac{1}{\nu} \left( 1 + W \left( -\exp \left( -\frac{\nu^2 T^2}{2\eta} s - 1 \right) \right) \right). \]

Then 

\[ \partial J(\tau, x) = k - TV'(\tau - Tx) \quad \text{for } \tau_m(x) \leq \tau \leq \tau_l(x). \]

Second, define 

\[ x^* = x_* + \frac{\tau_l(x_*)}{T}. \]

Then for \( x \geq x^* \), we have 

\[ J(\tau, x) = a + kx + V(\tau - Tx) \quad \text{for } \tau_m(x) \leq \tau \leq \tau_l(x). \]

Proof. We know that for \( \tau \leq \tau_m(x) \), \( \partial J(\tau, x) = s \). By continuity of \( \partial J(\tau, x) \), we have for some \( a \)

\[ \lim_{\tau \downarrow \tau_m(x)} \partial J(\tau, x) = s. \]

Since \( \partial J(\tau, x) \) is continuous for all \((\tau, x)\) by Theorem 3.4.2, we have that if \( \tau_m(x) > 0 \) then

\[ \lim_{\tau \downarrow \tau_m(x)} \partial J(\tau, x) = \lim_{\tau \uparrow \tau_m(x)} \partial J(\tau, x) = 0. \]

Recall that \( \partial \dot{J}(\tau, x) \leq 0 \) by Theorem 3.4.2. For all \( x \) such that \( \tau_m(x) = 0 \), we have

\[ 0 \geq \partial \dot{J}(\tau_m(x), x) = \gamma \sigma^2 \left( x - \frac{\tau_m(x)}{T} \right) - \nu \partial J(\tau_m(x), x) = \gamma \sigma^2 \left( x - \frac{\tau_m(x)}{T} \right) - \nu s \]

Note that for \( x \) such that \( \tau_m(x) > 0 \), this inequality becomes an equality. Hence for \( x \geq x_* \), we have the formula for \( \tau_m(x) \) follows from

\[ \gamma \sigma^2 \left( x - \frac{\tau_m(x)}{T} \right) - \nu s = 0. \]
Then from Theorem 3.4.2 we have the formula for \( \tau_l(x) \) and \( \tau_l(x^*) \), as desired.

Observe that we have from (3.17) that

\[
\partial J(\tau, x) = s + \int_{\tau_m(x)}^{\tau} \gamma \sigma^2 (x - \frac{u}{T}) - \nu \partial J(u, x) \, du. \tag{3.25}
\]

Define

\[ V(\tau, x) = kx + V(\tau - Tx) \]

by an abuse of notation. A simple computation will show that \( \partial V(\tau_m(x), x) \) satisfies the boundary conditions

\[
\partial V(\tau_m(x), x) = s, \quad \partial \dot{V}(\tau_m(x), x) = 0.
\]

Furthermore, by the definition of \( \tau_m \) in Theorem 3.4.2, we have

\[
\tau_m(x) - Tx = -Tx^* = \tau_m \quad \text{for} \ x \geq x^*
\]

and so \( \partial V(\tau, x) \) and \( \partial J(\tau, x) \) satisfy the same equation (3.25) for \( x \geq x^* \). Hence,

\[
\partial V(\tau, x) = \partial J(\tau, x) \quad \text{for} \ x \geq x^* \tag{3.26}
\]

Notice that by definition, \( \tau_l \) solves

\[
\partial V(\tau_l + Tx, x) = -TV'(\tau_l) + k = 0
\]

and we have have (3.25) that \( \tau_l - \tau_m \) is fixed so thus

\[
\tau_l(x) - Tx = \tau_l \quad \text{for} \ x \geq x^*.
\]

Hence, \( x \) minimizes the function \( V(\tau_l(x), \cdot) \). Then rewriting \( \dot{J}(\tau, x) \), we obtain

\[
\dot{J}(\tau, x) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} - x \right)^2 - \nu \left[ J(\tau, x) - J(\tau, \tilde{L}(\tau)) \right] \quad \text{for} \ \tau_m(x) \leq \tau \leq \tau_l(x)
\]

where \( \tilde{L}(\tau, x) \) minimizes \( J(\tau, \cdot) \). It is easy to verify that \( \dot{V}(\tau, x) \) satisfies the above equation as well. Recall that \( \tilde{L}(\cdot) \) and \( \tau_l(\cdot) \) are inverses of one another. For \( x \geq x^* \) and \( \tau_m(x) \leq \tau \leq \tau_l(x) \), we have \( \tilde{L}(\tau) \geq x^* \) (see Fig. 3.4). Rewriting, we obtain

\[
\dot{J}(\tau, x) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau}{T} - x \right)^2 - \nu \int_{\tilde{L}(\tau)}^{x} \partial J(\tau, y) \, dy
\]
and by the above and (3.26) we have

\[ \dot{J}(\tau, x) = \dot{V}(\tau, x) \quad \text{for} \quad x \geq x^* \text{ and } \tau_m(x) \leq \tau \leq \tau_l(x). \]

Finally, we set \( a \) such that

\[ J(\tau_m(x^*), x^*) = a + V(\tau_m(x^*), x^*) \]

Therefore, we have

\[ J(\tau_m(x), x) = J(\tau_m(x^*), x^*) + \int_{x^*}^{x} \left( T\dot{J}(\tau_m(y), y) + \partial J(\tau_m(y), y) \right) dy \]

\[ = V(\tau_m(x^*), x^*) + \int_{x^*}^{x} \left( TV(\tau_m(y), y) + \partial V(\tau_m(y), y) \right) dy \]

\[ = V(\tau_m(x), x) \quad \text{for} \quad x \geq x^* \]

and so

\[ J(\tau, x) = J(\tau_m(x), x) + \int_{\tau_m(x)}^{\tau} \dot{J}(u, x) du \]

\[ = V(\tau_m(x), x) + \int_{\tau_m(x)}^{\tau} \dot{V}(u, x) du \]

\[ = V(\tau, x) \quad \text{for} \quad x \geq x^* \text{ and } \tau_m(x) \leq \tau \leq \tau_l(x). \]

Hence, we have the desired result. \( \square \)

### 3.4.2 General Case

In this section, we solve (3.20) for general \( \nu \). The solution applies to a large class of \( \nu(dy) \) but is presented separately as it does not cover the large-fill distribution case. The solution for \( \tau \geq \tau_l \) and \( \tau < \tau_m \) are already given by Proposition 3.4.1. We then must solve for \( V \) within the range \( \tau_m \leq \tau < \tau_l \). We first define the function \( U \) by the following integro-differential equation,

\[ -U'(t) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau_l - t}{T} \right)^2 + \int_{[0,t]} \nu(ds) \left[ U(t - s) + \frac{k}{T} (t - s) \right] 
+ U(0) \int_{(t,\infty)} \nu(ds) - \nu U(t) - \frac{k\nu}{T} t \quad \text{(3.27)} \]
where we now write $\mathcal{P}$ for the total mass of the finite, non-negative measure $\nu(dy)$ for notational clarity. A simple calculation shows that

$$-U'(t) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau_l - t}{T} \right)^2 + \int_0^\infty \nu(ds) \left[ U(t - (s \wedge t)) - U(t) - \frac{k}{T} (s \wedge t) \right]$$

and so

$$U(t) = V(\tau_l - t) \quad \text{for } t \leq \tau_l - \tau_m.$$

For convenience, define the function $h$ to be

$$h(t) = \frac{\gamma \sigma^2}{2} \left( \frac{\tau_l - t}{T} \right)^2 - \int_0^\infty \nu(ds) \left[ \frac{k}{T} (s \wedge t) \right] \quad \text{for } t \geq 0$$

and $h(t) = 0$ for $t < 0$. Before we begin, we give the following definition,

**Definition 3.4.4.** For a function $h$, we say $h \in M^2$ if

$$\int_R |h(t)| t^2 dt < \infty.$$  

Similarly, for a non-negative measure $\nu$, we say $\nu \in M^2$ if

$$\int_R \nu(dt) t^2 dt < \infty.$$  

**Theorem 3.4.5.** Let $\nu$ be a non-trivial, non-negative, finite-mass, $M^2$ measure with support only to $R^+$. Then

$$U(t) = U(0) - \int_0^t ds \int_R du \left( \int_R d\omega \frac{2\pi i \omega e^{2\pi i u \omega}}{2\pi i \omega - \hat{\nu}(\omega) + \mathcal{P}} \right) h(s - u)$$

where

$$\hat{\nu}(\omega) = \int_R \nu(dt) e^{-2\pi i \omega t}$$

is the Fourier transform of $\nu$.

**Proof.** Like $V$, we observe that the solution $U$ to (3.27) is unique up to an additive constant, which we fix such that $U(0) = 0$. We will solve $U$ in terms of it’s Fourier transform,

$$\hat{U}(\omega) = \int_R U(t) e^{-2\pi i \omega t} dt.$$
Taking the Fourier transform of both sides of (3.27), we obtain

\[ 2\pi i \omega \hat{U}(\omega) = \hat{h}(\omega) + \hat{\nu}(\omega) \hat{U}(\omega) - \nu \hat{U}(\omega). \]

Here, \( \hat{h} \) and \( \hat{\nu} \) are the respective Fourier transforms. Hence, we have

\[ \hat{U}(\omega) = \left[ 2\pi i \omega - \hat{\nu}(\omega) + \nu \right]^{-1} \hat{h}(\omega). \]

If we can show that the term

\[ 2\pi i \omega - \int_{\mathbb{R}} e^{-2\pi i \omega t} \nu(dt) + \int_{\mathbb{R}} \nu(dt) \] (3.28)

does not vanish for \( \omega \in \mathbb{R} \), then the solution makes sense (at least as a distribution). Unfortunately, the solution has one zero at \( \omega = 0 \) but has no other real zeros. The former claim is obvious. To prove the latter claim, first fix \( \omega \not= 0 \) and observe that the term (3.28) can be split into its real and imaginary parts

\[ i \left( 2\pi \omega + \int_{\mathbb{R}} \sin(2\pi \omega t) \nu(dt) \right) - \int_{\mathbb{R}} \cos(2\pi \omega t) \nu(dt) + \int_{\mathbb{R}} \nu(dt) \]

The real part will vanish only if the support of \( \nu \) is contained in the set of \( t \) where \( \omega t \in \mathbb{Z} \) but then the complex part will never vanish. We have furthermore that the reciprocal of the term (3.28) has an order-one zero at \( \omega = 0 \),

\[ 2\pi i \omega - \int_{\mathbb{R}} e^{-2\pi i \omega t} \nu(dt) + \int_{\mathbb{R}} \nu(dt) \sim 2\pi i \kappa \omega \quad \omega \sim 0 \]

where \( \kappa \) is nonzero and given by

\[ \kappa = \left( 1 + \int_{\mathbb{R}_+} t \nu(dt) \right) \geq 1. \]

Since this is the only zero, the function \( \hat{g}(\omega) \) defined by

\[ \hat{g}(\omega) = \frac{2\pi i \omega}{2\pi i \omega - \hat{\nu}(\omega) + \nu} \]

is finite for all \( \omega \in \mathbb{R} \). The transformed equation is then

\[ \hat{U}(\omega) = \frac{\hat{g}(\omega)}{2\pi i \omega} \hat{h}(\omega) \]
The solution to $U$ is (up to a constant)

$$-1_{\mathbb{R}_+} \ast g \ast h = - \int_{\mathbb{R}} ds \int_{\mathbb{R}} du 1_{\mathbb{R}_+}(t-s)g(s-u)h(u) = - \int_{-\infty}^{t} ds \ (g \ast h)(s)$$

Then by the appropriate choice of the additive constant, we have

$$U(t) = U(0) - \int_{0}^{t} ds \ (g \ast h)(s). \quad (3.29)$$

This is simply the desired result. From Lemma 3.4.6 and the fact that $\nu \in M^2$, we have $\hat{\nu} \in C^2$. Indeed, since $\nu$ is a finite-measure, we have that $\hat{\nu}$ is a proper function. Therefore, $\hat{g} \in C^2$ and by another application of Lemma 3.4.6, we have $g \in M^2$. Observe that $h(t) \leq C(1+t^2)$ for some sufficiently large $C$ and so $g \ast h(t)$ is well-defined and continuous for all $t$ (it is, in fact, differentiable). Therefore, the integral in (3.29) is well-defined and we have that $U$ is well-defined.

Lemma 3.4.6. For an arbitrary distribution $\varphi$ on $\mathbb{R}$, we have $\varphi \in M^2$ if and only if $\hat{\varphi}$ is $C^2$ (in the sense of distributions).

Finally, the boundary conditions for $U$ from Proposition 3.4.1 are

$$-U'_{\tau_l - \tau_m} = \frac{k - s}{T}, \quad U''_{\tau_l - \tau_m} = 0, \quad -U'(0) = \frac{k}{T}.$$ 

Example 3.4.7. We tackle the special case

$$\nu(dt) = ae^{-bt}1_{t \geq 0} dt.$$ 

In this case,

$$h(t) = \frac{\gamma\sigma^2}{2} \left( \frac{\tau_l - t}{T} \right)^2 + \frac{ak}{b^2T} \left[ 1 - e^{-bt} \right]$$

and

$$\hat{\nu}(\omega) = \frac{a}{b + 2\pi i \omega}$$

so that we have

$$[2\pi i \omega - \hat{\nu}(\omega) + \nu]^{-1} = \frac{\alpha_0}{2\pi i \omega} + \frac{\alpha_1}{2\pi i \omega + \beta_1}$$

74
where

\[ \alpha_0 = \frac{b^2}{a + b^2}, \quad \beta_1 = \frac{a + b^2}{b}, \quad \alpha_1 = \frac{a}{a + b^2}. \]

Then a simple calculation gives us

\[
U(t) = U(0) - \alpha_0 \int_{\mathbb{R}} 1_{\mathbb{R}_+}(t-u)h(u)\,du - \alpha_1 \int_{\mathbb{R}} 1_{\mathbb{R}_+}(t-u)e^{-\beta_1 u}h(u)\,du
\]

\[
= U(0) - \alpha_0 \int_0^t h(u)\,du - \alpha_1 \int_0^t e^{-\beta_1 (t-u)}h(u)\,du.
\]

Unfortunately, no closed-form solutions for \(\tau_l, \tau_m,\) and \(k\) are known as the miraculous cancellation of the \(t^2\) terms that occur in (3.24) does not occur here. However, we can use the explicit form for \(U\) to solve for \(\tau_l, \tau_m,\) and \(k\) numerically.

Example 3.4.8. There is a large class of \(\nu\) for which the answer can be written semi-explicitly. This is the class of measures

\[
\nu(dt) = \frac{a}{p!} t^p e^{-bt} 1_{t \geq 0} dt
\]

where \(p \in \mathbb{N}\) and \(a, b > 0\). Then an elementary calculation yields

\[
\int_0^\infty s^p e^{-bs} (s \wedge t) \,ds = \int_0^\infty \left( \frac{u}{b} \right)^p e^{-u} \left( \frac{u}{b} \wedge t \right) \frac{du}{b}
\]

\[
= \frac{1}{b^{2+p}} \left[ \int_0^{bt} \frac{u^{p+1}}{p!} e^{-u} \,du + bt \int_{bt}^\infty \frac{u^p}{p!} e^{-u} \,du \right]
\]

\[
= \frac{1}{b^{2+p}} \frac{1}{p!} \left[ \gamma(p+2, bt) + bt \Gamma(p+1, bt) \right]
\]

\[
= \frac{1}{b^{2+p}} e^{-bt} \left[ (p+1) \left( e^{bt} - \sum_{k=0}^{p+1} \frac{(bt)^k}{k!} \right) + bt \sum_{k=0}^{p} \frac{(bt)^k}{k!} \right]
\]

\[
= \frac{1}{b^{2+p}} e^{-bt} \left[ (p+1)e^{bt} + \sum_{k=0}^{p} \left[ \frac{1}{k!} - \frac{p+1}{(k+1)!} \right] (bt)^{k+1} - 1 \right]
\]

where \(\gamma(\cdot, \cdot)\) and \(\Gamma(\cdot, \cdot)\) are the lower and upper incomplete Gamma functions. Then

\[
h(t) = \frac{\gamma^2}{2} \left( \frac{\tau_l - t}{T} \right)^2 - \frac{ak}{T} \frac{1}{b^{2+p}} e^{-bt} \left[ (p+1)e^{bt} + \sum_{k=0}^{p} \left[ \frac{1}{k!} - \frac{p+1}{(k+1)!} \right] (bt)^{k+1} - 1 \right].
\]

We have that \(\nu = \frac{a}{b^{2+p}}\) and by Integration by Parts that

\[
\hat{\nu}(\omega) = \frac{a}{(2\pi i \omega + b)^{p+1}}
\]

75
and so we can write
\[ [2\pi i\omega - \hat{\nu}(\omega) + \mathcal{V}]^{-1} = \frac{(2\pi i\omega + b)^{p+1}}{(2\pi i\omega + \nu)(2\pi i\omega + b)^{p+1} - a}. \]

The denominator and numerator are (real) polynomials in $2\pi i\omega$. Since the degree of the denominator is larger than that of the numerator, the Fundamental Theorem of Algebra and the Method of Partial Fractions gives us the representation
\[ [2\pi i\omega - \hat{\nu}(\omega) + \mathcal{V}]^{-1} = \sum_{k=0}^{m} \frac{\alpha_k}{(2\pi i\omega + \beta_k)^{q_k}} \]
where $\alpha_k$ and $\beta_k$ are real, $q_k \in \mathbb{N}_+$ for each $k = 0, \ldots, m$, and $\sum_k q_k = p + 1$.

To solve the problem, note that if
\[ \hat{g}(\omega) = \frac{1}{(i\omega + \beta)^q} \]
then
\[ g(t) = \frac{(t^+)^q}{q!} e^{-\beta t}. \]
and thus
\[ U(t) = U(0) - \sum_{k=0}^{m} \alpha_k \int_0^t \frac{(t-s)^{q_k}}{q_k!} e^{-\beta_k(t-s)} h(s) ds. \]

We can apply numerical integration to obtain a solution and solve for $\tau_l$, $\tau_m$, and $k$ numerically.

**Example 3.4.9.** The above analysis can be generalized to the class of measures
\[ \nu(dt) = \sum_{k=0}^{n} \frac{\alpha_k}{p_k!} t^{p_k} e^{-b_k t} 1_{t \geq 0} dt. \]
This sum of gamma distributions can be used to approximate virtually any distribution of practical importance. Furthermore, it has a near closed-form solution. We can write
\[ [2\pi i\omega - \hat{\nu}(\omega) + \mathcal{V}]^{-1} = \frac{P(2\pi i\omega)}{Q(2\pi i\omega)} \]
where $P$ and $Q$ are (real) polynomials
\[
\deg P = \sum_{k=0}^{n} (p_k + 1) \quad \deg Q = 1 + \sum_{k=0}^{n} (p_k + 1)
\]
and

$$\mathcal{P} = \sum_{k=0}^{n} \frac{a_k}{b_{P_k} + \mathcal{T}}.$$ 

Since $\deg Q > \deg P$ we have by the Fundamental Theorem of Algebra and the Method of Partial Fractions that

$$[2\pi i \omega - \hat{\nu}(\omega) + \mathcal{P}]^{-1} = \sum_{k=0}^{m} \frac{\alpha_k}{(i \omega + \beta_k) q_k}$$

where $\alpha_k$ and $\beta_k$ are real, $q_k \in \mathbb{N}_+$ for each $k = 0, \ldots, m$, and

$$\sum_{k=0}^{m} q_k = 1 + \sum_{k=0}^{n} (p_k + 1).$$

The expression for $U$ remains as in Example 3.4.8 except with the new $\alpha_k$, $\beta_k$, $q_k$ and $m$.

### 3.5 Conclusion

We proved the analogous results in the continuous for the discrete case in Theorem 3.2.1. We are able to show that this is the limit of discrete results in Theorem 3.2.2, which allows us to transfer the properties from the discrete result Theorem 2.2.1. For the stationary limit, we are able to derive a general characterization of the optimal solution in Proposition 3.4.1 and a closed-form solution for the large-fill case Theorem 3.4.2. Furthermore, we have that the solution in Theorem 3.4.2 is equal to the general continuous case when we are sufficiently far from the “boundary conditions” of Theorem 3.2.1. For the general (non large-fill) case, we can derive the solution in terms of a Fourier transform for a number of very general classes of $\nu$ (see Example 3.4.7, Example 3.4.8, and Example 3.4.9).
Chapter 4

Options Hedging under Illiquidity

The subsequent sections are as follows. We motivate our assumptions and formally set up the problem in Section 4.1. Closed-form solutions for two general cases are presented in Section 4.2. In Section 4.3, we give a number of applications and implications of our result. In Section 4.4, we give a discrete-time formulation of the problem. This is necessary for performing discretized numerical simulations on TAQ data. The discrete-time solution is used in the hedging simulation in Section 4.5.

4.1 Problem Setup

In this section, we first give a heuristic justification for our objective (4.6). The resulting optimal policy for this objective is then rigorously proved in Section 4.2. We give the market impact model, our model for a European option, and our formal objective in the next three subsections.

4.1.1 Market Impact Model

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a complete probability space with \(\mathcal{F}_t\) being the filtration generated by \(W_t\). Consider the trading model

\[
X_t = X_0 + \int_0^t \theta_u \, du
\]
where $X_t$ is the number of shares held by the agent and $\theta_t$ is the intensity of trading. The fundamental price is given by

$$P_t = P_0 + \nu(X_t - X_0) + \sigma W_t$$

(4.1)

where $\nu > 0$ is the coefficient of permanent impact, $\sigma > 0$ is the absolute volatility of the fair value, and $W_t$ is a standard one-dimensional Brownian Motion. Using a linear Brownian Motion rather than a geometric one is appropriate over the short time horizons considered in the paper and leads to dramatic simplifications in our result. Throughout this paper we neglect interests rates, which is appropriate for the short time-periods under consideration.

To model temporary impact, we assume that the $\theta_t$'th share (at time instant $t$) will cost a linear premium $\lambda \theta_t$ over the fair value

$$\tilde{P}_t(\theta_t) - P_t = \lambda \theta_t$$

where $\lambda > 0$ is the market impact parameter. Therefore, the total temporary impact (rate) of buying $\theta_t$ shares is

$$\int_0^{\theta_t} [\tilde{P}_t(\xi) - P_t] d\xi = \lambda \int_0^{\theta_t} \xi d\xi = \frac{\lambda \theta_t^2}{2}.$$  

(4.2)

We can think of temporary impact as coming from a limit-order book with constant depth $1/\lambda$ and instant resilience (see, for example, Alfonsi and Schied [2009], Predoiu et al. [2010]). In such a model, purchasing $\theta_t$ shares would consume all the shares priced from $P_t$ to $P_t + \lambda \theta_t$ on the book, thus pushing the execution cost of the last share up by $\lambda \theta_t$. The limit orders that were eaten up are replaced immediately after execution. Thus our temporary impact coefficient $\lambda$ is related to the corresponding term $\eta$ in Almgren and Chriss [2001] by $\lambda = 2\eta$.

### 4.1.2 European Option

We think of hedging a European contingent claim over a finite time horizon $[0, T]$. For a hypothetical small trader whose execution has no price-impact and trades in a complete
market, the value of the option’s price \( g(t, P_t) \) is a function of the time \( t \) and the price \( P_t \) of the underlying asset,

\[
g : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}.
\]

Hence, the agent’s total portfolio value at time \( t \) is given by \( X_tP_t + g(t, P_t) \). We assume the price at \( t = T \) is given by \( g_0 : \mathbb{R} \rightarrow \mathbb{R} \) and the value for \( t \in [0, T) \) is prescribed by Feynman-Kac to be the solution of the PDE

\[
\dot{g}(t, p) + \frac{\sigma^2}{2} g''(t, p) = 0 \quad \text{for } t, p \in [0, T) \times \mathbb{R} \quad \text{and} \quad g(T, \cdot) = g_0.
\]

(4.3)

Here \( \dot{g} \) represents derivatives with respect to \( t \) and \( g' \) and \( g'' \) represent derivatives with respect to \( p \). We can view this as the Black-Scholes option-pricing PDE for the arithmetic Brownian Motion \( P_t \) and zero interest rates so that \( g \) has the interpretation of the option price in the corresponding (no impact-cost) complete-market. For notational simplicity, we define

\[
\Gamma(t, p) = g''(t, p)
\]

to be the gamma of the option.

For a large trader who does face market-impact, the terminal wealth (or revenue) \( R_T \) is the sum of the option’s value, the stock’s value, and the cost of acquiring the position,

\[
R_T = g(T, P_T) + X_TP_T - \int_0^T \int_0^T \theta_t \tilde{P}_t(\xi)d\xi \, dt.
\]

(4.4)

Using integration by parts, the stock dynamics (4.1), the total temporary impact (4.2), and Feynman-Kac (4.3), we may rewrite the terminal wealth as

\[
R_T = R_0 + \int_0^T [X_t + g'(t, P_t)] \, dP_t - \frac{\lambda}{2} \int_0^T \theta_t^2 dt,
\]

where

\[
R_0 = g(0, P_0) + X_0P_0.
\]

It is now convenient to introduce the variable

\[
Y_t = X_t + g'(t, P_t)
\]

80
with the interpretation as the portfolio's **net delta exposure**. This variable replaces net shares held \( X_t \) as a more relevant variable. \( Y_t \) also has the interpretation of **distance from being delta-hedged** as the ideal net delta-exposure is 0. Hence, we may decompose our wealth process into three intuitive parts

\[
R_T = R_0 + \int_0^T \underbrace{Y_u \sigma dW_u}_\text{Intraperiod Fluctuation} + \underbrace{Y_u \nu \theta_u du}_\text{Permanent Impact} - \underbrace{\frac{\lambda \theta^2_u du}{2}}_\text{Temporary Impact}.
\] (4.5)

If we interpret the interval \([0, T]\) as the trading day, we have the wealth \( R_T \) as the sum of the fluctuation during the trading day and the liquidity cost from permanent and temporary impacts.

Finally, we make the assumption

**Assumption 4.1.1.** Let \( g(t, p) \) have a gamma \( \Gamma(t, p) \) that is bounded everywhere and Lipschitz in \( p \) with a Lipschitz constant that is independent of \( t \).

The above assumption holds, e.g., for the Delta and Gamma of calls and puts except at expiry and so this formula will be valid for hedging except on the expiration day of the option. The case of hedging at expiry is well-known for being difficult because of the divergence in Gamma and these complications are beyond the scope of this paper.

### 4.1.3 Objective

**Assumption 4.1.2.** Assume \( \xi \) and \( \Delta P_T \) are \( L^2 \) random variables satisfying

1. the conditional distribution \( \mathbb{P} [\xi \mid \mathcal{F}_T] \) depends only on \( P_T \).
2. the conditional distribution \( \mathbb{P} [\Delta P_T \mid \mathcal{F}_T] \) is independent of \( \mathcal{F}_T \) and has mean zero.

We define our linear-quadratic objective to be

\[
\inf_{\theta \in \Theta} \mathbb{E} \left[ \frac{\gamma}{2} (Y_T \Delta P_T - \xi)^2 + \int_0^T \frac{\gamma \sigma^2}{2} Y_u^2 - Y_u \nu \theta_u + \frac{\lambda \theta^2_u}{2} du \right],
\] (4.6)
where $\xi$ and $\Delta P$ are to be defined later and the control set $\Theta$ is given by

$$
\Theta := \{ \theta \text{ predictable} : E \int_0^T \theta_s^2 ds < \infty \text{ and } \theta_t \leq C(1 + |Y_t|) \text{ a.s. for all } t \}.
$$

This objective is standard in the literature and can be thought of as a finite-horizon adaptation of Garleanu and Pedersen [2009]. The relationship to utility optimization in the portfolio theory literature is well-known and given in Fabozzi [2007, Chapter 2]. The first term in (4.6) gives a terminal (time $t = T$) penalty for being mishended. We allow this terminal penalty to be quite flexible so that it can accommodate multiple interpretations (through multiple interpretations of $\xi$ and $\Delta P$: see Examples 4.1.3, 4.1.4, and 4.1.5). The running terms penalize for intraperiod misheding fluctuations, and permanent and temporary impact. Equation (4.6) has a natural interpretation of balancing the temporary and permanent liquidity costs with the penalty for variance of the intraperiod and terminal misheding errors.

The state variables have dynamics

$$
dP_t = \nu \theta_t dt + \sigma dW_t
$$

(4.7)

$$
dY_t = (1 + \nu \Gamma(t, P_t)) \theta_t dt + \sigma \Gamma(t, P_t) dW_t.
$$

(4.8)

(Recall that $\Gamma$ is the gamma of the option). It is easy to see that $P_t$ is a continuous semimartingale and hence predictable. We see from (4.1.1) that since $\Gamma$ is bounded that $Y_t$ is well-defined for all $\theta \in \Theta$. We see from (4.7) that every share purchased $\theta_t$ pushes $P_t$ up by $\nu$ because of the permanent impact. We also see from (4.8) that every share purchased $\theta_t$ increases his net delta position $Y_t$ by $1 + \nu \Gamma(t, P_t)$: 1 for the increase in the stock position and $\nu \Gamma(t, P_t)$ for the permanent impact’s effect on the option’s delta.

**Example 4.1.3.** We could choose to interpret $T$ as the maturity of the option. In this case, the misheding penalty is proportional to the time to maturity (i.e. the time remaining to potentially accumulate misheding). Hence, one might choose to not include a terminal term ($\Delta P_T = 0$ and $\xi = 0$) and no additional penalty is incurred for being mishedged at maturity

$$
E \left[ \frac{1}{2} (Y_T \Delta P_T - \xi)^2 \right] = 0.
$$
However, the individual trader may choose to incorporate a penalty because he finds it too risky to be too mishedged even at expiry (see Remark 4.2.3).

**Example 4.1.4.** For general intraday trading, $T$ would represent the close of trading. However, there is an unhedgable potential overnight jump and trading does not resume until the next morning. Therefore, we might choose to benchmark the objective to the opening of trade the next morning. The choice of benchmarking to the morning is natural because the objective can be reset for the next trading day. (Of course, the agent may choose to not incorporate a terminal penalty, even on the day of the option’s maturity, as in Example 4.1.3).

For practical applications, adding a terminal mishedging penalty is more realistic. Without one, the agent’s incentive to be hedged would (unrealistically) vanish as he approaches the closing bell. This extra penalty reflects the need to be well-hedged for the overnight jump, during which, he is unable to trade.

To make this more precise, let $T'$ be the opening of trading the next morning, $T' \geq T$, and let $\Delta T = T' - T$ denote the overnight time interval. Between $T$ and $T'$, the agent is not allowed to trade but the underlying price continues to evolve. From (4.4), we obtain

$$
R_{T'} = g(T', P_{T'}) + X_T P_{T'} - \int_0^T \int_0^{\theta_t} P_t(\xi) d\xi dt \\
= R_T + g(T', P_{T'}) - g(T, P_T) + X_T (P_{T'} - P_T) \\
= R_T + \int_T^{T'} \left[ g'(t, P_t) - g'(T, P_T) \right] dP_t + Y_T (P_{T'} - P_T).
$$

Then we can set the terms in the terminal penalty to

$$
\xi = -\int_T^{T'} \left[ g'(t, P_t) - g'(T, P_T) \right] dP_t
$$

$$
\Delta P_T = P_{T'} - P_T.
$$

and the terminal penalty reduces to the quadratic (in $Y_T$)

$$
\mathbb{E} \left[ \frac{\gamma}{2} (Y_T \Delta P_T - \xi)^2 \right] = \frac{\gamma}{2} \mathbb{E} \left[ \sigma_T^2 Y_T^2 - 2 \mathbb{E} [\Delta P_T \xi | \mathcal{F}_T] Y_T + \mathbb{E} [\xi^2 | \mathcal{F}_T] \right].
$$
where $\sigma_T^2 = \sigma^2 \Delta T$. Observe that the distribution of $\xi$ does not depend on $P_T$ and so satisfies Assumption 4.1.2.

**Example 4.1.5.** One highly illustrative subcase of Example 4.1.4 is when we can approximate the option as having a constant gamma, i.e.

$$g'(t, P_t) \approx g'(t, P_0) + \Gamma(P_t - P_0) \quad \text{with } \Gamma \in \mathbb{R} \text{ constant}.$$  

In other words, we assume $\Gamma(t, p) \approx \Gamma$. Under this assumption, the terms in the terminal penalty simplify down to

$$\xi = - \int_T^{T'} \Gamma(P_t - P_T) dP_t$$

$$\Delta P_T = P_{T'} - P_T .$$

and so we can use Itô Calculus to derive

$$-E [\Delta P_T \xi \mid P_T = p] = E \left[ \sigma^2 \Gamma \int_T^{T'} g'(u, P_u) - g'(T, P_T) du \mid P_T = p \right] = 0$$

(see Subsection 4.2.3 for the full calculation). Following directly from Example 4.1.4, the terminal objective would be

$$E \left[ \frac{\gamma}{2} (Y_T \Delta P_T - \xi)^2 \right] = \frac{\gamma}{2} E \left[ \sigma_T^2 Y_T^2 + \xi^2 \right] .$$

### 4.2 Continuous-time Solution

Let $J(t, p, y)$ be a $C^{1,2,2}([0, T] \times \mathbb{R} \times \mathbb{R})$ function. Then from the Martingale Principle of Optimal Control, we have the HJB equation is

$$0 = \inf_\theta \left[ \frac{\gamma \sigma_T^2}{2} y^2 - y \nu \theta + \frac{\lambda}{2} \theta^2 + \dot{J} + (1 + \nu \Gamma) \theta \partial_y J + \nu \theta \partial_p J ight.$$

$$+ \sigma^2 \Gamma \partial_y \partial_p J + \frac{(\sigma \Gamma)^2}{2} \partial_y^2 J + \frac{\sigma^2}{2} \partial_p^2 J \right]$$

$$= - \frac{1}{2\lambda} \left[ \nu (y - \partial_p J) - (1 + \nu \Gamma) \partial_y J \right]^2 + \frac{\gamma \sigma_T^2}{2} y^2 + \dot{J}$$

$$+ \sigma^2 \Gamma \partial_y \partial_p J + \frac{(\sigma \Gamma)^2}{2} \partial_y^2 J + \frac{\sigma^2}{2} \partial_p^2 J .$$

(4.9)
We will not solve the equation in full generality but will stick to two major sub-cases. For these, it suffices to make the Ansatz,

$$J(t, p, y) = \frac{A_2(T - t)}{2} y^2 + A_1(T - t, p)y + A_0(T - t, p).$$

We will later verify that this Ansatz holds in the proofs of Theorem 4.2.2 and Theorem 4.2.4. Observe that $A_2$ has no dependence on $p$. This requires Assumption 4.1.2 and will be key to our decomposition. Then the equation breaks up into the following three equations

$$\dot{A}_2 = \gamma \sigma^2 - \frac{1}{\lambda} \left[ \nu (1 - A'_1) - (1 + \nu \Gamma)A_2 \right]^2$$

$$\dot{A}_1 = \frac{\sigma^2}{2} A''_1 + \frac{1}{\lambda} \left[ \nu (1 - A'_1) - (1 + \nu \Gamma)A_2 \right] \left[ \nu A'_0 + (1 + \nu \Gamma)A_1 \right]$$

$$\dot{A}_0 = -\frac{1}{2\lambda} \left[ \nu A'_0 + (1 + \nu \Gamma)A_1 \right]^2 + A'_1 \sigma^2 \Gamma + \frac{\sigma^2}{2} \left[ \Gamma^2 A_2 + A''_0 \right].$$

(Here $A'$ and $\dot{A}$ denote the derivative with respect to $p$ and $t$, respectively). Observe that the equation for $\dot{A}_2$ (4.11a) cannot hold in general: $A_2$ does not have a dependence on $p$ but $A_1$ does. The equation will nonetheless hold in the two cases of interest. These equations are subject to the boundary conditions

$$A_2(0) = \gamma E[\Delta P_T^2]$$

$$A_1(0, p) = -\frac{\gamma}{2} E[\Delta P_T \xi | P_T = p]$$

$$A_0(0, p) = \frac{\gamma}{2} E[\xi^2 | P_T = p].$$

The trading strategy is then

$$\theta_t = \frac{\nu Y_t - (1 + \nu \Gamma(t, P_t)) (A_2(T - t)Y_t + A_1(T - t, P_t)) - \nu (A'_1(T - t, P_t)Y_t + A'_0(T - t, P_t))}{\lambda}.$$

Hence, the optimal policy comes from solving for $A_1$ and $A_2$. However, they exhibit closed-form solutions in two important broad cases. Before we move on to these two cases, we first define

$$K = 1 + \nu \Gamma \quad \kappa = \frac{\sqrt{\gamma}}{\lambda} \quad d = \frac{|K|}{\lambda \kappa} \left( \frac{\gamma \sigma^2_T}{K} - \nu \right).$$
4.2.1 Constant Gamma Approximation

On short time scales, we can safely assume that the option has a constant gamma. The approximation says that for small intraperiod fluctuations of $P_t - P_0$, the option’s delta varies linearly with the stock price. We make the the same approximation as in Example 4.1.5

$$g'(t, P_t) \approx g'(t, P_0) + \Gamma(P_t - P_0) \quad \text{with } \Gamma \in \mathbb{R} \text{ constant}.$$ (4.14)

As in Example 4.1.5, the terminal penalty will be

$$E\left[\frac{\gamma}{2} (Y_T \Delta P_T - \xi)^2\right] = \frac{\gamma}{2} E [\sigma_Y^2 Y_T^2 + \xi^2]$$

and does not depend on $P_T$. Furthermore, the dynamics of (4.8) are now independent of $P_t$,

$$dP_t = \nu \theta_t dt + \sigma dW_t$$

$$dY_t = (1 + \nu \Gamma) \theta_t dt + \sigma \Gamma dW_t$$

where $\Gamma$ is now a constant. The assumption considerably simplifies the problem by dropping the dependence on the state variable $P_t$ and eliminating the function $A_1$.

In this subsection, we will make the additional assumption

**Assumption 4.2.1.** Assume that either

1. $d \geq -1$ holds or

2. $d < -1$ and $\kappa |K| T + \text{arctanh}(d) \leq 0$.

**Theorem 4.2.2.** Let Assumption 4.2.1 hold. Then the optimal trading intensity is given by

$$\theta_t = -\kappa h(\kappa K(T - t))Y_t$$ (4.15)

where the function $h$ is defined by

$$h(x) = \begin{cases} 
\tanh(x + \text{arctanh}(d)) & |d| < 1 \\
1 & d = \pm 1 \\
\coth(x + \text{arccoth}(d)) & |d| > 1.
\end{cases}$$ (4.16)
Under the optimal trading strategy, $Y_T \neq 0$ a.s. Indeed, $\mathbb{P}[|Y_T| > M] > 0$ for all $M > 0$. That is, because of the market-impact costs, the position is not perfectly delta hedged, even at the terminal time and there is a chance (albeit small) that $Y_T$ is far away from 0.

Proof. Proof of Theorem 4.2.2 The Martingale Principle of Optimal Control tells us that

$$M_t = \int_0^t \frac{\sigma^2 Y_u^2}{2} - Y_u \nu \theta_u + \frac{\lambda}{2} \theta_u^2 du + J(t, Y_t)$$

is a submartingale for all $\theta_t$ and a martingale under the optimal control. The HJB equation (4.9) simplifies to

$$0 = \frac{\gamma \sigma^2}{2} y^2 + J(t, y) + \frac{(\Gamma\sigma)^2}{2} J''(t, y) - \frac{1}{2\lambda} (y\nu - KJ'(t, y))^2$$

(4.17)

$$\theta(t, y) = \frac{1}{\lambda} (y\nu - KJ'(t, y))$$

(4.18)

where $J'$ and $J''$ are derivatives with respect to $y$. The ansatz (4.10) then simplifies to

$$J(t, y) = A_2(T-t)\frac{y^2}{2} + A_0(T-t).$$

and (4.11a) and (4.11c) simplify to

$$\dot{A}_2 = \gamma \sigma^2 - \frac{1}{\lambda} (\nu - KA_2)^2,$$

$$\dot{A}_0 = \frac{(\sigma\Gamma)^2}{2} A_2.$$

We can see that $\dot{A}_2$ is a parabola opening downward in $A_2$ with vertex at the point $(\frac{\nu}{K}, \gamma \sigma^2)$ and left and right critical points are $A_2^-$ and $A_2^+$, respectively where

$$A_2^\pm = \frac{\nu}{K} \pm \frac{\lambda \kappa}{|K|}.$$

If $A_2(0) = \gamma \sigma^2 \in (A_2^-, A_2^+)$ then $\dot{A}_2(T-t) > 0$ and so $A_2(T-t) \nearrow A_2^+$ as $T-t \to \infty$.

This corresponds to the case when $-1 < d < 1$ where the solution is given by

$$A_2(T-t) = \frac{\nu}{K} + \frac{\lambda \kappa}{|K|} \tanh (\kappa |K|(T-t) + \text{arctanh}(d))$$

$$A_0(T-t) = \frac{(\Gamma\sigma)^2}{2K} \left\{ \nu(T-t) + \frac{\lambda}{K} \left[ \log \circ \cosh(\kappa|K|(T-t) + \text{arctanh}(d)) + \frac{1}{2} \log(1-d^2) \right] \right\}.$$

(4.19)
Similarly, if $A_2(0) = \gamma \sigma_T^2 \in \{A_2^-, A_2^+\}$ then $A_2$ is fixed at the critical point. This corresponds to the case $|d| = 1$ and the solution is given by
\begin{align*}
A_2(T-t) &= \frac{\nu}{K} + \frac{\lambda \kappa d}{|K|} - \sigma^2 \gamma \\
A_0(T-t) &= \frac{\Gamma^2 \sigma^2}{2} \sigma^2 \gamma (T-t).
\end{align*}
(4.20)

Finally, if $A_2(0) > A_2^+$ then $\dot{A}_2(T-t) < 0$ and so $A_2 \searrow A_2^+$. This corresponds to the case $d > 1$ where the solution is given by
\begin{align*}
A_2(T-t) &= \frac{\nu}{K} + \frac{\lambda \kappa}{|K|} \coth (\kappa|K|(T-t) + \text{arccoth}(d)) \\
A_0(T-t) &= \frac{(\Gamma \sigma)^2}{2K} \left\{ \nu(T-t) + \frac{\lambda}{K} [\log \sinh (\kappa|K|(T-t) + \text{arccoth}(d)) + \frac{1}{2} \log(d^2 - 1)] \right\}. \\
(4.21)
\end{align*}

These three cases cover case 1 of Assumption 4.2.1. Under case 2 of Assumption 4.2.1, equation (4.21) also holds and $A_2(T-t)$ is well-defined for $t \in [0,T]$ since $\kappa|K|(T-t) + \text{arccoth}(d) \leq 0$ for all $t \in [0,T]$.

Thus we have a solution $J$ to the HJB equation (4.18) and we see that $M_t$ has non-negative drift for all $\theta$ and is a local martingale for the optimal $\theta$ given in (4.15). To conclude the proof, we need to show that $M_t$ is a submartingale for all $\theta \in \Theta$ and a martingale for the optimal $\theta$. Thus, it suffices to show $\text{E}[M]_T < \infty$ for all $\theta \in \Theta$. Observe that the function $A_2(t)$ is uniformly bounded by some $C > 0$ independent of $t \in [0,T]$ so that
\begin{align*}
\text{E}[M]_T &= \text{E} \int_0^T |J'(t,Y_t)|^2 \, dt \\
&\leq C^2 \text{E} \int_0^T |Y_t|^2 \, dt.
\end{align*}
(4.22)

Itô’s Lemma gives us
\begin{align*}
\text{E}Y_t^2 &= Y_0^2 + \text{E} \left[ \int_0^t 2Y_s dY_s + d[Y]_s \right] \\
&= Y_0^2 + \text{E} \left[ \int_0^t C' \theta_s Y_s + \sigma^2 \Gamma^2 \, ds \right] \\
&\leq Y_0^2 + \text{E} \left[ \int_0^t C'' |Y_s|^2 + \sigma^2 \Gamma^2 \, ds \right]
\end{align*}
for some $C', C'' > 0$ independent of $t \in [0, T]$. Thus, a standard application of Gronwall’s Lemma yields $E[M]_T < \infty$ and we have (4.15) is the optimal policy.

The second part also follows from the fact that $Y_T$ is given by a linear SDE. Its solution

$$Y_t = e^{-\int_0^t \kappa K h(\kappa K(T-s)) ds} \left( Y_0 + \Gamma \sigma \int_0^t e^\int_s^t \kappa K h(\kappa K(T-u)) du dW_s \right)$$

has a density that is non-singular with respect to the Lebesgue measure on $\mathbb{R}$ as $h$ is bounded.

In Almgren and Chriss [2001] $\kappa$ is an ‘urgency parameter’ which dictates the speed of liquidation: the higher $\kappa$ the faster the initial liquidation. In Garleanu and Pedersen [2009], this term gives the relative intensity of trading\(^1\), that is, the higher $\kappa$, the faster the agent trades towards the Merton-optimal portfolio.

The agent’s trading target for shares $X_t$ is the Black-Scholes delta hedge,

$$\text{target}_t = -g'(t, P_t).$$

He constantly trades towards this target but is prevented from holding the exact Black-Scholes delta hedge by the convex temporary impact stemming from limited liquidity. Hence, the trading intensity $\theta_t$ is proportional to the degree of mishedge $Y_t$ and the urgency parameter $\kappa$. There is a greater penalty to being mishedged with higher underlying volatility $\sigma$ and risk aversion $\gamma$ so these parameters increase urgency. Similarly, a more illiquid market (higher $\lambda$) makes trading more costly, which decreases trading intensity.

The solution falls into three cases depending on whether $h(\cdot)$, the trading intensity proportion, is decreasing, flat, or increasing in time, i.e. whether the agent trades less intensely, is flat, or more intensely (respectively) towards the end of the trading period. This, in turn, depends on the value of a constant $d$ which gives the relative size of the overnight jump $\Delta P_T$ versus permanent impact and noise $dP_t$.

The interpretation of $d$ is clearer when there is no permanent impact ($\nu = 0$) and we first handle this case. In this case, the cost of being mishedged comes purely from running

\(^1\)The $a/\lambda$ in Proposition 5 of Garleanu and Pedersen [2009] is equivalent to $\kappa h$ in our setup.
and terminal mishedging terms. Intraperiod trading can be thought of as primarily hedging out the intraperiod fluctuations \( dp_t \) while trading near the end of the period is primarily for hedging the terminal jump \( \Delta P_T \). When \( d < 1 \), the majority of the fluctuations occur during the day, and the solution is marked by \( t \mapsto h(\kappa K(T-t)) \) decreasing as \( t \) approaches \( T \). That is, the agent is more concerned about hedging intraperiod fluctuations \( dp_t \) and relaxes hedging intensity to reduce hedging cost as the end of the period approaches. This is the case in Example 4.1.3 where \( d = 0 \). The second case is when \( d = 1 \) and hedging risk for the intraperiod and end-of-day are weighted equally. Then the agent’s trading intensity is constant in time. The final case is when \( d > 1 \) and the terminal jump is large compared to the daily fluctuations. So the agent trades more intensely with time and \( t \mapsto h(\kappa K(T-t)) \) increases towards \( d \) as \( t \) approaches \( T \). This case could occur in Example 4.1.4.

Remark 4.2.3. If we consider the case when \([0, T]\) is the trading day, the last case (4.21) when \( d > 1 \) is the most realistic. Options market-makers typically increase their hedging towards the close of trading to minimize overnight exposure. The trader may choose a lower \( d \) as options expiry approaches, selecting \( d = 0 \) on the day of options expiry. However, this strategy may be deemed too risky and the trader may choose \( d > 0 \) even at expiry to avoid the risk of being mishedged.

When there is permanent impact \((\nu > 0)\), the running penalty comes from two sources in (4.6), the running mishedging term and the permanent impact. The running mishedging term is as above but the permanent impact term is more subtle. If the trader is in a net long delta position, he sells the position to remain hedged, incurring a capital loss from the downward pressure on his inventory. A buy to hedge out a net short delta position incurs a similar loss and so the permanent impact creates a cost for being mishedged in either direction. The terms for \( A_2 \) in (4.19), (4.20), and (4.21) have an extra term which accounts for the permanent impact. The size of \( d \) balances the running costs (both from running fluctuations and the permanent impact effect) with the terminal cost of being mishedged.
Figure 4.1: Comparative statics of \((KA_2(T - t) - \nu)/\lambda\) as a function of \(t\). Parameters (unless otherwise specified): \(\Gamma = 5.0, \sigma = 0.2, \lambda = 0.2, \nu = 0.2, \gamma = 2.0, \sigma_T = 0.4, T = 1.0\).

<table>
<thead>
<tr>
<th>Units of Fundamental Quantities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_t) sh</td>
</tr>
<tr>
<td>(P) $/sh</td>
</tr>
<tr>
<td>(\theta) sh/sec</td>
</tr>
</tbody>
</table>

Table 4.1: Units for constants: sh is a share, sec is a second, and $ is a dollar.
4.2.2 Comparative Statics

We give some comparative statics for our solution. The effect of the temporary impact term $\lambda$ is the easiest to understand. As the cost of hedging decreases, it becomes optimal to trade more aggressively to minimize mishedging error (see Figure 4.1c). The effect of risk aversion $\gamma$ is similar. As $\gamma$ increases, so does the penalty for being mishedged and there is a stronger incentive to trade more aggressively (see Figure 4.1e). To understand the effect for intraperiod volatility $\sigma$, recall that the agent is trying to hedge both intraperiod volatility $\sigma$ (via intraperiod trading) and volatility from the terminal jump $\sigma_T$ (via trading near the end). As intraperiod volatility $\sigma$ increases while the terminal jump size $\sigma_T$ remains constant, the agent optimally trades more aggressively initially while still ending at the same trading intensity at time $t = T$ (see Figure 4.1b). The effect of increasing terminal jump volatility $\sigma_T$ while keeping intraperiod volatility constant is roughly the opposite (see Figure 4.1f): it increases trading near $t = T$ while trading during the rest of the period remains relatively unaffected. Note that with higher $\sigma_T$, there is a need to trade slightly more intensely even at the beginning of the period $t = 0$ as this impacts the mishedging at the end of the period $t = T$.

$\Gamma$ and the permanent impact $\nu$ have more complex effects, which are qualitatively similar to each other. There are two competing effects at play, whose importance differs with the time to maturity. Firstly, there is an adverse-impact effect stemming from the permanent impact. To see this, let $Y_t < 0$ so the agent optimally wishes to purchase shares. This purchase will push up each share’s value by $\nu$, hence resulting in a decrease of $\nu Y_t < 0$ in the stock value of the agent’s portfolio. In general, rebalancing under permanent impact always hurts his existing position, giving the agent an incentive to trade less aggressively in the hope that subsequent price-fluctuations will automatically rehedge his portfolio. The second effect is dubbed the leverage-effect. With $\nu = 0$, the purchase of a single share increases the trader’s net delta position $Y$ by 1. However, if $\nu > 0$ then his net delta position actually increases by $K = 1 + \nu \Gamma$ due to the permanent impact’s effect on the option’s price $\nu \Gamma$. When $\Gamma > 0$, the effect of a purchase on the portfolio’s net delta
position is leveraged by the permanent impact on the option’s delta. We can think of this as effectively lowering the cost of trading, (i.e. lowering the cost of achieving a fixed change in delta). Hence, the agent optimally increases the intensity of trading. When $\Gamma < 0$, the effect of a sale or purchase is diminished, if not reversed. We can think of this as reducing the effectiveness of trading on the net delta position or increasing the cost of rebalancing. Either way, the agent optimally decreases the intensity of trading.

The relative importance of the two effects can be seen in the expression

$$\theta_t = \frac{1}{\lambda} \cdot \left[ \nu Y_t - K J'(t, Y_t) \right].$$

The leverage-effect acts through $K J'$ while the adverse impact acts through $\nu Y_t$. In Figure 4.1, $d > 1$ and the pressure to be hedged ($K J'$) increases towards the end of the period. Hence, near the beginning of the period, the adverse impact is the overriding concern and trading intensity is less aggressive with higher $\nu$ because of the adverse-impact effect (Figure 4.1d) Near the end, the agent is more concerned about hedging the terminal jump. This combined with the ease of hedging from the leverage effect acting through $K J'(t, Y_t)$ implies the agent trades more aggressively with higher $\nu$ near the end. The effect of $\Gamma$ is similar except that it only directly affects the leverage effect. With higher $\Gamma$, the increase in the leverage effect allows the agent to optimally trade more aggressively, particularly near the end of the day (Figure 4.1a). However, like varying $\sigma_T$, because $\Gamma$ affects mishedging near the closing bell, it has an indirect effect on the trading intensity during the middle of the day as well.

### 4.2.3 General Options, No Permanent Impact

We can relax the constant gamma assumption and allow for general options. However, to make the solution tractable, we need to dispense with permanent impact ($\nu = 0$). We are still able to obtain a fairly explicit solution for the control $\theta$ in (4.12) through $A_2$ and $A_1$. Recall that we have made the assumptions Assumption 4.1.1 on our option.
Theorem 4.2.4. The optimal control is given by

$$\theta_t = -\frac{A_2(T - t)Y_t + A_1(T - t, P_t)}{\lambda}$$ \hspace{1cm} (4.23)$$

with

$$A_2(T - t) = \begin{cases} 
\lambda \kappa \tanh (\kappa(T - t) + \operatorname{arctanh}(d)) & d < 1 \\
\gamma \kappa \sigma_T^2 & d = 1 \\
\lambda \kappa \coth (\kappa(T - t) + \operatorname{arccoth}(d)) & d > 1 
\end{cases}$$

and

$$A_1(T - t, p) = \begin{cases} 
\frac{1}{\sqrt{1 - d^2}} \sech (\kappa(T - t) + \operatorname{arctanh}(d)) \mathbb{E}[F(p + \sigma B_{T-t})] & d < 1 \\
\exp \left( -\frac{1}{\lambda} \gamma \kappa \sigma_T^2 (T - t) \right) \mathbb{E}[F(p + \sigma B_{T-t})] & d = 1 \\
\frac{1}{\sqrt{d^2 - 1}} \csch (\kappa(T - t) + \operatorname{arccoth}(d)) \mathbb{E}[F(p + \sigma B_{T-t})] & d > 1 
\end{cases}$$

where we assume $B$ is a new Brownian Motion, the constants $\kappa$ and $d$ are defined by

$$\kappa = \sigma \sqrt{\frac{\gamma}{\lambda}} \quad d = \frac{\gamma \sigma_T^2}{\lambda \kappa} = \kappa \Delta T$$

and the function $F(\cdot)$ is defined by

$$F(p) = -\frac{\gamma}{2} \mathbb{E} \left[ \Delta P_T \xi \, \bigg| \, P_T = p \right].$$

Proof. Proof In this case, the equations (4.11) then reduce to

$$\dot{A}_2 = \gamma \sigma^2 - \frac{1}{\lambda} A_2^2$$

$$\dot{A}_1 = \frac{\sigma^2}{2} A_1'' - \frac{1}{\lambda} A_2 A_1$$

$$\dot{A}_0 = -\frac{A_1^2}{2\lambda} + \sigma^2 \Gamma A_1' + \frac{\sigma^2}{2} \left[ \Gamma^2 A_2 + A_0'' \right].$$

The solution for $A_2$ then follows immediately. (Observe that the range of tanh includes $(0, 1)$ and the range of coth includes $(1, \infty)$). Feynman Kac then gives us

$$A_1(T - t, p) = \mathbb{E} \left[ \exp \left( -\int_t^T \frac{A_2(T - s)}{\lambda} ds \right) F(p + \sigma B_{T-t}) \right]$$ \hspace{1cm} (4.24)
where $B$ is any Brownian Motion. We have by Assumption 4.1.1 that $g''$ is bounded hence $g'$ is Lipschitz. Since we have

$$F(p) = -\frac{\gamma}{2}E[\Delta P_T \xi \mid P_T = p]$$

$$= \frac{\gamma}{2}E \left[ \sigma^2 \int_T^{T'} g'(u, P_u) - g'(T, P_T) \, du \mid P_T = p \right]$$

we can show that $F$ is Lipschitz as well. Hence $A_1$ given by an expectation in (4.24) is well-defined. Observe that

$$\text{sech}(\text{arctanh}(d)) = \sqrt{1 - d^2} \quad d < 1$$

$$\text{csch}(\text{arccoth}(d)) = \sqrt{d^2 - 1} \quad d > 1$$

from which our expressions for $A_1$ then follow. Finally, since the PDE for $A_0$ is linear, it is straightforward to verify it exists and so we may define $J$ as in (4.10).

For the verification argument, define

$$M_t = \frac{\sigma^2 \gamma}{2} Y^2 u - Y_u \nu \theta_u + \frac{\lambda}{2} \theta^2 u + J(t, P_t, Y_t)$$

Since $J$ solves the HJB equation (4.9), we have that $M_t$ has non-negative drift for all $\theta \in \Theta$ and zero drift for the optimal policy $\theta$ given in (4.23). Again, we shall show that for all $\theta \in \Theta$, the expected quadratic variation $E[M_T] < \infty$. We have

$$E[M_T] = E \int_0^T \left| A_2(T - t) Y_t + A_1(T - t, P_t) \right|^2 dt$$

\[ \leq C'E \int_0^T |1 + Y_t + P_t|^2 dt, \quad (4.25) \]

where $C'$ is a positive constant. Since $A_1$ is bounded and $A_2$ is Lipschitz in $p$, we have for all $\theta \in \Theta$ that

$$E \int_0^T |1 + Y_t + P_t|^2 \leq E|1 + Y_0 + P_0|^2 + 2E \int_0^T (1 + Y_s + P_s) \left[ 1 + \nu (1 + \Gamma(s, P_s)) \right] \theta_s$$

$$+ \sigma^2 (1 + \Gamma(s, P_s))^2 \, ds$$

\[ \leq E|1 + Y_0 + P_0|^2 + CE \int_0^T |1 + Y_s + P_s|^2 \, ds \quad (4.26) \]
by the definition of $\Theta$ and the fact that $\Gamma(t, p)$ is bounded (4.1.1). A standard application of Gronwall’s Lemma yields $E[M]_T < \infty$. 

The optimal control differs from the constant gamma case in that the target is no longer the delta-neutral portfolio. There is a new term $A_1(T - t, P_t)$, which biases the target portfolio to account for the third-derivative of the option price, or the delta of the gamma. If this quantity vanishes, then $F(p)$ vanishes and so does $A_1$. To see this, consider the setup of Example 4.1.4 and observe that by Itô’s Isometry that

$$F(p) = \frac{\gamma}{2} E \left[ \sigma^2 \int_T^{T'} g'(u, P_u) - g'(T, P_T) \, du \mid P_T = p \right].$$

$F$ has an interpretation as the average change in delta of the option over the period $T$ to $T'$. We saw in Example 4.1.5 that $F$ and hence $A_1$ vanish for options with constant gamma.

More generally, if the option has vanishing delta of gamma (that is, $g'''$ vanishes) then for $t > T$, we have by Itô’s Lemma that

$$E [g'(t, P_t) - g'(T, P_T)] = E \left[ \int_T^t g''(u, P_u) \sigma dW_u + \frac{1}{2} g'''(u, P_u) \sigma^2 du \right] = 0.$$

Hence, $F$ and $A_1$ also vanish. For an option with positive delta of gamma ($g''' > 0$), an agent who is delta-neutral at the close $T$ will actually have a positive expected overnight delta exposure $F$. He corrects for this by adding a negative bias to his trading target for shares $X_t$ is

$$\text{target}_t = -g'(t, P_t) - \frac{A_1(T - t, P_t)}{A_2(T - t)}.$$

The effect is the opposite if $g''' < 0$. Observe that for constant $p$, the magnitude of $A_1$ increases as $t \uparrow T$ (see Figure 4.2). This follows since the bias corrects for an ‘overnight’ delta slippage and is less relevant far away from the closing bell.

We plot the offset $A_1$ in Figure 4.2. Compared with Figure 4.2 we omit the plots for $\nu$ (as $\nu = 0$) and $\Gamma$ (as $\Gamma$ can only affect $A_1$ through a nonzero $\nu$). As expected, a higher $\sigma$ reduces the tracking bias $A_1$ as it increases the penalty for not being delta-hedged, although this difference disappears as $t \to T$ (see Figure 4.2a.) Varying $\lambda$ and $\gamma$ have inverse
effects: increasing $\gamma$ is similar to increasing $\sigma$ (increasing the penalty for mishedging) while increasing $\lambda$ has the inverse effect (increasing the market impact costs). The results are plotted in Figure 4.2b and Figure 4.2c. Finally, increasing $\sigma_T$ also increases the running costs of being mishedged (it increases $A_2$ in Figure 4.1f). This in turn decreases $A_1$ (see (4.24) and Figure 4.2d).

Figure 4.2: Comparative statics of $A_1(T-t)$ as a function of $t$. Parameters (unless otherwise specified): $\sigma = 0.2, \lambda = 0.2, \gamma = 2.0, \sigma_T = 0.4, T = 1.0$. We normalize $E[F(p + \sigma B_{T-t})]$ to be 1.

### 4.3 Asymptotics and Applications

#### 4.3.1 Small Market-Impact Cost Limit

We are interested in the limit as both the temporary and permanent impacts vanish. For instance, in the constant gamma case (Theorem 4.2.2) the trading intensity $\theta_t$ is propor-
tional to the distance from being delta hedged \( Y_t = X_t + g'(t, P_t) \) as given in

\[
\theta_t = -\kappa \coth(\kappa K(T-t) + \text{arccoth}(d)) Y_t.
\]

We observe in Figure 4.1d that the profile of trading intensity flattens as a function of time with decreasing \( \nu \) and from Figure 4.1c, we see that the trading intensity shifts upward with decreasing \( \lambda \). Hence, we observe the trading intensity both increasing and flattening as market-impact decreases.

In the limit as \( \lambda \downarrow 0 \) and \( \nu \downarrow 0 \), we have \( K \downarrow 1 \), \( \kappa \uparrow \infty \). If \( \sigma_T > 0 \) then \( d \uparrow \infty \) as well. Therefore \( A_2(T-t) \uparrow \infty \) and \( \theta_t/Y_t \downarrow -\infty \) and the process \( Y_t \) is driven back to zero very quickly. That is, the agent aggressively drives \( X_t \) towards \( g'(t, P_t) \). Hence, we recover the Black-Scholes equation in the limit of vanishing transaction costs. Similarly, for general options and no permanent impact (Theorem 4.2.4), we still have \( A_2(T-t) \uparrow \infty \) and \( A_1(T-t,p) \to 0 \) for all \( p \) and \( t < T \) and so again \( \theta_t/Y_t \downarrow -\infty \). The following theorem makes this idea precise.

**Theorem 4.3.1.** Assume that the assumptions of either Theorem 4.2.2 or Theorem 4.2.4 hold. In the limit of vanishing market-impact costs, we recover the Black-Scholes delta hedge. That is \( \|Y\| \to 0 \) as \( \lambda, \nu \to 0 \) where \( \| \cdot \| \) is the \( L^2(\mathbb{P} \times \mu) \)-norm and \( \mu \) is the Lebesgue measure on \([0,T]\).

**Proof.** If we make the assumptions of Theorem 4.2.2, we have from (4.26) that

\[
E[Y_t^2] \leq (Y_0 + \sigma^2 \Gamma^2 T) \exp \left( -\int_0^t \kappa K \coth(\kappa K(T-s) + \text{arccoth}(d)) \, ds \right).
\]

and the result then follows from noting that \( \kappa \uparrow \infty \), \( K \uparrow 1 \), and \( d \uparrow \infty \) in our limit. In the setup of Theorem 4.2.4, we the result follows similarly by (4.26) and the fact that \( A_2(T-t) \uparrow \infty \) and \( A_1(T-t,p) \to 0 \) in the limit. 

4.3.2 Small Intraperiod Mishedging Penalty Limit

We consider the case when the penalty for the terminal jump is large compared to the intraperiod hedging penalty. Mathematically, this is the limit when \( \sigma \to 0 \) and \( \sigma \Gamma \) and \( \sigma_T \)
are held constant. The objective in this limit is given by

\[
\inf_{\theta \in \Theta} E \left[ \frac{\gamma \sigma^2 T}{2} Y_T^2 - \int_0^T Y_u \nu \theta_u + \frac{\lambda}{2} \theta^2_u \, du \right]
\]

where we have removed the intraperiod mishedging penalty. We include this result because (4.28) clearly demonstrates the increasing nature of the functions \( \theta_t/Y_t \) as \( t \to T \) in terms of algebraic, rather than transcendental, functions.

**Theorem 4.3.2.** We assume that the option has constant gamma as in Theorem 4.2.2. Then for small intraperiod mishedging, the optimal trading intensity \( \theta_t \) can be written as

\[
\theta_t = -\frac{\kappa d}{1 + \kappa K d (T - t)} Y_t.
\]

**Proof.** Proof If we make the Ansatz (4.10) then again \( A_1 = 0 \) and we separate the resulting HJB by powers of \( y \). This yields

\[
\theta(t, y) = \frac{1}{\lambda} (\nu - KA_2(T - t)) y
\]

and

\[
\dot{A}_2 = -\frac{1}{\lambda} [(1 + \nu \Gamma) A_2 - \nu]^2 \quad A_2(0) = \gamma \sigma^2 T
\]

\[
\dot{A}_0 = \frac{\Gamma^2 \sigma^2}{2} A_2 \quad A_0(0) = 0.
\]

By a similar argument as in Theorem 4.2.2, we obtain that

\[
A_2(T - t) = \frac{1}{K} \left[ \nu + \frac{\lambda \kappa d}{1 + \kappa K (T - t)} \right]
\]

\[
A_0(T - t) = \frac{\Gamma^2 \sigma^2}{2K} \left[ \nu(T - t) + \frac{\lambda}{K} \log(1 + \kappa K (T - t)) \right].
\]

The result then follows. \( \Box \)

For a sense of how the execution of the policy in Theorem 4.3.2 compares to the full case in Theorem 4.2.2, see Figure 4.3. Observe that without an intraperiod mishedging penalty, the trading intensity ratio \(-\theta_t/Y_t \) (see (4.28)) is lower than with the intraperiod penalty (see (4.15)), especially during the beginning of the period. The extra trading comes
from hedging out fluctuations during the period. Intuitively, the optimal investor trades less aggressively when only faced with a terminal and no running mishedging penalty. As in the \( d > 1 \) case in Theorem 4.2.2, trading becomes more aggressive towards the end of the period.

If we take the interpretation of \([0, T]\) as the trading day, then this approximation is valid when the overnight jump is large compared to the daily fluctuations. This may be the case in advance of a major post closing-bell announcement.

![Figure 4.3: Comparison of trading intensity with and without intraperiod mishedging penalty. The parameter values are defined as in Figure 4.1](image)

4.3.3 Restriction on the direction of trading

If executed through a broker, the delta-hedging problem is more challenging. Regulatory policy mandates that brokers can only either buy or sell for any given client order. This is to prevent market-manipulation and to protect clients from potentially unscrupulous dealers. We solve the optimal delta-hedging problem under these constraints.

Without loss of generality, we impose a buy-only restriction on trading. That is, we add the constraint \( \theta_t \geq 0 \) a.s. to the minimization of our objective (4.6),

\[
\inf_{\theta \in \Theta^+} \mathbb{E} \left[ \frac{\gamma}{2} (Y_T \Delta P_T - \xi)^2 + \int_0^T \frac{\gamma^2}{2} Y_u^2 - Y_u \nu \theta_u + \frac{\lambda}{2} \theta_u^2 \right],
\]

(4.29)

where \( \Theta^+ = \{ \theta \in \Theta, \theta \geq 0 \} \). While the problem is still Markovian, a closed-form solu-
tion is no-longer readily available. We use a policy-improvement algorithm that assumes a Markovian optimal policy $\theta_t = \theta(t, Y_t)$ to solve this problem. As a check, the same policy-improvement code was used to solve the Simplified Model in Section 4.3.2 and compared against the analytic solution found in Section 4.3.2. It obtained continuation-value functions $J$ accurate to within $\sim .1\%$.

A comparison of the restricted and unrestricted cases are plotted in Figure 4.4. Observe the asymmetry in trading policy $\theta(t, y)$ in Figure 4.4c. For negative values of $y$, $\theta(t, y)$ is positive to reduce the delta-hedging error, that is the agent still purchases stock as in the unrestricted case (compare with Figure 4.4a). For positive values of $y$, an unrestricted agent would sell shares but a restricted agent cannot do so, and the binding constraint forces $\theta$ to be zero (see Figure 4.4c). There is a no-trade curve $y(t) \leq 0$ such that the agent trades if and only if his current net delta position is below this level, i.e. $\theta_t = 0$ when $Y_t \geq y(t)$ and $\theta_t > 0$ when $Y_t < y(t)$. This curve is marked in the plot in Figure 4.4c. Observe that $y(0) < 0$ and $t \mapsto y(t)$ is an increasing function. Hence, far from the terminal time ($t \ll T$), the agent does not purchase stocks even when he is net short delta. The result is interpreted thus: a selling-restricted agent is hesitant to purchase stock only to see his delta position become positive before $T$ due to stock-price fluctuations. However, $y(T) = 0$ so this effect disappears as $t \to T$: as the time remaining for $Y_t$ to fluctuate above 0 runs out, the agent’s rule becomes “buy if I am net short delta.”

In the restricted case, the asymmetry in execution strategy for $\theta$ translates into an asymmetry for the value function $J$ (compare the restricted case Figure 4.4d with the unrestricted one Figure 4.4b). For negative values of $Y_t$, $J$ behaves similarly in both the selling-restricted and non-restricted cases as the optimal strategies are similar. For positive values of $Y_t$, $J$ is significantly higher in the selling-restricted case as the control cannot be exercised to reduce the hedging error. The difference between the two cases represents the premium of being able to sell.
Figure 4.4: Plots of the continuation-value function $J$ and the optimal policy $\theta$ for the unrestricted case (Figure 4.4b and Figure 4.4a) and the case when trading is restricted to purchases $\theta \geq 0$ (Figure 4.4d and Figure 4.4c). In the restricted case, we also plot the no-trade boundary in black, above which $\theta(t, y) = 0$ (see Figure 4.4c). Parameters: $\Gamma = 5, \sigma = .4, \lambda = .2, \nu = .2, \gamma = 2, \sigma_T = .4, T = 1$.

4.3.4 Stock Pinning

The increase in trading intensity for high gamma near the expiry is the mechanism behind stock pinning. A put or call option that is about to expire with a large outstanding institutional interest and with a stock price near a strike level can induce a so-called ‘pinning effect’ whereby the stock price at the close of trading on expiry is ‘pinned’ to the strike level. Empirically, this manifests itself in the observation that the distribution of end-of-day stock prices on these days deviates substantially from the distribution of end-of-day stock price on any other days. That is, the stock price clumps around the strike
level at the close of such an option expiry [Avellaneda and Lipkin, 2003].

A call or put option near expiry with the underlying prices near its strike exhibits a large positive gamma. When institutional investors are net short gamma (perhaps because they collectively short the put or call) option market makers are net long gamma. The market-makers will hedge their position by buying stock as the price dips and selling it as the price increases, thus stabilizing the price. Then as Figure 4.1a shows, a large gamma implies the agent trades intensely as expiry approaches and as his trades are stabilizing, the permanent impact pins the stock price at the strike. Since the gamma peaks near expiry with the underlying price near the strike, the effect manifests itself as pinning near the strike.

Our model provides some insight into this model. Observe that for a call option, the delta is given by
\[
C'(t,p) = \mathbb{E}[1_{P_T > 0} \mid P_t = p] = N\left(\frac{p}{\sigma\sqrt{T-t}}\right)
\]
and so the gamma is given by
\[
C''(t,p) = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \exp\left(-\frac{p^2}{2\sigma^2(T-t)}\right).
\]
(4.30)
For small fluctuations of \(p\), the gamma appears to be flat. Let \(g = NC\) where \(N\) is a large positive number representing the number of call options held. We make the constant gamma assumptions of Theorem 4.2.2, where now the gamma will be \(\Gamma \approx NC''(t,p)\) for the gamma of the entire position. It turns out that this assumption is valid for large \(N\) and moderately large values of \(T - t\). For simplicity, we take \(d = 1\) so that \(\theta_t = \kappa Y_t\) and we rewrite (4.7) and (4.8) as
\[
dP_t = \frac{1}{\Gamma} (dY_t - \theta_t \, dt)
\]
\[
dY_t = -\kappa(1 + \nu\Gamma)Y_t \, dt + \sigma\Gamma dW_t.
\]
Hence, we have that the stationary distribution of \(Y_t\) is
\[
Y_t \sim \mathcal{N}\left(0, \frac{(\sigma\Gamma)^2}{2\kappa(1 + \nu\Gamma)}\right).
\]
That is, both $\theta_t$ and $Y_t$ are distributed with standard deviation proportional to $\sqrt{N}$ in the limit as $N \to \infty$. Hence, we have that the standard deviation of $P_T \sim 1/\sqrt{N}$ as $N \to \infty$. In other words, if $P_0$ starts close to the strike then $P_t$ will remain not deviate far and the price is pinned to the strike. We can see from (4.30) that near the strike ($p = 0$) the gamma of the call option is flat in moneyness. Hence, our constant gamma approximation is valid in the limit as $N \to \infty$ as the pinned stock only “sees” a flat gamma surface.

4.3.5 Intraday Trading Patterns

If we continue with the interpretation of the trading period as the trading day, we see that it is consistent with the intraday trading pattern. The typical daily profile is $U$-shaped, with the most intense trading occurring during the opening and close (see Figure 4.5a). Compare Figure 4.5a with the graph of the average trading volume from delta hedging in Figure 4.5b as predicted by our model,

$$E|\theta_t|$$

with $Y_0 \sim N(0, \Sigma_0)$. The distributional assumption on $Y_0$ comes from the overnight fluctuations in the underlying affecting the delta of the option. The expression for $E|\theta_t|$ is given in (4.31). The parameters for the plot are chosen to reflect realistic values and are explained in Example 4.3.3.

We see that our model predicts a trading rate that is roughly 20% of the trading for (NYSE:BA). More importantly, the trading volume profile roughly corresponds to the double-humped trading pattern that is observed in the empirical data. Near the open, the high $E|\theta_t|$ comes from the high initial variance in $Y_t$ as the trader trades down his net delta position from the previous evening’s jump. Near the close, the high $E|\theta_t|$ comes from the increase in $h$ near $t = T$, which represents aggressive trading in anticipation of the coming evening’s overnight jump. Clearly the intraday trading volume is determined by many additional factors. However, on stocks with a large outstanding interest, Figure 4.5 illustrates how delta hedging may account for a significant fraction of the trading.
Example 4.3.3. We choose $\sigma = .04$ based on the corresponding value for Boeing (NYSE:BA) in Figure 4.2 (see the chart for units). We set $\Sigma_0 = \Gamma \cdot (1.5 \cdot \sigma_T)$ with $\sigma_T$ from Figure 4.2. This corresponds to the change in the option’s delta from an overnight jump in the stock of size $1.5 \cdot \sigma_T$. If the book were perfectly hedged the night before, we would see a jump of size $\Gamma \sigma_T$ in the option value. The factor 1.5 accounts for the book not being perfectly hedged at the close of the previous evening. We set $T = 6.5$ hours, the length of the trading day.

$\lambda = 10^{-4}$ corresponds to a temporary impact of 10 cents for trading 20% of ADV or 32 shares per second. We choose the permanent impact $\nu = 10^{-5}$ corresponds to a permanent impact of 1 cent for trading 32 shares. For the computation of $d$, we choose $\sigma_T = 25$ to penalize not just for the unhedgable overnight fluctuation but for the hedging costs to be incurred for future days starting the next morning. This penalty accounts for We choose $\Gamma = 10^5$ to corresponds to the gamma position of the entire book. In the expression for the gamma of an option (4.30), $\Gamma(3.5 \text{ hours}, 0) \approx .1$ is the gamma of an At-The-Money option on Boeing (NYSE:BA) about 3 hours from expiry. Hence, for nearest month options we expect the gamma of an individual option gamma to be between .01 and .1 (compare with Figure 4.2). Therefore, our assumption on the net book $\Gamma$ corresponds to a position with one to ten million At-The-Money contracts. Finally, we choose $\gamma \sim 10^8$. This can be thought of as corresponding to a relative risk aversion of 1 for an agent with a portfolio of a million dollars.

We now give the derivation of $E|\theta_t|$ based on the assumptions of Theorem 4.2.2. It is clear that $Y_t$ is normally distributed with mean zero. From (4.8) and (4.15), we obtain a PDE for the variance of $Y_t$ as

$$\frac{d}{dt}EY_t^2 = -2\kappa Kh(\kappa K(T - t))EY_t^2 + (\Gamma \sigma)^2.$$  

We are interested in the more realistic case when $d > 1$ (see Remark 4.2.3) whose solution is given by

$$EY_t^2 = e^{-2H(t)}(\Gamma \sigma_T)^2 + (\Gamma \sigma)^2 e^{-2H(t)} \int_0^t e^{2H(s)} \, ds.$$
where

\[ H(t) = kK \int_0^t \coth(kK(T - s) + \text{arccoth}(d)) \, ds \]

\[ = - \log \left[ \sinh(kK(T - t) + \text{arccoth}(d)) \csc(hKT + \text{arccoth}(d)) \right]. \]

Since we have

\[ \int_0^t e^{2H(s)} \, ds = \sinh^2(kKT + \text{arccoth}(d)) \int_0^t \csc^2(\kappa K(T - s) + \text{arccoth}(d)) \, ds \]

\[ = \frac{\sinh^2(kKT + \text{arccoth}(d))}{\kappa K} \left[ \coth(\kappa K(T - t) + \text{arccoth}(d)) - \coth(\kappa KT + \text{arccoth}(d)) \right] \]

we can compute the variance of \( Y_t \) as

\[ \text{E}Y_t^2 = \frac{\sinh^2(kK(T - t) + \text{arccoth}(d))}{\sinh^2(kKT + \text{arccoth}(d))} \cdot \left\{ (\Gamma \sigma_T)^2 + \frac{(\Gamma \sigma_T)^2}{\kappa K} \sinh^2(kKT + \text{arccoth}(d)) \right. \]

\[ \left. \quad \left[ \coth(\kappa K(T - t) + \text{arccoth}(d)) - \coth(\kappa KT + \text{arccoth}(d)) \right] \right\}. \]

Since \( \theta_t \) is normally distributed, we have that

\[ \text{E} [\theta_t] = \sqrt{\frac{2}{\pi \kappa \coth(\kappa K(T - t) + \text{arccoth}(d))}} \sqrt{\text{E}Y_t^2}. \] (4.31)

This is plotted in Figure 4.5b with parameters chosen as in Example 4.3.3.

### 4.4 Discrete-Time Formulation and Solution

In this section, we assume a constant \( \Gamma \) model as in Theorem 4.2.2. We need a discrete-time formulation and solution to the hedging problem in order to perform the simulations on TAQ data in Section 4.5. The simplest discrete strategy would be to evaluate the continuous-time strategy at each time point in question, buying \( \theta_t \Delta t \) shares over the interval \([t, t + \Delta t]\). In effect, this is a forward Euler discretization of the underlying dynamics. Unfortunately, this strategy results in unrealistic and costly overshooting.
To illustrate this, consider the mean value of $Y_t$, $\mathbb{E}[Y_t]$. Under the continuous-time policy of Theorem 4.2.2, it would behave as

$$
\bar{Y}_{\Delta t} = Y_0 - \int_0^{\Delta t} \kappa K h(\kappa K(T - s)) \bar{Y}_s \, ds
$$

and converges exponentially towards 0 but remains positive if $Y_0 > 0$. On the other hand, under a discrete Euler policy where we purchase $\theta_0 \Delta t$ shares, $\bar{Y}_t$ would behave as

$$
\bar{Y}_{\Delta t} = Y_0 (1 - \kappa K h(\kappa KT) \Delta t)
$$

Depending on the size of $\Delta t$, this may not be positive for fixed $Y_0 > 0$. So while this strategy converges to the continuous strategy as $\Delta t \to 0$, for fixed $\Delta t$, it may overshoot zero, a result which is known in the literature (see, for example, Ascher and Petzold [1998]). Indeed, simulations using the discretized Euler-method exhibited this overhedging. In order to obtain well-behaved discrete time solutions we must pose the discrete-time problem directly. The overshooting is so severe as to lead to unstable solutions! It is especially prominent when temporary impact $\lambda$ is small compared to $K^2 A_2(T - t) \Delta$ (see the denominator of discrete-time solution (4.33)). The continuous-time solution is important as it is closed-form and provides important intuition for understanding the problem dynamics. The discrete-time solution resembles the discretized continuous solution but makes an important correction for overshooting (4.33) that is important for practical applications.

### 4.4.1 Discrete-Time Formulation

We will discretize the dynamics by imposing a lattice $\mathcal{T}_N$ of $N$ points of width $\Delta t = T/N$. That is $\mathcal{T}_N = \{\frac{T}{N}(\mathbb{Z} \cap [0,N))\}$. At each time $t \in \mathcal{T}_N$, the agent sets his strategy for the entire timestep $[t, t + \Delta t)$. Under such conditions, it is only necessary to consider the implied discrete-time process. Hence, the agent’s stock position is given by

$$
X_t = X_0 + \sum_{u \in \mathcal{T}_N, u < t} \theta_u
$$

while the dynamics for the stock price becomes

$$
P_t = P_0 + \nu X_t + \sum_{u \in \mathcal{T}_N, u < t} \sigma \Delta W_u
$$
where $\Delta W_{kT/N} = W_{(k+1)T/N} - W_{kT/N}$. Then the new discrete-time value function becomes

$$J(t, y) = \inf_\theta E \left[ \sum_{u \in T_N, u \geq t} \left( \frac{\gamma \sigma^2}{2} Y_u^2 - Y_u \nu \theta_u + \frac{\lambda}{2} \theta_u^2 \right) \Delta t + \frac{\gamma \sigma^2}{2} Y_T^2 \bigg| Y_t = y \right].$$

### 4.4.2 Problem Solution

Again, we make a quadratic Ansatz that for $T - t \in T_N$,

$$J(T - t, y) = A_2(T - t + \Delta t) \frac{y^2}{2} + A_0(T - t + \Delta t). \quad (4.32)$$

**Theorem 4.4.1.** The discrete-time optimal trading intensity $\theta = \theta(t, Y_t)$ is given by

$$\theta(t, y) = \frac{\nu - KA_2(T - t)}{\lambda + KA(T - t) \Delta t} y$$

and the continuation value $J$ is given by (4.32) where

$$A_2(T - t + \Delta t) = A_2(T - t) + \gamma \sigma^2 \Delta t + \frac{(\nu - KA_2(T - t))^2}{\lambda + KA_2(T - t) \Delta t} \Delta t \quad A_2(0) = \gamma \sigma^2$$

$$A_0(T - t + \Delta t) = A_0(T - t) + \frac{1}{2} \Gamma^2 \sigma^2 A_2(T - t) \Delta t \quad A_0(0) = 0.$$  

**Proof.** Proof. The discrete-time HJB equation becomes

$$J(T - t, y) = \inf_{\theta_t} \left[ \frac{\gamma \sigma^2}{2} y^2 - y \nu \theta_t + \frac{\lambda}{2} \theta_t^2 \right] \Delta t + \frac{A_2(T - t)}{2} \left[ (y + K \theta_t \Delta t)^2 + (\Gamma \sigma)^2 \Delta t \right] + A_0(T - t) = 0.$$

Separating both sides by powers of $y$ and using (4.32) yields the desired result. \hfill \Box

### 4.5 Simulation Using TAQ Data

To test the hedging strategy, we use stock TAQ Data to simulate delta hedging a call option using our trading strategy and compare that to the benchmark Black-Scholes strategy. We choose five companies: Boeing (NYSE:BA), Bank of America (NYSE:BAC), Microsoft (NASDAQ:MSFT), Pfizer (NYSE:PFE), and Wal-Mart (NYSE:WMT), each representing a different sector of the economy. Daily NBBO data was collected for the 251 trading days from April 3rd, 2000 to March 30th, 2001 (inclusive) and the mid-price prevailing at
Table 4.2: Summary Statistics of Stocks: the average closing stock price $P_T$, absolute volatility in dollars per share per square-root seconds $\sigma$, the average overnight volatility in dollars per share $\sigma_T$, and the average $\Gamma$ of a call-option that matures in 5 days and was At-The-Money at the previous close.

<table>
<thead>
<tr>
<th>Stock</th>
<th>$P_T$</th>
<th>$\sigma$</th>
<th>$\sigma_T$</th>
<th>$\Gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>53.27084</td>
<td>0.0433451</td>
<td>0.8270762</td>
<td>0.0515023</td>
</tr>
<tr>
<td>BAC</td>
<td>49.5763872</td>
<td>0.0243846</td>
<td>0.7866773</td>
<td>0.0780991</td>
</tr>
<tr>
<td>MSFT</td>
<td>64.9043047</td>
<td>0.0122172</td>
<td>1.4496147</td>
<td>0.1173921</td>
</tr>
<tr>
<td>PFE</td>
<td>43.5416817</td>
<td>0.0164636</td>
<td>0.6695676</td>
<td>0.1008441</td>
</tr>
<tr>
<td>WMT</td>
<td>52.5306878</td>
<td>0.0208141</td>
<td>0.9813519</td>
<td>0.0773531</td>
</tr>
</tbody>
</table>

We use the price data to calibrate our the discrete-time model (see Section 4.4) under minutely rehedging. We then use the data to simulate delta hedging a derivatives’ position with a flat gamma. We chose $\Gamma$ to corresponded to that of one million call options that were At-The-Money at the previous trading day’s close and mature 5 days after that day’s close. These quantities are computed using the Black-Scholes formula.

We can think of the flat gamma derivative as approximating that of an At-The-Money call option under Approximation (4.14). A call-option’s gamma profile becomes very narrow in moneyness near the option’s expiry so that small fluctuations in the stock price will greatly affect the gamma, thus violating the assumptions of the model. This feature is not unique to our setup but stems from the ‘kink’ in the ‘hockey-stick shape’ payoff of the call and the associated difficulty of hedging At-The-Money call options near expiry is a widely recognized problem among practitioners. We sidestep this issue in our simulation by choosing a call option that expires in 5 days.

We set the initial position $X_0$ so that the option is initially delta hedged at the previous close. This (optimistically) simulates the position of a trader who hedges an option across multiple days. His initial position in the morning may not be hedged due to the overnight fluctuation.

We follow the logic of Example 4.3.3 in choosing our parameter values. We choose
\[ \gamma \sim 10^{-6}. \] We assume no permanent impact \( \nu = 0 \) and a temporary impact of the form \( \lambda = \lambda_p \sigma \) where \( \lambda_p > 0 \) is a proportionality constant and \( \sigma \) is the absolute volatility of the stock. This accounts for the well-known stylized fact that, caeteris paribus, market impact is higher for stocks with greater volatility [Ané and Geman, 2000, Cont, 2001, Jain and Joh, 1988]. We choose \( \lambda_p = 10^{-2} \) so that \( \lambda \approx 10^{-4} \) as in Example 4.3.3. The parameter \( \sigma \) was chosen based on empirical values while \( \sigma_T \) was chosen to be 10 times its empirical value to induce the agent to hedge aggressively towards the close. Given our choice of parameter values, the characteristic time-scale at which trading intensity peaks near the market close is

\[
\frac{1}{\kappa} = \sqrt{\frac{\lambda_p}{\gamma} \frac{1}{\sigma}}
\]

which is on the order of 10 to 20 minutes. In other words, for our liquid stocks, the agent trades with a constant proportional intensity until the last hours of the trading day.

We track the performance of two different agents, one using the proposed hedging intensity of Section 4.4 (Model) and another whose strategy is to maintain a Black-Scholes delta hedged (BS). The Black-Scholes trader will not be perfectly hedged due to fluctuations in the subsequent minute interval before rehedging. However, (BS) will likely maintain a tighter hedge on average as he would be perfectly hedged if the stock did not fluctuate in the subsequent interval whereas the other trader (Model) is not even hedged at the start of the interval to begin with.

The dollar terminal exposure, dollar running exposure, and dollar impact costs are given in Figure 4.3. The formulas for these quantities are given by

\[
\text{terminal exposure} = |\sigma_T Y_T| \\
\text{intraday exposure} = \sqrt{\sum_{t \in T} \sigma^2 Y_t^2 \Delta t} \\
\text{impact cost} = \sum_{t \in T} \lambda \theta_t^2 \Delta t.
\] (4.34)

Each term has dimension of dollars to make them directly comparable. Hedging according to the model in Section 4.4, (Model) saves $10K a day in market impact costs, while taking
on a comparable amount of risk.

(b) Intensity of trading $E|\theta_t|$ for one trading day with parameters given in Example 4.3.3. In our example, the trading predicted by our model accounts for 20% of trading flow.

Figure 4.5: Intraday trading intensity, actual and projected in our model from delta hedging.
Black-Scholes

<table>
<thead>
<tr>
<th></th>
<th>Terminal</th>
<th>Intraday</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>9.965e+02</td>
<td>9.661e+03</td>
<td>1.235e+04</td>
</tr>
<tr>
<td>BAC</td>
<td>1.283e+03</td>
<td>9.300e+03</td>
<td>1.865e+04</td>
</tr>
<tr>
<td>MSFT</td>
<td>5.758e+03</td>
<td>1.387e+04</td>
<td>6.022e+04</td>
</tr>
<tr>
<td>PFE</td>
<td>1.736e+03</td>
<td>7.984e+03</td>
<td>1.583e+04</td>
</tr>
<tr>
<td>WMT</td>
<td>2.142e+03</td>
<td>1.114e+04</td>
<td>2.564e+04</td>
</tr>
</tbody>
</table>

Model, $\gamma = 0.5 \times 10^{-6}$

<table>
<thead>
<tr>
<th></th>
<th>Terminal</th>
<th>Intraday</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>1.086e+03</td>
<td>2.751e+04</td>
<td>5.036e+02</td>
</tr>
<tr>
<td>BAC</td>
<td>1.392e+03</td>
<td>3.095e+04</td>
<td>6.780e+02</td>
</tr>
<tr>
<td>MSFT</td>
<td>5.858e+03</td>
<td>4.969e+04</td>
<td>2.019e+03</td>
</tr>
<tr>
<td>PFE</td>
<td>1.838e+03</td>
<td>2.572e+04</td>
<td>5.163e+02</td>
</tr>
<tr>
<td>WMT</td>
<td>2.242e+03</td>
<td>3.641e+04</td>
<td>9.844e+02</td>
</tr>
</tbody>
</table>

Model $\gamma = 1.0 \times 10^{-6}$

<table>
<thead>
<tr>
<th></th>
<th>Terminal</th>
<th>Intraday</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>1.043e+03</td>
<td>2.359e+04</td>
<td>6.779e+02</td>
</tr>
<tr>
<td>BAC</td>
<td>1.339e+03</td>
<td>2.641e+04</td>
<td>8.921e+02</td>
</tr>
<tr>
<td>MSFT</td>
<td>5.808e+03</td>
<td>4.267e+04</td>
<td>2.585e+03</td>
</tr>
<tr>
<td>PFE</td>
<td>1.789e+03</td>
<td>2.205e+04</td>
<td>6.686e+02</td>
</tr>
<tr>
<td>WMT</td>
<td>2.193e+03</td>
<td>3.116e+04</td>
<td>1.281e+03</td>
</tr>
</tbody>
</table>

Model, $\gamma = 2.0 \times 10^{-6}$

<table>
<thead>
<tr>
<th></th>
<th>Terminal</th>
<th>Intraday</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>1.020e+03</td>
<td>2.034e+04</td>
<td>9.253e+02</td>
</tr>
<tr>
<td>BAC</td>
<td>1.311e+03</td>
<td>2.258e+04</td>
<td>1.192e+03</td>
</tr>
<tr>
<td>MSFT</td>
<td>5.783e+03</td>
<td>3.667e+04</td>
<td>3.401e+03</td>
</tr>
<tr>
<td>PFE</td>
<td>1.763e+03</td>
<td>1.894e+04</td>
<td>8.820e+02</td>
</tr>
<tr>
<td>WMT</td>
<td>2.167e+03</td>
<td>2.672e+04</td>
<td>1.700e+03</td>
</tr>
</tbody>
</table>

Table 4.3: Summary Statistics of Simulation

Summary Statistics of Simulation: Average dollar exposure versus dollar market impact by stock for a simulated Black-Scholes trader and one that follows our Model. See (4.34) for details of the terms.
Appendix A

Martingale Principle of Optimal Control

This is a simplified exposition of the Martingale Principal of Optimal Control used in Chapter 4. We use the notation of Section 4.2.

Theorem A.0.1 (Martingale Principal of Optimal Control). Assume that

\[ M_t = \int_0^t \frac{\sigma^2 \gamma}{2} Y_u^2 - Y_u \nu \theta_u + \frac{\lambda}{2} \theta_u^2 du + J(t, P_t, Y_t) \]

is a submartingale for all \( \theta \in \Theta \) and a martingale for \( \theta^* \in \Theta \). Then \( \theta^* \) is the optimal policy.

Hence, if we are able to find a solution \( J \) to (4.9), then we would have that for an arbitrary \( \theta_t \), the drift of \( M_t \) would be non-negative and it would be zero for \( \theta_t = \theta^*_t \). A solution \( (A_2, A_1, A_0) \) to the equations (4.11) would yield the desired solution \( J \) to (4.9). These solutions are found in Theorem 4.2.2 and Theorem 4.2.4 for the desired subcases. Hence, for both theorems we have that \( M_t \) is a local martingale under the optimal policy. It remains to show that it is a strict martingale under the optimal policy \( \theta^* \). This follows from showing that the expectation of the quadratic variation is finite,

\[ \mathbb{E}[|M_t|] < \infty \]
which is given in (4.22) and (4.25) in Theorem 4.2.2 and Theorem 4.2.4, respectively.
References


T. Li. Optimal limit-versus-market order slicing under a vwap bechmark - continuous case, 2012.


M. Mnif and A. Sulem. Optimal risk control and dividend policies under excess of loss reinsurance. *Stochastics An International Journal of Probability and


