The University of Warwick

Warwick Business School

Masters in Financial Mathematics

Optimal Portfolio Liquidation with Reinforcement Learning Applications

Supervisor:

Dr Florian Theil

Author:

Robert Buck

Academic year 2020/2021

Acknowledgments

Firstly, I would like to thank my dissertation supervisor Florian Theil who was incredibly helpful with respect to ensuring that the mathematics throughout this dissertation was rigorous. Secondly, my family who have supported me though out my academic journey, without whom, I would not have achieved a fraction of what I have. Finally, to my loving girlfriend who has put up with a library obsessed boyfriend for the past year. This dissertation is as much all of your achievement as it is mine.

Abstract

The aim of this dissertation is to examine the optimal liquidation policy for an agent who seeks to maximise their terminal wealth while minimising deviations from a target liquidation schedule. This model is constructed in the context of an agent who must liquidate a fixed inventory over a finite time horizon using both market and limit orders. I initially formulate this problem in an optimal control and stopping context. The corresponding quasi variational inequality (QVI) is then derived for the given value function. This is solved via a finite difference scheme giving the optimum depths to place limit orders and times to execute market orders. The results are then discussed and the impact when the units of inventory to liquidate approaches infinity is discussed. The assumptions that enable the QVI to be solved are then relaxed. As such a reinforcement learning technique is then implemented to find an optimal liquidation policy. Finally, the liquidation policy given by solving the QVI and that produced by the reinforcement learning method are compared.

Contents

1 Introduction

Modern financial markets are increasingly dominated by algorithmic trading strategies. A study by Bigiotti, Alessandro; Navarra, Alfredo [1] in 2016 showed that over 80% of volume in forex markets was performed by algorithms rather than humans. Indeed, algorithmic trading is becoming increasingly popular with not only institutional investors but also retail clients with websites such as Quantopian [2] making it increasingly easy for investors to learn about these strategies. Algorithmic strategies follow pre-defined rules and allowing for consistency and speed of execution far greater than anything that could be achieved manually [3]. A clear application of these algorithmic strategies is with respect to the classic problem in finance of optimal portfolio liquidation. In this problem an agent has a portfolio of financial assets and is looking to sell these assets in an optimal way. The success of the liquidation schedule chosen is then evaluated according to certain performance criteria, e.g. the price sold at, how quickly the inventory was sold e.c.t. In financial markets liquidity providers are compensated via the bid-ask spread. Therefore, any market agent seeking to liquidate a position faces a trade off between immediacy and cost represented via the choice of either submitting a limit or market order. As such the agent's optimal liquidation policy will depend on their sensitivity to market risk and execution costs in addition to the dynamics of the market. However, finding this optimal liquidation schedule is often not a trivial task.

Typically the problem of optimal liquidation is formulated as either an optimal control problem or a combination of optimal control and optimal stopping. This distinction is typically determined by how the liquidation problem is formulated, e.g. if only limit orders are allowed or if market orders can also be used. A solution can then be found by either applying the dynamic programming principle or by solving the corresponding Hamiltonian Jacobi Bellman equation (HJB)/quasi variational inequality (QVI) depending on the context, see [7] and [8]. Indeed, solving the corresponding QVI is one of the approach taken in

this dissertation.

The key tasks that this dissertation seeks to tackle are as follows:

Formulate the market dynamics and performance criteria and solve the corresponding QVI in order to derive the optimal liquidation policy given by stochastic optimal control and stopping, see section 5.

Evaluate what occurs to the QVI as the units of inventory to liquidate approaches infinity, see section 5.5.

Relaxing the assumptions that allow the QVI to be solved and attempting to still find an optimal liquidation policy, section 6 tackles this problem. With section 6.2 detailing the distinctions between how the optimal liquidation problem is formulated and solved here vs in section 5. Finally, section 6.3 compares the solution given by this algorithm to that given by solving the QVI.

A complete overview of all chapters is detailed below:

In Chapter 3 I outline the mathematical definitions, propositions and theorems used throughout this dissertation. An emphasis is placed her on jump processes due to their usefulness in modeling the arrival of market orders.

In Chapter 4 I will then derive the Hamiltonian Jacobi Bellman equation for the general case of optimal control before moving to derive the QVI for the case of optimal control and stopping. This is done in order to show the reader how solving the QVI will ensure that the optimal policy is found.

Chapter 5 starts by specifying the model used to formulate the optimal portfolio liquidation problem using both limit and market order. The corresponding quasi variational inequality for this optimal control and stopping problem is then derived using the theorems discussed

in Chapter 4. A numerical solution to the quasi variational inequality is found under a given parametisation and the results discussed. Finally, the limiting properties of the QVI as the units of inventory approach infinity is discussed.

Chapter 6 looks to relax some of the assumptions underlying the results seen in Chapter 5. In order to do this a variation of a random search algorithm commonly used in the reinforcement learning is implemented. The optimal combination of market and limit orders given by this method are then compared to those produced by solving problem using the method in Chapter 5.

Chapter 7 gives the concluding remarks, summing up the results from the dissertation and giving a brief discussion on further research that could be done.

2 Literature Review

Optimally entering and liquidating financial positions is a heavily studied area of quantitative finance. As such there have been many proposed solutions to the problem under varying constraints and market dynamics.

Possibly the most iconic liquidation schedule within quantative finance is that given by Almgren & Chriss [4]. In this problem the agent can only post market orders but can control the speed at which they are posted. However, the market orders placed by the agent impact the asset price both temporarily and permanently. The agent will then seek to find a liquidation policy which maximises total returns for a given level of risk aversion, as represented by the variance of returns. As such an efficient frontier for optimal execution can be formed and from this a closed form solution for optimal execution times is found. This liquidation schedule will often be used as a benchmark to evaluate alternative liquidation schedules.

Although Almgren & Chriss find an optimal liquidation schedule by implementing a similar method used in Markowitz portfolio optimisation [5], a more typical approach taken in the literature is to formulate the market dynamics in such a way that the Hamiltonian Jacobi Bellman (HJB) equation exists. The corresponding HJB equation or quasi variation inequality (QVI) can then be solved either analytically or numerically to determine the optimal policy (actions to take by the agent). Indeed, this is the approach taken by Huitema [7]. Here the market dynamics are formulated such that the order pressure from the agent has both a temporary and permanent impact on the asset price. The agent then seeks to maximise their utility from the sale of theses assets, where the utility function is given by a CARRA utility function. The agent will attempt to maximise their utility function by posting an optimal combination of limit and market orders. Under these market dynamics and possible actions Huitema formulates the corresponding HJB equation. This is then solved

numerically via a finite difference scheme in order to determine the optimal policy.

Cartea and Jaimungal [8] similarly tackle the problem of an optimal limit & market order policy via constructing market dynamics such that the HJB equation exists. The key difference being the formulation of the problem as that of optimal stopping and control. As such they solve the QVI rather than just the HJB. However a key distinction between Cartea and Jaimungal method rather than that of Huitema is the use of a deterministic target liquidation schedule. Specifically, in the case of Cartea and Jaimungal the agent seeks to maximise their wealth while minimising the deviations of their inventory from said target schedule. Added complexities can also be introduced by allowing there to be more than one venue in which to liquidated the asset see "Generalized Optimal Liquidation Problems Across Multiple Trading Venues" [9] and "Optimal order placement in limit order markets" [10]. Furthermore, the complexities caused by multiple trading venues can be compounded should one of these venues be a dark pool. Where a dark pool can be thought of as a private trading venue with often non transparent order follow, see Lin [11] for further details. The impact of these dark pools can be seen primarily via problems associated with adverse selection, see "Simultaneous Trading in 'Lit' and Dark Pools" [12] and "Optimal liquidation in dark pools" [13].

In previous discussed scenarios the agent does not have a directional view on the market. However, an opinion on future asset returns can be incorporated into an optimal trading policy. Cartea, Jaimungal, Kinzebulatovc [15] do just that. In their paper "Algorithmic Trading with Learning" the agent has a dynamic view of future asset prices. Depending on this view the agent will then seek to find the ideal combination of limit and market orders to optimally enter their desired position. Again the optimal policy is given via solving the corresponding QVI numerically.

The literature surrounding the use of reinforcement learning to the problem of optimal

liquidation is somewhat lacking. The problem is typically solved via dynamic programing. An example of this is given in the paper "Online Learning in Limit Order Book Trade Execution" by Akbarzedeh, Tekin and Schaar [14]. Here the impact of a market order on the mid price is learnt i.e. the probability of the mid price decreasing x ticks if y units of inventory are sold. The sample probabilities are then fed into a separate algorithm which solves this problem via dynamic programming.

An alternative approach is to use the techniques present in reinforcement learning to incrementally improve on an already pre determined liquidation schedule. This is the approach used by Hendricks and Wilcox [16]. Here the liquidation schedule used by Almgren & Chriss is used as a base case. A reinforcement learning technique known as Q-learning, which was first introduced by Chris Watkins [17] in 1989, is then implemented to adjust the volume of inventory sold from this base schedule.

Both approaches to implementing reinforcement learning for optimal portfolio liquidation are valid. However, neither tries to directly determine the optimal policy from sampled data, which is the approach taken in this dissertation.

3 Mathematical Background

This section will detail the mathematical background required. A basic understanding of stochastic processes, stochastic differential equations and partial differential equations is assumed. An emphasis here is placed on jump processes due to their usefulness in modelling the impact of orders being filled on the agents wealth.

3.1 Key Definitions, Propositions and Theorems

Throughout this dissertation we work on a filtered probability space denoted by $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\leq t\leq T}, \mathbb{P})$. For all such spaces we assume the so called usually conditions hold. These are:

1. $\mathcal{F}$ is $\mathbb{P}$ complete i.e. for all $\omega' \subset \omega \in \mathcal{F}$ s.t. $\mathbb{P}(\omega) = 0$ then $\omega' \in \mathcal{F}$ and so $\mathbb{P}(\omega') = 0$.
2. $\mathcal{F}_0$ contains all $\mathbb{P}$ null sets and so for all $\{t : 0 \leq t \leq T\}$ $\mathcal{F}_t$ also contains all $\mathbb{P}$ null sets.
3. The filtration is right-continuous i.e. $\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s = \mathcal{F}_t$.

This technical condition allows for the construction of stochastic processes which are càdlàg, i.e. right continuous with a left limit.

Definition 3.1 (Infinitesimal generator) *The infinitesimal generator of a stochastic process $X = (X_t)_{0\leq t\leq T}$ at time t, given by $\mathcal{L}_t$, is defined as:*

$$\mathcal{L}_t f(x) = \lim_{h\downarrow 0} \frac{\mathbb{E}[f(X_{t+h}) \mid X_t = x] - f(x)}{h}.$$

$f(x)$ must be a function that is twice differentiable in x. Intuitively the infinitesimal generator can be thought of as the generalisation of a derivative of a function for stochastic processes.

Definition 3.2 (Multivariate Counting Process) *The m-dimensional $\mathcal{F}$-adapted stochastic process $\mathbf{N} = (\mathbf{N}_t)_{0\leq t\leq T}$, where $m \in \mathbb{N}$, is said to be a multivariate counting process if the following properties hold almost surely:*

1. $\mathbf{N}_0 = \mathbf{0}$

2. $\mathbf{N} \in \mathbb{N}_0^m$

3. $\mathbf{N}_t \cdot \mathbf{N}_t \geq \mathbf{N}_s \cdot \mathbf{N}_s \qquad \forall\, t \geq s$

One of the most prominent examples of counting process is the (one dimensional) Poisson process.

Definition 3.3 (Poisson Process) *A (one dimensional) Poisson Process* $N = (N_t)_{0 \leq t \leq T} \in \mathbb{N}_0$ *is a counting process which satisfies the following properties:*

1. $N_0 = 0$ *a.s.*

2. *The increments are independent i.e.* $N_t - N_s$ *is independent of* $N_v - N_u$ *where* $t > s > v > u$.

3. *The increments* $N_t - N_s$ *are Poisson distributed with parameter* $\lambda(t-s)$ *i.e.*

$$\mathbb{P}(N_t - N_s = n) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}$$

4. *The increments are stationary,* $N_{t+s} - N_s \overset{d}{=} N_t$ *for all* $t, s \geq 0$.

A consequence of this definition is that the time between arrivals of N *are independent and exponentially distributed with* $\mathbb{E}(N_t) = \lambda t$.

Proposition 3.1 (Compensator of a Counting Process) *For all m-dimensional multivariate counting processes, where* $m \in \mathbb{N}$, *there exists a unique m-dimensional right-continuous predictable and increasing process,* $\mathbf{A} = (\mathbf{A}_t)_{0 \leq t \leq T}$, *with* $\mathbf{A}_0 = \mathbf{0}$ *a.s. such that* $\hat{\mathbf{N}} = \mathbf{N} - \mathbf{A}$ *is a local martingale.*

Remark 3.1 (Intensity of a Counting Process) *For a m-dimensional counting process,* $\mathbf{N} = (\mathbf{N}_t)_{0 \leq t \leq T}$, *the m-dimensional intensity rate,* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)_{0 \leq t \leq T}$, *is the stochastic process associated with* $\mathbf{N}$ *such that:*

$$\mathbf{A}_t = \int_0^t \boldsymbol{\lambda}_u du.$$

Where $\mathbf{A} = (\mathbf{A}_t)_{0 \leq t \leq T}$ *is the compensator of the counting process.*

When applying this property to a one dimensional Poisson process the following result is achieved.

Proposition 3.2 (Compensated Poisson Process) . *The compensated Poisson process:*

$$\hat{N} = (\hat{N}_t)_{0 \leq t \leq T} \qquad where \qquad \hat{N}_t = N_t - \lambda t$$

is a martingale.

Definition 3.4 *The stochastic integral of a m-dimensional $\mathcal{F}$-predictable process $\mathbf{g} = (\mathbf{g}_t)_{0 \leq t \leq T}$ with respect to a m-dimensional counting process, $\mathbf{N} = (\mathbf{N}_t)_{0 \leq t \leq T}$, is defined as follows:*

$$\mathbf{Y}_t = \int_0^t \mathbf{g}_{u^-} d\hat{\mathbf{N}}_u.$$

Taking the left limit of g_t ensures the preservation of the local martingale property and so $\mathbf{Y} = (\mathbf{Y}_t)_{0 \leq t \leq T}$ is also a local martingale.

Again applying this result to a one dimensional Poisson process the following result is achieved.

Definition 3.5 *The stochastic integral of a with respect to a compensated Poisson process is defined as follows:*

$$Y_t = \int_0^t g_{u^-} d\hat{N}_u$$

In this case the left limit of the function g means that the integral will be a martingale.

Theorem 3.1 (Multi-dimensional Itô's formula) *Define $\mathbf{X}_t$ to be a diffusion process satisfying the following stochastic differential equation:*

$$d\mathbf{X}_t = \boldsymbol{\mu}(t, \mathbf{X}_t)dt + \boldsymbol{\sigma}(t, \mathbf{X}_t)d\mathbf{W}_t.$$

Where $\mathbf{W}$ is a n-dimensional Brownian motion, i.e. each of the n Brownian motions in the vector are independent. $\boldsymbol{\mu}(t, \mathbf{X}_t)$ is a m-dimensional column vector of the drifts. $\boldsymbol{\sigma}(t, \mathbf{X}_t)$ is a $m \times n$-matrix representing the volatilities. As such X is a m-dimensional stochastic process. Then define the stochastic process $Y = (Y_t)_{0 \leq t \leq T}$ with $\mathbf{Y}_t = f(t, \mathbf{X}_t)$. Where $f(t, \mathbf{x})$

is once differentiable in t and twice differentiable in $\mathbf{x}$. $\mathbf{Y}$ *will be an Itô process and satisfy the following stochastic differential equation:*

$$d\mathbf{Y}_t = (\partial_t f(t, X_t) + \boldsymbol{\mu}(t, \mathbf{X}_t)' \mathbf{D} f(t, X_t) + \tfrac{1}{2} \operatorname{Tr} \boldsymbol{\sigma}(t, X_t)' \boldsymbol{\sigma}(t, X_t) \mathbf{D}^2 f(t, X_t)) dt$$

$$+ \mathbf{D} f(t, \boldsymbol{X}_t)' \sigma(t, \mathbf{X}_t) d\mathbf{W}_t$$

Here $\mathbf{D} f(t, \boldsymbol{X}_t)$ *denotes the vector of first derivative w.r.t* $\mathbf{X}_t$ *and* $\mathbf{D}^2 f(t, \mathbf{X}_t)$ *denotes the matrix of second derivatives w.r.t.* $\mathbf{X}_t$. *Specifically,* $\mathbf{D}^2 f(t, \mathbf{X}_t)_{j,k} = \partial_{x_j, x_k} f(t, \mathbf{X}_t)$.

The generator function for $\mathbf{X}$ on the function $f(t, \mathbf{x})$ will then be as follows:

$$\mathcal{L}_t^{\mathbf{X}} f(\mathbf{x}) = \boldsymbol{\mu}(t, \mathbf{x})' \mathbf{D} f(\mathbf{x}) + \tfrac{1}{2} \operatorname{Tr} \boldsymbol{\sigma}(t, \mathbf{x})' \boldsymbol{\sigma}(t, \mathbf{x}) \mathbf{D}^2 f(\mathbf{x})$$

Theorem 3.2 (Itô's formula for Poisson Process) *Given a stochastic process* $Z = Z_{0 \le t \le T}$ *where* $Z_t = f(t, Y_t)$. *Let* $Y = (Y_t)_{0 \le t \le T}$ *to be a stochastic process which satisfies the following stochastic differential equation:*

$$dY_t = g_{t^-} d\hat{N}_t.$$

Where N *is a one-dimensional Poission process with intensity rate* λ *and* $\hat{N}$ *is the compensated version of* N. *Note* f *is a function once differentiable in* t *and* Y. *Then*

$$dZ_t = \big(\partial_t f(t, Y_t) - \lambda g_t \partial_y f(t, Y_{t^-})\big) dt + \big[f(t, Y_{t^-} + g_{t^-}) - f(t, Y_{t^-})\big] dN_t$$

$$= \big\{\partial_t f(t, Y_t) + \lambda\big(\partial_y [f(t, Y_{t^-} + g_{t^-}) - f(t, Y_{t^-})] - g_t \partial_y f(t, Y_t)\big)\big\} dt + [f(t, Y_{t^-} + g_{t^-}) - f(t, Y_{t^-})] d\hat{N}_t.$$

Where you can interpret the first part of this formula as the change in drift and the second as the adjustment in Z *whenever* $\hat{N}$ *arrives.*

It can be clearly be inferred that the generator for the process Y on the function $f(t, y)$ is as follows:

$$\mathcal{L}_t f(y) = \lambda\big([\partial_y f(t, y + g_{t^-}) - f(t, y)] - g_t \partial_y f(t, y)\big)$$

When producing our market model we will be dealing with stochastic processes that have both diffusion and jumps. As such we must know how to apply Itô's in this circumstance.

Theorem 3.3 (Itô's formula for multi-dimensional jump diffusion processes) *Suppose* $\mathbf{X}$ *is a stochastic process which solves the following stochastic differential equation:*

$$d\mathbf{X}_t = \boldsymbol{\mu}(t, \mathbf{X}_t)dt + \boldsymbol{\sigma}(t, \mathbf{X}_t)d\mathbf{W}_t + \boldsymbol{\gamma}(t^-, \mathbf{X}_{t^-})\mathbf{N}_t.$$

Where $\mathbf{W}$ *is a* n*-dimensional Brownian motion, i.e. each of the* n *Brownian motions in the vector are independent.* $\boldsymbol{\mu}(t, \mathbf{X}_t)$ *is a* m*-dimensional column vector of the drifts.* $\boldsymbol{\sigma}(t, \mathbf{X}_t)$ *is a* $m \times n$*-matrix representing the volatilities.* $\boldsymbol{\gamma}(t^-, \mathbf{X}_{t^-})$ *is a* $m \times p$*-matrix of jump sizes and* $\mathbf{N}$ *is a* p*-dimensional counting process with intensities given by the* p*-dimensional process* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)_{0 \le t \le T}$*. As such* X *is a* m*-dimensional stochastic process. Then define the stochastic process* $\mathbf{Y} = (\mathbf{Y}_t)_{0 \le t \le T}$ *with* $\mathbf{Y}_t = f(t, \mathbf{X}_t)$*. Where* $f(t, \mathbf{x})$ *is once differentiable in* t *and twice differentiable in* $\mathbf{x}$*. Then with a slight abuse of notation:*

$$\begin{aligned} d\mathbf{Y}_t = &\Big\{ \partial_t f(t, \mathbf{X}_t) + \boldsymbol{\mu}(t, \mathbf{X}_t)' \, \mathbf{D}_{\mathbf{x}} f(t, \mathbf{X}_t) + \frac{1}{2} \operatorname{Tr} \boldsymbol{\sigma}(t, \mathbf{X}_t)' \boldsymbol{\sigma}(t, \mathbf{X}_t) \, \mathbf{D}^2_{\mathbf{xx}} f(t, \mathbf{X}_t) \\ &+ \sum_{j=1}^{p} \lambda_j(t, \mathbf{X}_t) \big[f\big(t, \mathbf{X}_t + \boldsymbol{\gamma}_{t^-,j}(t^-, \mathbf{X}_{t^-})\big) - f(t, \mathbf{X}_t) \big] \Big\} dt \\ &+ \mathbf{D}_{\mathbf{x}} f(t, \mathbf{X}_t)' \boldsymbol{\sigma}(t, \mathbf{X}_t) d\mathbf{W}_t \\ &+ \sum_{j=1}^{p} \big[f\big(t, \mathbf{X}_t + \boldsymbol{\gamma}_{t^-,j}(t^-, \mathbf{X}_{t^-})\big) - f(t, \mathbf{X}_t) \big] d\hat{\mathbf{N}}^j_t \end{aligned}$$

Where γ_j *denotes the vector corresponding to the* j^{th} *column of* $\boldsymbol{\gamma}$*,* λ_j *corresponds to the* j^{th} *element of* $\boldsymbol{\lambda}$ *and* $\hat{\mathbf{N}}^j$ *corresponds to the* j^{th} *element of* $\hat{\mathbf{N}}$*.*

It can again be seen that the generator of $\mathbf{X}$ with respect to $f(t, \mathbf{y})$ is as follows:

$$\begin{aligned} \mathcal{L}_t f(t, \mathbf{x}) = &\boldsymbol{\mu}(t, \mathbf{x})' \, \mathbf{D}_{\mathbf{x}} f(t, \mathbf{x}) + \frac{1}{2} \operatorname{Tr} \boldsymbol{\sigma}(t, \mathbf{x})' \boldsymbol{\sigma}(t, \mathbf{x}) \, \mathbf{D}^2_{\mathbf{xx}} f(t, \mathbf{x}) \\ &+ \sum_{j=1}^{p} \lambda_j(t, \mathbf{x}) \big[f\big(t, \mathbf{x} + \boldsymbol{\gamma}_{t^-,j}(t^-, \mathbf{x})\big) - f(t, \mathbf{x}) \big]. \end{aligned}$$

4 Stochastic Optimal Control & Stopping

Stochastic optimal control refers to the sub-field of optimal control which deals with dynamic systems. These systems have stochastic elements either due to uncertainty in the observations or noise that drives the underlying system. Optimal stopping refers to the problem of choosing a time to perform an action in order to maximise some reward. Both stochastic optimal control and optimal stopping has many application to a diverse range of fields such as mathematics, economics and finance. Specifically, in this dissertation I am looking to find the optimum time to issue a market orders, an optimal stopping problem, and the optimal depth to post a limit orders, a stochastic optimal control problem.

4.1 Controlled Jump Diffusion Process

In the problem examined in this dissertation an agent will be exposed to uncertainty driven by both diffusion and jump processes. Specifically, the asset mid price will be driven by a diffusion process and the impact of orders being filled on the agents wealth is given by a jump processes. As such we must examine how to apply stochastic control in these circumstances. Specifically, let $\mathbf{X^u} = (\mathbf{X}^{\mathbf{u}}_t)_{0\leq t\leq T}$ denote a m-dimensional controlled process, representing the wealth process of the agent, which obeys the following stochastic differential equation:

$$d\mathbf{X}^{\mathbf{u}}_t = \boldsymbol{\mu}^{\mathbf{u}}_t dt + \boldsymbol{\sigma}^{\mathbf{u}}_t d\mathbf{W}_t + \boldsymbol{\gamma}^{\mathbf{u}}_{t^-} d\mathbf{N}^{\mathbf{u}}_t. \tag{4.1}$$

$\mathbf{N^u} = (\mathbf{N}^{\mathbf{u}}_t)_{0\leq t\leq T}$ denotes a collection of counting process of dimension p with intensities $\boldsymbol{\lambda}^{\mathbf{u}} = (\boldsymbol{\lambda}^{\mathbf{u}}_t)_{0\leq t\leq T}$ and $\mathbf{W^u} = (\mathbf{W}^{\mathbf{u}}_t)_{0\leq t\leq T}$ is a m-dimensional Brownian motion. Here $\mathbf{u} = (\mathbf{u}_t)_{0\leq t\leq T}$ is the m-dimensional control process. With a slight abuse of notation the coefficients on the diffusion process are then defined as follows:

The controlled m-dimensional drift: $\boldsymbol{\mu}^{\mathbf{u}}_t = \boldsymbol{\mu}(t, \mathbf{X}^{\mathbf{u}}_t, \mathbf{u}_t)$

The controlled $m \times m$-dimensional volatility: $\boldsymbol{\sigma}^{\mathbf{u}}_t = \boldsymbol{\sigma}(t, \mathbf{X}^{\mathbf{u}}_t, \mathbf{u}_t)$

The controlled $m \times p$-dimensional jump size: $\boldsymbol{\gamma}^{\mathbf{u}}_t = \boldsymbol{\gamma}(t, \mathbf{X}^{\mathbf{u}}_t, \mathbf{u}_t)$.

In addition, the intensity of the counting process is assumed to have the following form:

$$\boldsymbol{\lambda}_t^{\mathbf{u}} = \boldsymbol{\lambda}(t, \mathbf{X}_t^{\mathbf{u}}, \mathbf{u}_t)$$

Note that in this general formulation means that the agent has the ability to influence the drift, volatility, jump size and jump arrival. We now characterise the performance criteria of the agent as

$$H^{\mathbf{u}}(t, \mathbf{x}) = \mathbb{E}_{t,\mathbf{x}}\left[G(\mathbf{X}_T^{\mathbf{u}}) + \int_t^T F(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s)ds\right]. \tag{4.2}$$

The agent then seeks to maximise this performance criteria via choosing $\mathbf{u}$ this gives the value function

$$H(t, \mathbf{x}) = \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} H^{\mathbf{u}}(t, \mathbf{x}). \tag{4.3}$$

$G(\mathbf{x}) : \mathbb{R}^m \mapsto \mathbb{R}$ is a terminal payoff that depends only on the final value of the wealth process. F(t,$\mathbf{x}$,$\mathbf{u}$) : $\mathbb{R}_+ \times \mathbb{R}^{2m} \mapsto \mathbb{R}$ represents a running penalty or reward. The size of the jump is also assumed to be bounded i.e.

$$\sum_{j=1}^{p} \int_t^{\tau} \left[H\big(s, \mathbf{X}_s + \boldsymbol{\gamma}_{s^-,j}(s^-, \mathbf{X}_{s^-})\big) - H(s, \mathbf{X}_s)\right] d\hat{\mathbf{N}}_s^j \leq M, \quad M \in \mathbb{R}.$$

Note $\mathbb{E}_{t,\mathbf{x}}[\,\cdot\,] = \mathbb{E}[\,\cdot \,|\, \mathbf{X}_t = \mathbf{x}]$ i.e. the conditional expectation. $\mathcal{A}_{[t,T]}$ corresponds to the set of admissible processes at time t which ensures that (4.1) has a strong solution.

4.2 Optimal Control

4.2.1 Dynamic Programming Principle

In order to solve for the value function and optimal control, $\mathbf{u}$, the value function is broken down into a recursive problem. To show how this can be performed I use adapt the proof given by Touzi in his book Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE [18]. While Tozi derives the dynamic programming principle (DPP) and HJB for the case of a diffusion only I modify this proof so that it is also valid for jump diffusion processes. Specifically, the proof is as follows: consider an arbitrary stopping time

$\tau \in [t,T]$. For a given admissible strategy $\mathbf{u}$ using both the additive properties of the stochastic integral and the law of iterated expectations we can then then re-write (4.2) as

$$\begin{aligned} H^{\mathbf{u}}(t,\mathbf{x}) &= \mathbb{E}_{t,\mathbf{x}}\left[G(\mathbf{X}_T^{\mathbf{u}}) + \int_\tau^T F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right] \\ &= \mathbb{E}_{t,\mathbf{x}}\left[\mathbb{E}_{\tau,\mathbf{X}_\tau^{\mathbf{u}}}\left[G(\mathbf{X}_T^{\mathbf{u}}) + \int_\tau^T F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right] + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right] \\ &= \mathbb{E}_{t,\mathbf{x}}\left[H^{\mathbf{u}}(\tau,\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right]. \end{aligned} \tag{4.4}$$

We can now note that $H(t,\mathbf{x}) \geq H^{\mathbf{u}}(t,\mathbf{x})$ with equality holding if $\mathbf{u}$ is the optimal control, assuming that the optimal control is an admissible strategy. As such

$$H^{\mathbf{u}}(t,\mathbf{x}) \leq \mathbb{E}_{t,\mathbf{x}}\left[H(\tau,\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right]$$

then by taking the supremum over all admissible strategies

$$H(t,\mathbf{x}) \leq \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[H(\tau,\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right]. \tag{4.5}$$

In order to show that this inequality is in fact a equality we consider an ϵ-optimal control, $\mathbf{u}^{\boldsymbol{\epsilon}}$. The ϵ-optimal control is defined to be an admissible optimal control which obeys the following inequality:

$$H(t,\mathbf{x}) \geq H^{\mathbf{u}^{\boldsymbol{\epsilon}}}(t,\mathbf{x}) \geq H(t,\mathbf{x}) - \epsilon.$$

where $\epsilon \in \mathbb{R}$. Taking a modified ϵ-control of the form

$$\tilde{\mathbf{u}}^{\epsilon} = \mathbf{u}\mathbb{1}_{t\leq\tau} + \mathbf{u}^{\boldsymbol{\epsilon}}\mathbb{1}_{t>\tau}$$

where $\mathbf{u}$ is some arbitrary control that may be better or worse than $\mathbf{u}^\epsilon$ and τ a stopping time such that $\tau \in [t, T]$. Noting the following inequality

$$\begin{aligned} H(t,\mathbf{x}) \geq H^{\tilde{\mathbf{u}}^{\boldsymbol{\epsilon}}}(t,\mathbf{x}) &= \mathbb{E}_{t,\mathbf{x}}\left[H^{\tilde{\mathbf{u}}^\epsilon}(\tau, \mathbf{X}_\tau^{\tilde{\mathbf{u}}^\epsilon}) + \int_t^\tau F(s, \mathbf{X}_s^{\tilde{\mathbf{u}}^\epsilon}, \tilde{\mathbf{u}}_s^\epsilon)ds\right] \\ &= \mathbb{E}_{t,\mathbf{x}}\left[H^{\mathbf{u}^\epsilon}(\tau, \mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s)ds\right] \\ &\geq \mathbb{E}_{t,\mathbf{x}}\left[H^{\mathbf{u}}(\tau, \mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s)ds\right] - \epsilon \end{aligned}$$

taking the limit as ϵ approaches 0 from the right the following inequality is obtained

$$H(t,\mathbf{x}) \geq \mathbb{E}_{t,\mathbf{x}}\left[H^{\mathbf{u}}(\tau, \mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s)ds\right]$$

Since this inequlaity holds $\forall\, \mathbf{u} \in \mathcal{A}_{[t,T]}$:

$$H(t,\mathbf{x}) \geq \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[H(\tau, \mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s)ds\right] \tag{4.6}$$

Combining (4.5) and (4.6) gives the dynamic programming principle

Theorem 4.1 (Dynamic Programming Principle) *The value function given in (4.3) satisfies the following equality:*

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

for all $[t,\mathbf{x}] \in [0,T] \times \mathbb{R}^m$ *and stopping times* $\tau \in [t,T]$.

4.2.2 Hamilton Jacobi Bellman Equation

Intuitively the Hamilton-Jacobi-Bellman equation can be thought of as the infintesimal time equivalent of the dynamic programming principle. More formally

$$\tau = T \wedge \inf\{s > t : (s-t, ||\mathbf{X}_s^{\mathbf{u}} - \mathbf{x}||) \notin [0,h) \times [0,\epsilon)\}. \tag{4.7}$$

Where $||\cdot||$ is the euclidean norm. Intuitively τ is the minimum of an arbitrarily small time h and the time taken for X to escape a ball of size ϵ, assuming this is less than T. Now assuming that the value function is sufficiently smooth, once differentiable in t and twice differentiable in $\mathbf{x}$, then Itô's formula for a multidimensional jump diffusion processes can be applied. Hence

$$\begin{aligned} H(\tau, \mathbf{X}_\tau^{\mathbf{u}}) = H(t, \mathbf{x}) + \int_t^\tau \Big\{ & \partial_s H(s, \mathbf{X}_s^{\mathbf{u}}) + \boldsymbol{\mu}(s, \mathbf{X}_s^{\mathbf{u}})' \, \mathbf{D}_{\mathbf{x}} H(s, \mathbf{X}_s^{\mathbf{u}}) \\ & + \frac{1}{2} \operatorname{Tr} \boldsymbol{\sigma}(s, \mathbf{X}_s^{\mathbf{u}})' \boldsymbol{\sigma}(s, \mathbf{X}_s^{\mathbf{u}}) \, \mathbf{D}_{\mathbf{xx}}^2 H(s, \mathbf{X}_s) \\ & + \sum_{j=1}^{p} \lambda_j(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}) \big[H\big(s, \mathbf{X}_s^{\mathbf{u}} + \boldsymbol{\gamma}_{s^-,j}(s^-, \mathbf{X}_{s^-}^{\mathbf{u}}, \mathbf{u})\big) - H(s, \mathbf{X}_s^{\mathbf{u}}) \big] \Big\} ds \\ & + \int_t^\tau \mathbf{D}_{\mathbf{x}} H(s, \mathbf{X}_s^{\mathbf{u}})' \boldsymbol{\sigma}(s, \mathbf{X}_s^{\mathbf{u}}) d\mathbf{W}_s \\ & + \sum_{j=1}^{p} \int_t^\tau \big[H\big(s, \mathbf{X}_s + \boldsymbol{\gamma}_{s^-,j}(s^-, \mathbf{X}_{s^-})\big) - H(s, \mathbf{X}_s) \big] d\hat{\mathbf{N}}_s^j. \end{aligned} \tag{4.8}$$

Noting that $\mathcal{L}_t^{\mathbf{u}}$ represents the infinitesimal generator of $\mathbf{X}_t^{\mathbf{u}}$ (4.8) can be written in short hand as

$$\begin{aligned} H(\tau, \mathbf{X}_\tau^{\mathbf{u}}) = H(t, \mathbf{x}) + \int_t^\tau (\partial_s + \mathcal{L}_s^{\mathbf{u}}) H(s, \mathbf{X}_s^{\mathbf{u}}) ds + \int_t^\tau \mathbf{D}_{\mathbf{x}} H(s, \mathbf{X}_s^{\mathbf{u}})' \boldsymbol{\sigma}(s, \mathbf{X}_s^{\mathbf{u}}) d\mathbf{W}_s \\ + \sum_{j=1}^{p} \int_t^\tau \big[H\big(s, \mathbf{X}_s + \boldsymbol{\gamma}_{s^-,j}(s^-, \mathbf{X}_{s^-})\big) - H(s, \mathbf{X}_s) \big] d\hat{\mathbf{N}}_s^j. \end{aligned} \tag{4.9}$$

Now define $\mathbf{v} \in \mathcal{A}_{[t,T]}$ and constant over the interval $[t, \tau]$ and substituting (4.9) into (4.6) gives the following:

$$\begin{aligned} H(t,\mathbf{x}) &\geq \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\bigg[H(\tau,\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\bigg] \\ &\geq \mathbb{E}_{t,\mathbf{x}}\bigg[H(\tau,\mathbf{X}_\tau^{\mathbf{v}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg] \\ &= \mathbb{E}_{t,\mathbf{x}}\bigg[H(t,\mathbf{X}_\tau^{\mathbf{v}}) + \int_t^\tau (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^\tau \mathbf{D}_{\mathbf{x}}H(s,\mathbf{X}_s^{\mathbf{v}})'\boldsymbol{\sigma}(s,\mathbf{X}_s^{\mathbf{v}})d\mathbf{W}_s \\ &+ \sum_{j=1}^p \int_t^\tau \big[H\big(s,\mathbf{X}_s + \boldsymbol{\gamma}_{s^-,j}(s^-,\mathbf{X}_{s^-})\big) - H(s,\mathbf{X}_s)\big]d\hat{\mathbf{N}}_s^j + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg] \end{aligned} \tag{4.10}$$

Note both $\int_t^\tau \mathbf{D}_{\mathbf{x}}H(s,\mathbf{X}_s^{\mathbf{v}})'\boldsymbol{\sigma}(s,\mathbf{X}_s^{\mathbf{v}})d\mathbf{W}_s$ and $\int_t^\tau \big[H\big(s,\mathbf{X}_s+\boldsymbol{\gamma}_{s^-,j}(s^-,\mathbf{X}_{s^-})\big)-H(s,\mathbf{X}_s)\big]d\hat{\mathbf{N}}_s^j$ are bounded as $||\mathbf{X}_\tau^{\mathbf{v}} - \mathbf{x}|| \leq \epsilon$ and the jump size is bounded. This ensures that stochastic integral with respect to both the multidimensional Brownian motion and counting process are martingales. As such the conditional expectation will be zero. Therefore (4.10) can be written as

$$H(t,\mathbf{x}) \geq \mathbb{E}_{t,\mathbf{x}}\bigg[H(t,\mathbf{X}_\tau^{\mathbf{v}}) + \int_t^\tau (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg].$$

Finally using the pull out property of conditional expectations

$$0 \geq \mathbb{E}_{t,\mathbf{x}}\bigg[\int_t^\tau (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg].$$

Now note that via the definition of τ as h approach's 0 from the right τ in turn approaches t from the right a.s. As such $\tau = t + h$ as the probability that $\mathbf{X}$ will exit the ball approaches zero a.s. Therefore,

$$\begin{aligned} 0 &\geq \lim_{h\to 0^+} \mathbb{E}_{t,\mathbf{x}}\bigg[\int_t^\tau (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg] \\ &= \lim_{h\to 0^+} \mathbb{E}_{t,\mathbf{x}}\bigg[\int_t^{t+h} (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^{t+h} F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\bigg]. \end{aligned}$$

Now recalling that $||\mathbf{X}_\tau - \mathbf{x}|| \leq \epsilon$ ensures that the if the process does hit the ball it is bounded. As such the dominated convergence theorem can be applied. This combined with the mean value theorem and noting that $\mathbf{X}_t^{\mathbf{u}} = \mathbf{x}$ gives the following:

$$\begin{aligned}
0 &\geq \lim_{h\to 0^+} \mathbb{E}_{t,\mathbf{x}}\left[\int_t^{t+h} (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \int_t^{t+h} F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\right] \\
&= \mathbb{E}_{t,\mathbf{x}}\left[\lim_{h\to 0^+}\int_t^{t+h} (\partial_s + \mathcal{L}_s^{\mathbf{v}})H(s,\mathbf{X}_s^{\mathbf{v}})ds + \lim_{h\to 0^+}\int_t^{t+h} F(s,\mathbf{X}_s^{\mathbf{v}},\mathbf{v}_s)ds\right] \\
&= (\partial_t + \mathcal{L}_t^{\mathbf{v}})H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{v}_t)
\end{aligned}$$

It is now obvious that the restriction that $\mathbf{v}$ be constant over the interval $[t,\tau]$ allows for all $\mathbf{u} \in \mathcal{A}_{[t,T]}$ as h approaches 0^+. Therefore,

$$0 \geq \partial_t H(t,\mathbf{x}) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}} H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t)) \tag{4.11}$$

In order to show that this inequality is in-fact an equality let $\mathbf{u}^*$ be the optimal control. Then via the dynamic programming principle

$$H(t,\mathbf{x}) = \mathbb{E}_{t,\mathbf{x}}\left[H(\tau,\mathbf{X}_\tau^{\mathbf{u}^*}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}^*},\mathbf{u}^*{}_s)ds\right].$$

Now by applying a similar procedure to determine (4.11). Specifically, define τ as given in (4.7) apply Itô's formula for a jump diffusion process in order to decompose $H(\tau,\mathbf{X}_\tau^{\mathbf{u}^*})$ into $H(t,\mathbf{x})$ and the corresponding stochastic integrals with respect to time, $\mathbf{W}$ and $\hat{\mathbf{N}}$. Then take the limit as h approaches 0^+ and use the Mean Value Theorem giving:

$$\partial_t H(t,\mathbf{x}) + \mathcal{L}_t^{\mathbf{u}^*} H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t^*) = 0.$$

This is simply the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned}
\partial_t H(t,\mathbf{x}) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}} H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t)) = 0, \\
H(T,\mathbf{x}) = G(\mathbf{x}).
\end{aligned} \tag{4.12}$$

Where the boundary condition, $H(T, \mathbf{x}) = G(\mathbf{x})$, is given via the definition of the value function.

4.2.3 Verification Theorem

(4.12) gives a necessary condition of the value function. In order to show that the solution to the Hamilton-Jacobi-Bellman equation is in turn the solution to the value function the verification theorem is used.

Theorem 4.2 (Verification Theorem) *Let $\psi(t, \mathbf{x}) \in \mathcal{C}^{1,2}([0, T] \times \mathbb{R}^m)$ satisfies a quadratic growth condition i.e.*

$$|\psi(t, \mathbf{x})| \leq C(1 + |\mathbf{x}|^2), \quad \forall\, (\mathbf{x}, t) \in [0, T) \times \mathbb{R}^m.$$

i) Suppose that for all $\mathbf{u} \in \mathcal{A}_{[t,T]}$

$$\partial_t \psi(t, \mathbf{x}) + \mathcal{L}_t^{\mathbf{u}} H(t,\mathbf{x}) + F(t, \mathbf{x}, \mathbf{u_t}) \leq 0$$
$$G(\mathbf{x}) - \psi(T, \mathbf{x}) \leq 0$$

then

$$\psi(t, \mathbf{x}) \geq H^{\mathbf{u}}(t, \mathbf{x}), \quad \forall (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m$$

for all Markov controls $\mathbf{u} \in \mathcal{A}_{[t,T]}$.
ii) Suppose further that for every $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m$, there exists a measurable $\mathbf{u}^*(t, \mathbf{x})$ such that

$$\begin{aligned} 0 &= \partial_t \psi(t, \mathbf{x}) + \left(\mathcal{L}_t^{\mathbf{u}^*(t,\mathbf{x})} H(t, \mathbf{x}) + F(t, \mathbf{x}, \mathbf{u}^*(t, \mathbf{x})) \right) \\ &= \partial_t H(t, \mathbf{x}) + \sup_{\mathbf{u} \in \mathcal{A}_{[t,T]}} \left(\mathcal{L}_t^{\mathbf{u}} H(t, \mathbf{x}) + F(t, \mathbf{x}, \mathbf{u}_t) \right) \end{aligned}$$

with $\psi(T, \mathbf{x}) = G(\mathbf{x})$ and the stochastic differential equation

$$d\mathbf{X}_t^* = \boldsymbol{\mu}(t, \mathbf{X}_t^*, \mathbf{u}^*(t, \mathbf{X}_t^*))dt + \boldsymbol{\sigma}(t, \mathbf{X}_t^*, \mathbf{u}^*(t, \mathbf{X}_t^*))d\mathbf{W}_t + \boldsymbol{\gamma}(t, \mathbf{X}_t^*, \mathbf{u}^*(t, \mathbf{X}_t^*))d\mathbf{N}_t^{\mathbf{u}}$$

admits a unique solution and $\{\mathbf{u}^*(s, \mathbf{X}_s^*)\}_{t \le s \le T} \in \mathcal{A}_{[t,T]}$*, then*

$$H(t, \mathbf{x}) = \psi(t, \mathbf{x}), \quad \forall (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m$$

and $\mathbf{u}^*$ *is an optimal Markov control.*

Proof : A very similar approach as that given in Continuous-time Stochastic Control and Optimization with Financial Applications [19] is used. For i) since $\psi(t, \mathbf{x}) \in \mathcal{C}^{1,2}([0, T] \times \mathbb{R}^m)$ for any stopping time $\tau \in [0, \infty)$ and $s \in [t, \infty)$ by Itô's formula:

$$\begin{aligned}\psi(s \wedge \tau_n, \mathbf{X}^{\mathbf{u}}_{s \wedge \tau_n}) = \psi(t, \mathbf{x}) + \int_t^{s \wedge \tau_n} (\partial_r + \mathcal{L}^{\mathbf{u}}_r)\psi(r, \mathbf{X}^{\mathbf{u}}_r) dr + \int_t^{s \wedge \tau_n} \mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\sigma}(r, \mathbf{X}^{\mathbf{u}}_r) d\mathbf{W}_r \\ + \int_t^{s \wedge \tau_n} \mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\gamma}_r(r, \mathbf{X}^{\mathbf{u}}_r, \mathbf{u}) d\hat{\mathbf{N}}_r\end{aligned}$$

let

$$\tau_n = \inf \left\{ s \ge t : \int_t^s |\mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\sigma}(r, \mathbf{X}^{\mathbf{u}}_r)|^2 dr \ \vee \ \int_t^s |\mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\gamma}_r(r, \mathbf{X}^{\mathbf{u}}_r, \mathbf{u})|^2 d\hat{\mathbf{N}}_r \ge n \right\}$$

note τ_n approaches ∞ as n approaches ∞. As such $\int_t^{s \wedge \tau_n} \mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\sigma}(r, \mathbf{X}^{\mathbf{u}}_r) d\mathbf{W}_r$ and $\int_t^{s \wedge \tau_n} \mathbf{D}_{\mathbf{x}} \psi(r, \mathbf{X}^{\mathbf{u}}_r)' \boldsymbol{\gamma}_r(r, \mathbf{X}^{\mathbf{u}}_r, \mathbf{u}) d\hat{\mathbf{N}}_r$ are martingales. By taking expectations:

$$\mathbb{E}_{t,\mathbf{x}} \left[\psi(s \wedge \tau_n, \mathbf{X}^{\mathbf{u}}_{s \wedge \tau_n}) \right] = \psi(t, \mathbf{x}) + \mathbb{E}_{t,\mathbf{x}} \left[\int_t^{s \wedge \tau_n} (\partial_r + \mathcal{L}^{\mathbf{u}}_r)\psi(r, \mathbf{X}^{\mathbf{u}}_r) dr \right].$$

Given the condition

$$\partial_t \psi(t, \mathbf{x}) + \mathcal{L}^{\mathbf{u}}_t H(t, \mathbf{x}) + F(t, \mathbf{x}, \mathbf{u_t}) \le 0$$

it follows that

$$\mathbb{E}_{t,\mathbf{x}} \left[\psi(s \wedge \tau_n, \mathbf{X}^{\mathbf{u}}_{s \wedge \tau_n}) \right] \le \psi(t, \mathbf{x}) - \mathbb{E}_{t,\mathbf{x}} \left[\int_t^{s \wedge \tau_n} F(r, \mathbf{x}, \mathbf{u_r}) dr \right]. \tag{4.13}$$

Due to the quadratic growth condition

$$|\psi(s \wedge \tau_n, \mathbf{X}^{\mathbf{u}}_{s \wedge \tau_n})| \le C(1 + \sup_{\mathcal{A}_{[t,T]}} |\mathbf{X}^{\mathbf{u}}_s|^2)$$

and noting the following inequality:

$$\left| \int_t^{s\wedge\tau_n} F(r, \mathbf{x}, \mathbf{u_r})dr \right| \leq \int_t^{s\wedge\tau_n} |F(r, \mathbf{x}, \mathbf{u_r})|dr \leq \int_t^T |F(r, \mathbf{x}, \mathbf{u_r})|dr,$$

dominated convergence can then be applied to both sides of (4.13) giving

$$\mathbb{E}_{t,\mathbf{x}}\Big[\psi(s, \mathbf{X}_s^{\mathbf{u}})\Big] \leq \psi(t, \mathbf{x}) - \mathbb{E}_{t,\mathbf{x}}\Big[\int_t^s F(r, \mathbf{x}, \mathbf{u_r})dr\Big]. \tag{4.14}$$

Finally, by setting s to T and using the fact that $G(\mathbf{x}) - \psi(T, \mathbf{x}) \leq 0$ (4.14) becomes

$$\mathbb{E}_{t,\mathbf{x}}\Big[G(\mathbf{x})\Big] \leq \psi(t, \mathbf{x}) - \mathbb{E}_{t,\mathbf{x}}\Big[\int_t^T F(r, \mathbf{x}, \mathbf{u_r})dr\Big]$$

rearranging gives the desired inequality.

For ii) via the same localisation argument given in the proof of i)

$$\mathbb{E}_{t,\mathbf{x}}\Big[\psi(s, \mathbf{X}_s^{\mathbf{u}^*})\Big] = \psi(t, \mathbf{x}) - \mathbb{E}_{t,\mathbf{x}}\Big[\int_t^s (\partial_r + \mathcal{L}_r^{\mathbf{u}^*})\psi(r, \mathbf{X}_r^{\mathbf{u}^*})dr\Big] \tag{4.15}$$

then by using

$$0 = \partial_t\psi(t, \mathbf{x}) + \Big(\mathcal{L}_t^{\mathbf{u}^*(t,\mathbf{x})}H(t, \mathbf{x}) + F(t, \mathbf{x}, \mathbf{u}_t^*)\Big)$$

(4.15) becomes

$$\mathbb{E}_{t,\mathbf{x}}\Big[\psi(s, \mathbf{X}_s^{\mathbf{u}^*})\Big] = \psi(t, \mathbf{x}) - \mathbb{E}_{t,\mathbf{x}}\Big[\int_t^s F(r, \mathbf{x}, \mathbf{u}_r^*)dr\Big]$$

Setting s to T and using the fact that $\psi(T, \mathbf{x}) = G(\mathbf{x})$ gives the desired results

$$\psi(t, \mathbf{x}) = \mathbb{E}_{t,\mathbf{x}}\Big[G(\mathbf{x}) + \int_t^s F(r, \mathbf{x}, \mathbf{u}_r^*)dr\Big] = H^{\mathbf{u}^*}(t, \mathbf{x}) = H(t, \mathbf{x}).$$

The fact that $\mathbf{u}^*$ is optimal follows from the definition of $H(t, \mathbf{x})$ i.e. $H(t, \mathbf{x}) = \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} H^{\mathbf{u}}(t, \mathbf{x})$.

4.3 Optimal stopping

In our model the agent will have the option to execute a market order. The choice of when to execute said order can be modeled via an optimal stopping problem. Similar to a purely optimal control problem the addition of optimal stopping still admits a dynamic programming principle and in-addition have a infinitesimal version in the form of a dynamic programming equation (DPE). In order to formally show how the DPP and DPE is derived the same dynamics defined in subsection 4.1 are used. The value function is now defined as:

$$H(t,\mathbf{x}) = \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} H^{\mathbf{u},\tau}(t,\mathbf{x}). \tag{4.16}$$

Defining the performance criteria as:

$$H^{\mathbf{u},\tau}(t,\mathbf{x}) = \mathbb{E}_{t,\mathbf{x}}\left[G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right]. \tag{4.17}$$

Where τ,$\mathbf{u}$ refers to an admissible stopping time, τ, and control, $\mathbf{u}$. Note $\mathcal{T}_{[t,T]}$ refers to all admissible stopping times at time t. These will be all $\mathcal{F}$-stopping times bounded from below by t and above by T. $\mathcal{A}_{[t,T]}$ will contain all admissible controls, specifically these will be controls that ensure the stochastic differential equation in (4.1) has a strong solution. The value function is also assumed to be sufficiently smooth, once differentiable in t and twice in $\mathbf{x}$.

4.3.1 Dynamic Programming Principle

In order to derive the dynamic programming principle for the case of optimal control & stopping the idea of a stopping region must be introduced. The stopping region is defined as:

$$\mathcal{S} = \left\{(t^*,\mathbf{x}) \in [t,T]\times\mathbb{R}^m : H(t^*,\mathbf{x}) = G(\mathbf{x}) + \int_t^{t^*} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right\} \tag{4.18}$$

Intuitively the stopping region can be thought of as the area in which the stopping time has occurred. Consider now another arbitrary stopping time $\theta \in [t,T]$ and following a similar

procedure to that use in (4.4)

$$
\begin{aligned}
H^{\mathbf{u},\tau}(t,\mathbf{x}) &= \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + \left(G(\mathbf{X}^{\mathbf{u}}_{\boldsymbol{\theta}}) + \int_t^{\theta} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau\geq\theta}\right] \\
&= \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + \left(\mathbb{E}_{\theta,\mathbf{X}^{\mathbf{u}}_{\theta}}\left[G(\mathbf{X}^{\mathbf{u}}_{\theta}) + \int_t^{\theta} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right]\right)\mathbb{1}_{\tau\geq\theta}\right] \\
&= \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H^{\mathbf{u},\tau}(\theta,\mathbf{X}^{\mathbf{u}}_{\theta})\mathbb{1}_{\tau\geq\theta}\right].
\end{aligned}
\tag{4.19}
$$

With the second equality holding via the law of iterated expectations and the pull out property i.e.

$$
\begin{aligned}
&\mathbb{E}_{t,\mathbf{x}}\left[\mathbb{E}_{\theta,\mathbf{X}^{\mathbf{u}}_{\theta}}\left[G(\mathbf{X}^{\mathbf{u}}_{\theta}) + \int_t^{\theta} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right]\mathbb{1}_{\tau\geq\theta}\right] \\
&\qquad\qquad = \mathbb{E}_{t,\mathbf{x}}\left[\mathbb{E}_{\theta,\mathbf{X}^{\mathbf{u}}_{\theta}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\theta}) + \int_t^{\theta} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau\geq\theta}\right]\right].
\end{aligned}
$$

Now taking the same approach use to derive the DPP for the case of optimal control we note that $H(t,x) \geq H^{\boldsymbol{u},\tau}(t,\mathbf{x})$, therefore

$$
H^{\mathbf{u},\tau}(t,\mathbf{x}) \leq \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H(\theta,\mathbf{X}^{\mathbf{u}}_{\theta})\mathbb{1}_{\tau\geq\theta}\right].
$$

Taking the supremum over all admissible stopping times and controls

$$
H(t,\mathbf{x}) \leq \sup_{\tau\in\mathcal{T}_{[t,T]}}\sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}^{\mathbf{u}}_{\tau}) + \int_t^{\tau} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H(\theta,\mathbf{X}^{\mathbf{u}}_{\theta})\mathbb{1}_{\tau\geq\theta}\right] \tag{4.20}
$$

In order to prove this inequality is an equality again the concept of an ϵ-optimal control is used. Given the new notation this is defined as:

$$
H(t,\mathbf{x}) \geq \sup_{\tau\in\mathcal{T}_{[t,T]}} H^{\mathbf{u}^{\boldsymbol{\epsilon}},\tau}(t,\mathbf{x}) \geq H(t,\mathbf{x}) - \epsilon.
$$

where $\epsilon \in \mathbb{R}$. Taking a modified ϵ-control of the form:

$$
\tilde{\mathbf{u}}^{\epsilon} = \mathbf{u}\mathbb{1}_{t\leq\theta} + \mathbf{u}^{\boldsymbol{\epsilon}}\mathbb{1}_{t>\theta}
$$

where $\mathbf{u}$ is some arbitrary control that may be better or worse than $\mathbf{u}^\epsilon$ and θ a stopping time such that $\theta \in [t, T]$. The following inequality can then be derived:

$$\begin{aligned}
H(t,\mathbf{x}) &\geq \sup_{\tau\in\mathcal{T}_{[t,T]}} H^{\tilde{\mathbf{u}}^\epsilon,\boldsymbol{\tau}}(t,\mathbf{x}) \\
&= \sup_{\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}_\tau^{\tilde{\mathbf{u}}^\epsilon}) + \int_t^\tau F(s,\mathbf{X}_s^{\tilde{\mathbf{u}}^\epsilon},\tilde{\mathbf{u}}_s^\epsilon)ds\right)\mathbb{1}_{\tau<\theta} + H^{\tilde{\mathbf{u}}^\epsilon,\tau}(\theta,\mathbf{X}_\theta^{\tilde{\mathbf{u}}^\epsilon})\mathbb{1}_{\tau\geq\theta}\right] \\
&= \sup_{\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H^{\mathbf{u}^\epsilon,\tau}(\theta,\mathbf{X}_\theta^{\mathbf{u}^\epsilon})\mathbb{1}_{\tau\geq\theta}\right] \\
&\geq \sup_{\tau\in\mathcal{T}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H(\theta,\mathbf{X}_\theta^{\mathbf{u}})\mathbb{1}_{\tau\geq\theta}\right] - \epsilon
\end{aligned}$$

taking the limit as ϵ approaches 0 from the right the following inequality is obtained

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

Given that this inequlaity holds $\forall\, \mathbf{u} \in \mathcal{A}_{[t,T]}$ and $\forall\, \tau \in \mathcal{T}_{[t,T]}$:

$$H(t,\mathbf{x}) \geq \sup_{\tau\in\mathcal{T}_{[t,T]}} \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \mathbb{E}_{t,\mathbf{x}}\left[\left(G(\mathbf{X}_\tau^{\mathbf{u}}) + \int_t^\tau F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\right)\mathbb{1}_{\tau<\theta} + H(\theta,\mathbf{X}_\theta^{\mathbf{u}})\mathbb{1}_{\tau\geq\theta}\right] \quad (4.21)$$

Combining both (4.20) and (4.21) gives the DPP for optimal control stopping problems.

Theorem 4.3 (Dynamic Programming Principle for Optimal Control & Stopping) *The value function given in (4.21) satisfies the following equality:*

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

for all $[t,\mathbf{x}] \in [0,T]\times\mathbb{R}^m$ *and stopping times* $\theta \in [t,T]$.

4.3.2 Hamilton Jacobi Bellman Equation

Theorem 4.4 (Dynamic Programming Equation for Optimal Control & Stopping) *Assume that the value function* $H(t,\mathbf{x})$ *is once differentiable in* t *and twice in* $\mathbf{x}$ *and that*

$G : \mathbb{R}^m \to \mathbb{R}$ is continuous. Then H solves the quasi-variational inequality (QVI):

$$\max\left\{\partial_t H(t,\mathbf{x}) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} \left(\mathcal{L}_t^{\mathbf{u}} H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t)\right), G(\mathbf{x}) - H(t,\mathbf{x})\right\} = 0, \quad on \quad \mathcal{D}$$

where $\mathcal{D} = [0,T] \times \mathbb{R}^m$.

Proof : Here the proof followed by Tozi [18] is adapted such that it is valid for the case of optimal stopping and control rather than just stopping. First note that $\tau = t$ is an admissible strategy, as such by setting $\tau = t$ and using the pull out property of conditional expectations with respect to the performance criteria the following inequality is achieved

$$H(t,\mathbf{x}) \geq H^{\mathbf{u},t}(t,\mathbf{x}) = G(\mathbf{x}).$$

Therefore, $G(\mathbf{x}) - H(t,\mathbf{x}) \geq 0$. In order to show for any $(t_0,\mathbf{x}_0) \in \mathcal{D}$ this inequality holds a sequence of stopping times is defined as

$$\theta_h = \inf\{t > t_0 : (t, ||\mathbf{X}_t - \mathbf{x}_0||) \notin [t_0, t_0+h] \times 1\}. \tag{4.22}$$

Taking $\tau = t_0$ then using the dynamic programming principle for optimal control and noting that θ_h includes t_0 in its range of possible values

$$H(t_0,\mathbf{x}_0) \geq \mathbb{E}_{t_0,\mathbf{x}_0}\left[H(\theta_h, \mathbf{X}^{\mathbf{u}}_{\theta_h}) + \int_{t_0}^{\theta_h} F(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s)ds\right].$$

Now applying Itô's formula for a jump diffusion process note that

$$\begin{aligned} H(\theta_h,\mathbf{X}^{\mathbf{u}}_{\theta_h}) = H(t_0,\mathbf{x}_0) &+ \int_{t_0}^{\theta_h} (\partial_s + \mathcal{L}^{\mathbf{u}}_s) H(s,\mathbf{X}^{\mathbf{u}}_s)ds + \int_{t_0}^{\theta_h} \mathbf{D}_{\mathbf{x}} H(s,\mathbf{X}^{\mathbf{u}}_s)' \boldsymbol{\sigma}(s,\mathbf{X}^{\mathbf{u}}_s) d\mathbf{W}_s \\ &+ \sum_{j=1}^{p} \int_{t_0}^{\theta_h} \left[H\big(s, \mathbf{X}^{\mathbf{u}}_s + \boldsymbol{\gamma}_{s^-,j}(s^-, \mathbf{X}^{\mathbf{u}}_{s^-})\big) - H(s,\mathbf{X}^{\mathbf{u}}_s)\right] d\hat{\mathbf{N}}^j_s \end{aligned}$$

Given that $\mathbf{X}$ is bounded by a unit ball plus the potential of a bounded jump the stochastic integrals with respect to the Brownian motion and counting process will disappear under

expectation. Therefore,

$$0 \geq \mathbb{E}_{t_0,\mathbf{x}_0}\bigg[\int_{t_0}^{\theta_h} (\partial_s + \mathcal{L}_s^{\mathbf{u}})H(s,\mathbf{X}_s^{\mathbf{u}}) + F(s,\mathbf{X}_s^{\mathbf{u}},\mathbf{u}_s)ds\bigg].$$

dividing through by h, taking the limit as h approaches 0^+ and using the mean value theorem implies

$$0 \geq (\partial_t + \mathcal{L}_t^{\mathbf{u}})H(t_0,\mathbf{x}_0) + F(t_0,\mathbf{x}_0,\mathbf{u}_{t_0}).$$

Finally, noting that this inequality holds $\forall\, \mathbf{u} \in \mathcal{A}_{[t,T]}$ therefore,

$$\partial_t H(t_0,\mathbf{x}_0) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}}H(t_0,\mathbf{x}_0) + F(t_0,\mathbf{x}_0,\mathbf{u}_{t_0})). \tag{4.23}$$

A proof by contradiction is used to show that

$$\max\left\{\partial_t H(t,\mathbf{x}) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}}H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t))\,,\, G(\mathbf{x}) - H(t,\mathbf{x})\right\} \geq 0, \quad on \quad \mathcal{D}.$$

If this does not hold then there exists a point $(t_0,\mathbf{x}_0) \in \mathcal{D}$ such that both

$$\partial_t H(t_0,\mathbf{x}_0) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}}H(t_0,\mathbf{x}_0) + F(t_0,\mathbf{x}_0,\mathbf{u}_{t_0})) < 0 \tag{4.24}$$

and

$$G(\mathbf{x}_0) - H(t_0,\mathbf{x}_0) < 0. \tag{4.25}$$

It can now be shown that (4.24) and (4.25) contradicts the DPP (Theorem 4.3). Consider a new function φ_ϵ which perfectly approximates the value function at $(t_0,\mathbf{x}_0)$ but locally dominates it. Specifically:

$$\varphi_\epsilon(t,\mathbf{x}) := H(t,\mathbf{x}) + \epsilon\big(||\mathbf{x}-\mathbf{x}_0||^4 + |t-t_0|^2\big), \qquad \forall (t,\mathbf{x}) \in \mathcal{D}, \qquad \epsilon > 0$$

under (4.24) and (4.25) there exists a $h > 0$ and $\delta > 0$ such that

$$\begin{aligned} &H(t,\mathbf{x}) \geq G(\mathbf{x}) + \delta \quad and \quad \partial_t H(t,\mathbf{x}) + \sup_{\mathbf{u}\in\mathcal{A}_{[t,T]}} (\mathcal{L}_t^{\mathbf{u}} H(t,\mathbf{x}) + F(t,\mathbf{x},\mathbf{u}_t)) < 0 \\ &on \quad \mathcal{D}_h := [t_0, t_0+h] \times \mathcal{B}_h \quad where \quad \mathcal{B}_h = \{\mathbf{x} \in \mathbb{R}^m : ||\mathbf{x} - \mathbf{x}_0|| \leq h\}. \end{aligned} \tag{4.26}$$

φ_ϵ being locally larger than H implies

$$-\zeta := \max_{\partial\mathcal{D}_h}(H - \varphi_\epsilon) < 0 \tag{4.27}$$

with $\partial\mathcal{D}_h$ represents the boundary of the set $\mathcal{D}_h$. Take the stopping time, θ, given by

$$\theta := \inf\left\{t > t_0 : (t, \mathbf{X}_t) \notin \mathcal{D}_h\right\} \tag{4.28}$$

and an arbitrary stopping time $\tau \in \mathcal{T}_{[t,T]}$ then take $\psi = \tau \wedge \theta$ then

$$H(\psi, \mathbf{X}_\psi) - H(t_0, \mathbf{x}_0) = (H - \varphi_\epsilon)(\psi, \mathbf{X}_\psi) + (\varphi_\epsilon(\psi, \mathbf{X}_\psi) - \varphi_\epsilon(t_0, \mathbf{x}_0)). \tag{4.29}$$

Given that φ_ϵ and H are equal at $(t_0, \mathbf{x}_0)$ using Itô's formula and the fact that $\mathbf{X}_\psi$ is bounded then

$$\mathbb{E}_{t_0,\mathbf{x}_0}\left[\varphi_\epsilon(\psi, \mathbf{X}_\psi) - \varphi_\epsilon(t_0, \mathbf{x}_0)\right] = \mathbb{E}_{t_0,\mathbf{x}_0}\left[\int_{t_0}^{\psi} \partial_t H(t, \mathbf{X}_t) + \mathcal{L}_t^{\mathbf{u}} H(t, \mathbf{X}_t) dt\right]. \tag{4.30}$$

Using (4.29) and (4.30) gives

$$\begin{aligned} &H(\psi, \mathbf{X}_\psi) - H(t_0, \mathbf{x}_0) + \int_{t_0}^{\psi} F(t, \mathbf{X}_t, \mathbf{u}_t) dt \\ &= (H - \varphi_\epsilon)(\psi, \mathbf{X}_\psi) + \int_{t_0}^{\psi} \partial_t H(t, \mathbf{X}_t) + \mathcal{L}_t^{\mathbf{u}} H(t, \mathbf{X}_t) + F(t, \mathbf{X}_t, \mathbf{u}_t) dt \end{aligned} \tag{4.31}$$

Combining this result with (4.29) gives

$$\mathbb{E}_{t_0,\mathbf{x}_0}\left[H(\psi, \mathbf{X}_\psi) - H(t_0, \mathbf{x}_0) + \int_{t_0}^{\psi} F(t, \mathbf{X}_t, \mathbf{u}_t) dt\right] \leq \mathbb{E}_{t_0,\mathbf{x}_0}\left[(H - \varphi)(\psi, \mathbf{X}_\psi)\right] \leq -\zeta\, \mathbb{P}(\tau \geq \theta) \tag{4.32}$$

with the second inequality following from

$$(H-\varphi)(\psi,\mathbf{X}_\psi)\leq-\zeta\,\mathbb{1}_{\tau\geq\theta}$$

rearranging (4.32) for $H(t_0,\mathbf{x}_0)$ gives

$$\begin{aligned}H(t_0,\mathbf{x}_0)&\geq\zeta\,\mathbb{P}(\tau\geq\theta)+\mathbb{E}_{t_0,\mathbf{x}_0}\big[H(\psi,\mathbf{X}_\psi)+\int_{t_0}^{\psi}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\big]\\&=\zeta\,\mathbb{P}(\tau\geq\theta)+\mathbb{E}_{t_0,\mathbf{x}_0}\bigg[\bigg(H(\tau,\mathbf{X}_\tau)+\int_{t_0}^{\tau}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\bigg)\mathbb{1}_{\tau<\theta}\\&+\bigg(H(\theta,\mathbf{X}_\theta)+\int_{t_0}^{\theta}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\bigg)\mathbb{1}_{\tau\geq\theta}\bigg].\end{aligned}$$

Then by using the fact that as given in (4.26) on $\mathcal{D}_h$ $H(t,\mathbf{x})\geq G(\mathbf{x})+\delta$

$$\begin{aligned}H(t_0,\mathbf{x}_0)&\geq\zeta\,\mathbb{P}(\tau\geq\theta)+\mathbb{E}_{t_0,\mathbf{x}_0}\bigg[\bigg(G(\mathbf{X}_\tau)+\delta+\int_{t_0}^{\tau}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\bigg)\mathbb{1}_{\tau<\theta}\\&+\bigg(G(\mathbf{X}_\theta)+\delta+\int_{t_0}^{\tau}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\bigg)\mathbb{1}_{\tau\geq\theta}\bigg]\\&=\zeta\,\mathbb{P}(\tau\geq\theta)+\delta\\&+\mathbb{E}_{t_0,\mathbf{x}_0}\bigg[\bigg(G(\mathbf{X}_\tau)+\int_{t_0}^{\tau}F(t,\mathbf{X}_t,\mathbf{u}_t)dt\bigg)\mathbb{1}_{\tau<\theta}+H(\theta,\mathbf{X}_\theta)\mathbb{1}_{\tau\geq\theta}\bigg].\end{aligned}$$

Given the arbitrariness of τ and the fact that $\zeta\,\mathbb{P}(\tau\geq\theta)+\delta$ is positive, this contradicts the DPP.

5 Optimal Portfolio Liquidation with Limit and Market Orders

In this section I will outline the portfolio liquidation problem that will be studied in detail in this dissertation. Specifically, the dynamics of the market model, performance criteria and value function will all be parametised. I will then proceed to solve the Quasi Variational inequality for this market and examine the results. Thought this section I formulate the optimal controla and stopping problem in a similar way as that given by Cartea, Jaimungal and Penalva in their book Algorithmic and High-Frequency Trading [6].

5.1 Market Dynamics

In our market an agent looks to liquidate a given amount of inventory over a finite time horizon using a combination of both market orders and limit orders. The following notation will be used:

η - volume of inventory the agent wants to liquidate, where $\eta \in \mathbf{N}_0$.

T - Terminal date at which liquidation must have occurred, $T \in \mathbb{R}_+$.

$S = (S_t)_{0 \leq t \leq T}$ - Assets mid price, the arithmetic average of the bid and ask price. In this market model $S_t = S_0 + \sigma W_t$ where $W = (W_t)_{0 \leq t \leq T}$ is a Brownian motion and S_0, $\sigma \in \mathbb{R}_+$.

$\delta = (\delta_t)_{0 \leq t \leq T}$ - Depth at which a limit sell order is posted i.e. the agent will post a sell limit order at time t at $S_t + \delta_t$.

$M = (M_t)_{0 \leq t \leq T}$ - A Poisson process, with intensity λ which corresponds to the number of market buy orders that have occurred.

$P(\delta) = exp(-\kappa\delta)$, $\kappa > 0$. - The probability that a limit order of depth δ is filled conditional on a market buy order arriving.

$F = (F_t)_{0 \leq t \leq T}$ - Counting process for the agents market sell orders.

$\boldsymbol{\tau}$= $\{\tau_k : 1 \leq k \leq K\}$ - The increasing sequence of stopping times at which the agent executes a market order, where $K \leq \eta$. Here each stopping time represents the time to execute a single market order for a single unit of inventory.

$N^\delta = (N_t^\delta)_{0 \leq t \leq T}$ - The controlled counting process, controlled via the negative correlation with respect to the depth, corresponding to the number of the agents limit sell orders which are lifted.

ξ - The half spread, (best ask - best bid)/2, can be thought of as the cost of initiating a market order. This is assumed to always be constant, i.e. a market order will not walk the limit order book, a realistic assumption only if the agents market orders are a small proportion of total volume.

$Q_t^\delta = \eta - N_t^\delta - F_t$ - The remaining inventory to be sold.

$X^\delta = (X_t^\delta)_{0 \leq t \leq T}$ - Agents cash process which satisfies the following stochastic differential equation:

$$dX_t^{\boldsymbol{\tau},\delta} = (S_t + \delta_{t^-})dN_t^\delta + (S_t - \xi)dF_t^{\boldsymbol{\tau}} \tag{5.1}$$

i.e. for all τ_k in the set $\boldsymbol{\tau}$ the stopped process $X^{\tau_k,\delta}$ obeys the dynamics defined in (5.1).

5.2 Optimisation Problem

In this problem the agent seeks to maximise the amount of cash received from the sale of the inventory, given by the cash process X^δ. All inventory must be liquidated by time T. This is done via the agent choosing the optimal time to post market orders, as given by the set of stopping times $\boldsymbol{\tau}$ and optimal depth δ to post limit orders. When a market order is issued the agent receives the mid price minus the half spread, $S - \xi$. When a limit order is posted if it is filled the agent receives the mid price plus the depth the limit order is placed at, $S + \delta$. Noting that conditional on a market buy order arriving the probability of a limit order with depth δ being filled is $P(\delta) = \exp(-\kappa\delta)$. It is clear that when posting a limit order the agent wants to post at a δ as large as possible as this ensures the maximum price but not too large

as the order would never be filled. It is trivial to see that in the scenario described above there would be a strong incentive to not post a market order. This is because the agent will always receive a greater price per unit of inventory sold if a limit order is used. This coupled with the fact that $\mathbb{E}_{t,S}[S_u] = S_t \quad \forall\, t \leq u \in [0,T]$ means that it is never optimal to execute a market order except at the terminal time. If the optimal depth to post a limit order at time t is δ_t and $\mathbb{E}_{t,S}[S_u] \leq S_t - \xi - \delta_t \quad \forall\, t \leq u \in [0,T]$, i.e. there is a sufficiently large negative drift that effects the mid price, then it may be optimal to execute a market order before the terminal date. This is because by the time a limit order is filled on average the mid price may have decreased by more than the extra compensation you receive for posting a limit order adjusted for the cost of posting a market order. However, this is not the case in the model we have created. As such in order to ensure that it will be optimal to execute a market order at a time other than the terminal date a deterministic liquidity target schedule is specified. The agent is then penalised if inventory deviates from this liquidation schedule. If this penalisation is large enough and the agents inventory level deviates too greatly from the target schedule then it may make sense for the agent to issue a market order. Formally the agents performance criteria is defined as:

$$H^{\boldsymbol{\tau}}(t,x,S,q,\delta) = \mathbb{E}_{t,x,S,q}\left[X_\theta^{\boldsymbol{\tau},\delta} + Q_\theta^{\boldsymbol{\tau},\delta}(S_\theta - \xi) - \phi\int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_u)^2 du\right]. \tag{5.2}$$

Where $\mathfrak{q} = (\mathfrak{q}_t)_{0\leq t\leq T}$ denotes a deterministic liquidity target schedule. $\phi \in \mathbb{R}_+$ and is a parameter that represents how closely the agent wishes to stick to the target liquidation schedule. $\theta = T \wedge \inf\{t : Q_t^{\boldsymbol{\tau},\delta} = 0\}$ is a stopping time which ensures that should no inventory be left no more market or limit orders will be executed, this places a lower bound on $Q^{\boldsymbol{\tau},\delta}$ of 0. $\mathbb{E}_{t,x,S,q}[\cdot]$ denotes the expectation conditional on $X_{t^-}^{\boldsymbol{\tau},\delta} = x$, $S_{t^-} = S$, $Q_{t^-}^{\boldsymbol{\tau},\delta} = q$. Intuitively $X_\theta^{\boldsymbol{\tau},\delta}$ denotes the contribution from the inventory sold up until the termination time via the choice of δ and $\boldsymbol{\tau}$. The impact from liquidating all remaining inventory via a market sell order at T is denoted by $Q_\theta^{\boldsymbol{\tau},\delta}(S_\theta - \epsilon)$. Finally, $\phi\int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_u)^2 du$ penalises deviating from

the target liquidation schedule. The value function is then:

$$H(t,x,S,q) = \sup_{(\boldsymbol{\tau},\delta)\in\mathcal{A}_{[t,T]}} H^{\boldsymbol{\tau}}(t,x,S,q,\delta). \tag{5.3}$$

Where $\mathcal{A}_{[t,T]}$ denotes the admissible depths and set of admissible stopping times at time t. The admissible depths are given by $\delta_t \in [0, S_t]$ and each $\tau_k \in \boldsymbol{\tau}$ is given by $\tau_k \in \{s \in [t,T] : Q^{\delta}_s > 0\}$. From now on where it is clear the dependence on (t,x,S,q) will be omitted.

5.3 The Dynamic Programming Equation

Applying Theorem (4.4) to (5.3) the following quasi-variational inequality is obtained:

$$\begin{aligned} 0 = \max\Bigg\{ & \partial_t H + \tfrac{1}{2}\sigma\partial_{SS}H - \phi(q-\mathfrak{q}_{\mathfrak{t}})^2 \\ & + \sup_{\delta\,\in\,\mathcal{A}_{[t,T]}} \lambda\, e^{-\kappa\,\delta}\Big[H(t,x+(S+\delta),S,q-1) - H(t,x,S,q)\Big]; \\ & \Big[H(t,x+(S-\xi),S,q-1) - H(t,x,S,q)\Big]\Bigg\} \end{aligned} \tag{5.4}$$

with boundary and terminal conditions

$$\begin{aligned} H(t,x,S,0) &= x - \phi\int_t^T \mathfrak{q}^2 du \\ H(T,x,S,q) &= x + q(S-\xi) \end{aligned} \tag{5.5}$$

Where λ corresponds to the intensity of the counting process representing the arrival of market buy orders and $P(\delta) = e^{-\kappa\delta}$ is the conditional probability of one of the agents limit sell order being filled. It is clear that the infinitesimal generator is given by

$$\mathcal{L}^{\delta}_t = \tfrac{1}{2}\sigma\partial_{SS}H + \lambda\, e^{-\kappa\,\delta}\Big[H(t,x+(S+\delta),S,q-1) - H(t,x,S,q)\Big].$$

What is less clear is the second component of the QVI,

$$H(t,x+(S+\delta),S,q-1) - H(t,x,S,q).$$

This follows from the fact that the agent has a choice to either issue a limit or market order, this is represented by the max operator. When a market order is issued there is an increase in the wealth process equivalent to $S+\delta$ and a decrease in inventory of one unit. Intuitively these market orders will only be executed when the marginal gains from issuing one are equivalent to the gains from not issuing one i.e.

$$H(t, x+(S+\delta), S, q-1) = H(t, x, S, q).$$

Note that in the case of a single unit of inventory this equation becomes far more similar to that given in Theorem (4.4), specifically

$$G(x+S+\delta) = H(t, x, S, 1).$$

Within the continuation region, when a limit order is posted, the coefficients have the following interpretation.

∂_{SS} - Refers to the generator of the Brownian motion that drives the mid price.

$-\phi(q-\mathfrak{q}_{\mathrm{u}})^2$ - The impact of the running inventory penalty.

The supremum over δ represents the ability of the agent to choose the depth of limit order posted.

$\lambda\, e^{-\kappa\,\delta}$ - gives the rate at which each limit order posted by the agent is filled.

$H(t, x+(S+\delta), S, q-1) - H(t, x, S, q)$ - Gives the increase in the value function from the filling of a limit order, the inventory will decrease by one unit and the wealth process increase by $S+\delta$.

In order to simplify the QVI an ansatz is proposed regarding the form of the value function. Specifically, given the form of the boundary and terminal condition it is assumed

$$H(t, x, S, q) = x + qS + h(t, q). \tag{5.6}$$

Where $x + qS$ corresponds to the current value of the wealth process plus the book value of the current inventory. $h(t, q)$ represents the value of optimally liquidating the remaining inventory. Using (5.6) the QVI given in (5.4) becomes

$$0 = \max\left\{ \frac{d}{dt}h - \phi(q - \mathfrak{q}_t)^2 + \sup_{\delta \in \mathcal{A}_{[t,T]}} \lambda\, e^{-\kappa\,\delta}\Big[\delta + h(t, q-1) - h(t, q)\Big]; \right. \\ \left. - \xi + h(t, q-1) - h(t, q) \right\} \tag{5.7}$$

with boundary and terminal conditions

$$\begin{aligned} h(t, 0) &= -\phi \int_t^T \mathfrak{q}_u^2 du \\ h(T, q) &= -q\xi. \end{aligned} \tag{5.8}$$

By looking at the first order condition for the supremum

$$\begin{aligned} 0 &= \partial_\delta\{\lambda\, e^{-\kappa\delta}[\delta + h(t, q-1) - h(t, q)]\} \\ &= \lambda(-\kappa e^{-\kappa\delta}[\delta + h(t, q-1) - h(t, q)] + e^{-\kappa\delta}) \\ &= \lambda e^{-\kappa\delta}(-\kappa[\delta + h(t, q-1) - h(t, q)] + 1) \end{aligned}$$

the optimal depth in feedback form is then given by

$$\delta^* = \frac{1}{\kappa} + [h(t, q) - h(t, q-1)] \tag{5.9}$$

The interpretation being that $\frac{1}{\kappa}$ represents the optimal depth to post a limit order in order to maximise the profit from a round trip liquidating the mid-price, assuming a purchase at the mid-price. In other words the δ that maximises $\delta P(\delta)$. $h(t, q) - h(t, q-1)$ represents the reservation price i.e. the additional wealth required to ensure that if a unit of inventory is liquidated the value function remains unchanged. Specifically, the p that ensures $H(t, x + p, S, q-1) = H(t, x, S, q)$. The timing of market orders also has a feedback form given by:

$$h(\tau_q, q-1) - h(\tau_q, q) = \xi. \tag{5.10}$$

The interpretation being that a market order is executed whenever the value function is increase by a value equivalent to the half spread. Furthermore by observing (5.9) and (5.10) a lower bound on the depth of

$$\delta^* \geq \frac{1}{\kappa} - \xi$$

can be made. Noting now that in this model $\delta > 0$ as it is assumed a premium will always be paid for liquidity providers, as such $\xi < \frac{1}{\kappa}$. Substituting (5.9) into (5.7)

$$\begin{aligned} 0 = \max \Bigg\{ & \tfrac{d}{dt}h - \phi(q - \mathfrak{q}_t)^2 + \tfrac{e^{-1}\lambda}{\kappa}\, e^{-\kappa\,[h(t,q)-h(t,q-1)]}\,; \\ & - \xi + h(t, q-1) - h(t,q) \Bigg\} \end{aligned} \tag{5.11}$$

The DPE is further reduced by using the transformation

$$h(t,q) = \tfrac{1}{\kappa}\,\log \omega(t,q). \tag{5.12}$$

This gives the final reduced DPE

$$\begin{aligned} 0 = \max \Bigg\{ & (\tfrac{d}{dt} - \kappa\phi(q - \mathfrak{q}_t)^2)\omega(t,q) + e^{-1}\lambda\omega(t,q-1)\,; \\ & e^{-\kappa\xi}\omega(t,q-1) - \omega(t,q) \Bigg\} \end{aligned} \tag{5.13}$$

with boundary and terminal conditions

$$\begin{aligned} \omega(t,0) &= e^{-\kappa\phi\int_t^T \mathfrak{q}_u^2 du} \\ \omega(T,q) &= e^{-\kappa q\xi}. \end{aligned} \tag{5.14}$$

A market order will be executed if

$$e^{-\kappa\xi}\omega(t,q-1) > \omega(t,q). \tag{5.15}$$

Intuitively this means a market order will only be executed if the increase in ω is sufficiently large to overcome the cost of issuing the market order ξ.

5.4 Numerical Solution

The QVI given in (5.13) can be viewed as a system of PDEs. In order to solve these PDEs one works backwards first solving for $\omega(t,1)$ and then using this to solve $\omega(t,2)$ and so on. For example when $q=1$ (5.13) will take the following form:

$$0 = \max\Bigg\{(\tfrac{d}{dt} - \kappa\phi(1-\mathfrak{q}_t)^2)\omega(t,1) + e^{-1}\lambda e^{-\kappa\phi\int_t^T \mathfrak{q}_u^2 du}\,; \\ e^{-\kappa\xi}e^{-\kappa\phi\int_t^T \mathfrak{q}_u^2 du} - \omega(t,1)\Bigg\} \tag{5.16}$$

The solution to the ODE

$$(\tfrac{d}{dt} - \kappa\phi(1-\mathfrak{q}_t)^2)\omega(t,1) + e^{-1}\lambda e^{-\kappa\phi\int_t^T \mathfrak{q}_u^2 du} = 0$$

can be found via a finite difference method to give the optimal value of $\omega(t,1)$ in the continuation region. From this the optimal time to execute a market order can be determined via solving for the value of t that sets $e^{-\kappa\xi}e^{-\kappa\phi\int_t^T \mathfrak{q}_u^2 du} = \omega(t,1)$. In the case when $q=2$ the QVI becomes

$$0 = \max\Bigg\{(\tfrac{d}{dt} - \kappa\phi(2-\mathfrak{q}_t)^2)\omega(t,2) + e^{-1}\lambda\omega(t,1)\,; \\ e^{-\kappa\xi}\omega(t,1) - \omega(t,2)\Bigg\}. \tag{5.17}$$

Note that $\omega(t,1)$ appears in the QVI and so the numerical solution for $\omega(t,1)$ will be fed into (5.17) in order to solve for $\omega(t,2)$ numerically.

The approach taken to solve (5.13) written in pseudo code is as follows:

1. The time over which the inventory must be liquidated, T, is discretised via a uniform grid with intervals Δt. The i^{th} element in this grid is labeled as t_i.

2. The terminal condition, $\omega(T,q) = e^{-\kappa q\xi}$, and boundary condition, $\omega(t,0) = e^{-\kappa\phi\int_t^T \mathfrak{q}_u du}$, are set.

3. A loop with $N = T/\Delta t$ steps is started. This loop begins at the final time step N

and ends at 0 decreasing by one unit each time. For each element within this loop, i.e. $\forall i \in [0, N] \cap \mathbf{N}_0$, the following procedures occur:

3.1. The forward Euler method is implemented:

$$\omega(t_{i-1}, q) = \omega(t_i, q) - \Delta t \, \frac{d}{dt_i}\omega(t_i, q)$$

where

$$\frac{d}{dt_i}\omega(t_i, q) = \kappa\phi(q - \mathfrak{q}_u)^2\omega(t_i, q) - e^{-1}\lambda\omega(t_i, q).$$

3.2. A check is run to see if a market order should be executed, i.e. if we are within the stopping region, specifically if

$$e^{-\kappa\xi}\,\omega(t_i, q-1) > \omega(t_i, q).$$

This is done so that a record of when market orders are issued can be produced.

3.3. $\omega(t_{i-1})$ will then be equal to the maximum of $\omega(t_{i-1})$ produced by the forward Euler approach and $e^{-\kappa\xi}\,\omega(t_i, q-1)$.

5.4.1 Parameterisation

In order to solve for $\omega(t, q)$ the entire model must be parametised. The following model parameters are chosen:

$$T = 60sec, \quad \eta = 10, \quad \lambda = 50/min, \quad \kappa = 100, \quad S_0 = \$1.00, \quad \sigma = \$0.01,$$
$$\xi = 0.01, \quad \phi = 10^{-3}, \quad N\Delta t = 6000$$

Where $N\delta t$ is the number of time steps used in the finite difference scheme. The deterministic target schedule chosen is the Almgren-Chriss (AC) trading schedule with temporary and permanent market impact [4]. As mentioned in the introduction this is a trade schedule popular in academic literature and often seen as a benchmark to beat. Specifically, the AC trading schedule seeks to maximise profit via liquidation using purely market orders when

these market orders have both a temporary and permanent impact on price.

$$\mathfrak{q}_t = \frac{\zeta e^{\gamma\,(T-t)} - e^{-\gamma\,(T-t)}}{\zeta e^{\gamma\,T} - e^{-\gamma\,T}}\eta,$$

where

$$\gamma = \sqrt{\frac{\tilde{\phi}}{k}} \quad \textit{and} \quad \zeta = \frac{\alpha - \frac{1}{2}b + \sqrt{k\tilde{\phi}}}{\alpha - \frac{1}{2}b - \sqrt{k\tilde{\phi}}}$$

with

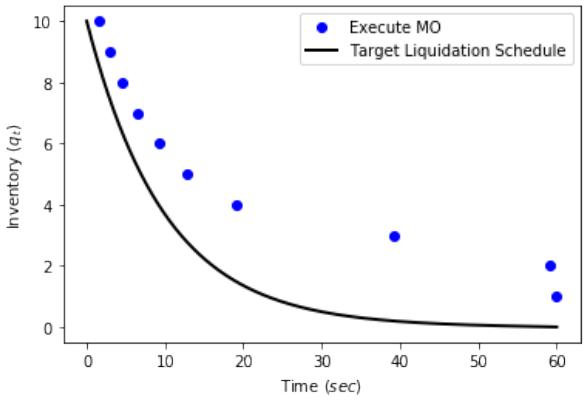
$$k = 0.001, \quad \tilde{\phi} = 10^{-5} \quad b = 0, \quad \alpha = +\infty.$$

A variety of plots are produced to investigate the outcome of the numerical solution under this parametisation are shown below.

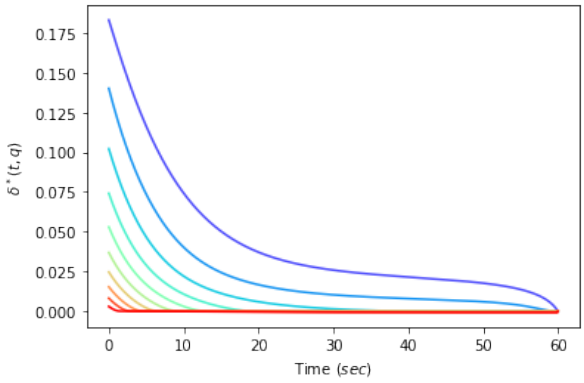
Figure 1a) gives a plot of the AC target schedule in black and the optimal time to issue a market order given your current stock of inventory. It can be seen that the optimal time to execute a market order is a decreasing function of the units of inventory held. Figure 1b) shows how the optimal depth δ^* varies with time. The different curves show δ^* for differing inventory where the inventory deceases as the curve move away from the origin, i.e. the dark blue curve is when $q = 1$ slightly lighter blue when $q = 2$ e.c.t. 1c) shows how the optimal time to execute a market order varies as ϕ, the coefficient on the deviation of the inventory from the target schedule, decreases. Specifically as the points go from yellow to pink ϕ decreases. It is clear as ϕ decreases the optimal time to execute a market order peels away from the target schedule. With the limit when $\phi = 0$ being when all market orders would be executed at T. Figure 1d) is generated by producing 1000 simulations of events while performing the optimal policy, optimal depths and times to issue market orders. 1d) shows the average inventory trajectory and compares it to the AC trading schedule. Here the heat map shows the relative frequency of inventory trajectories across all samples. Three sample paths are selected from these 1000 simulations and shown in Figure 2. These plots are relatively self explanatory. It is worth noting thought that the jumps in the sample paths

Figure 1

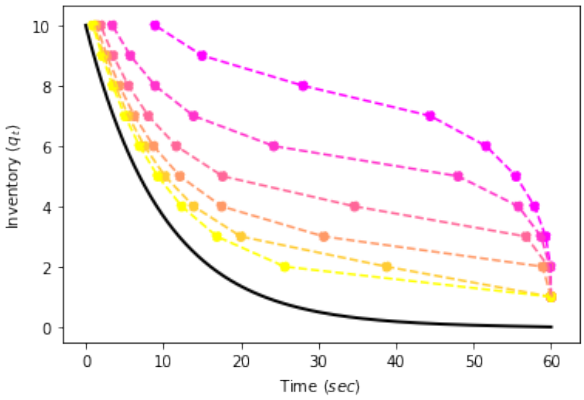
(a) Market Orders vs AC Trading Schedule



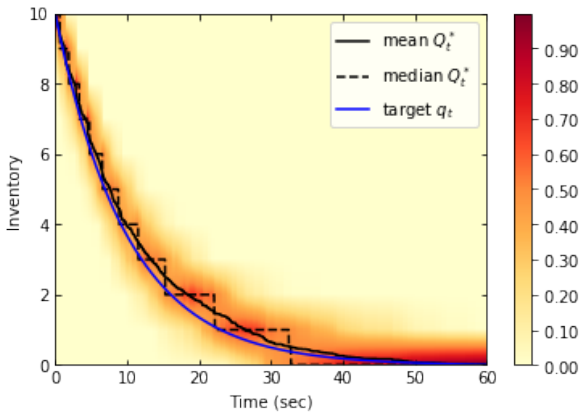
(b) Variation of Optimal Delta with Inventory Levels



(c) Impact of ϕ on market orders



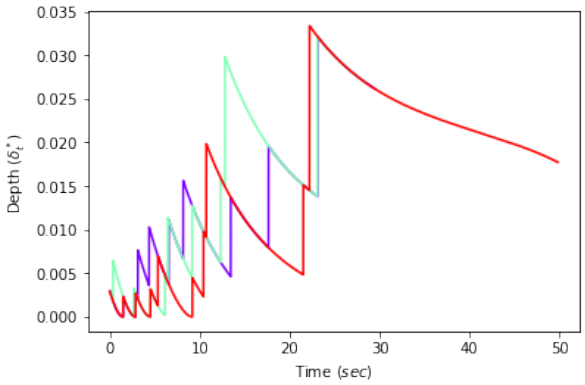
(d) Heat map of Inventory Trajectory



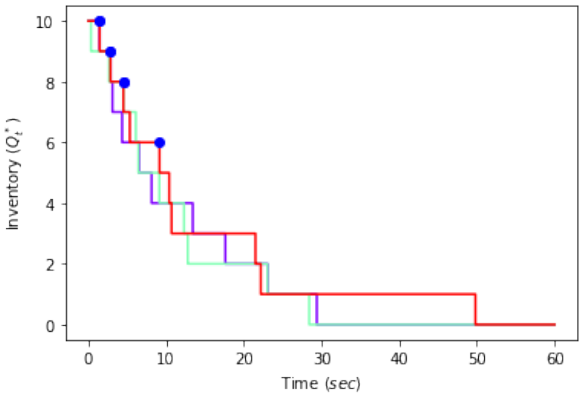
in figure 2a) occur when a unit of inventory is liquidated and the blue dots indicate where market orders are issued. As such when $\delta^*(t,q)$ transitions to $\delta^*(t,q-1)$ it jumps. This can be seen explicitly in 1b) where $\delta^*(q,t) \leq \delta^*(q-1,t)$ Furthermore, in 2d the sample paths flat line when there is no inventory left. Again this makes sense if we recall the anstaz given in (5.6). Specifically, the value function is driven by the wealth process book value of remaining inventory and the value given by optimally executing the remaining inventory. As such when there is no inventory left the value function will just be given by the wealth process. This will in turn be a constant as the wealth process can only change if there is inventory to liquidate. Furthermore, when $t = 0$ the value function of all sample paths will be the same with a value of 10.041. This is due to the wealth process for all sample paths being zero. The book value for all sample paths being \$10 and the value of optimal

Figure 2

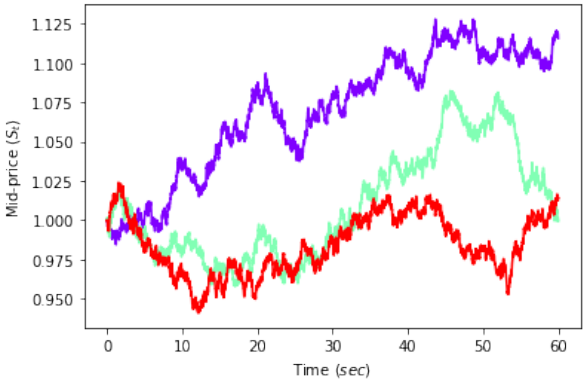
(a) Delta Sample Paths



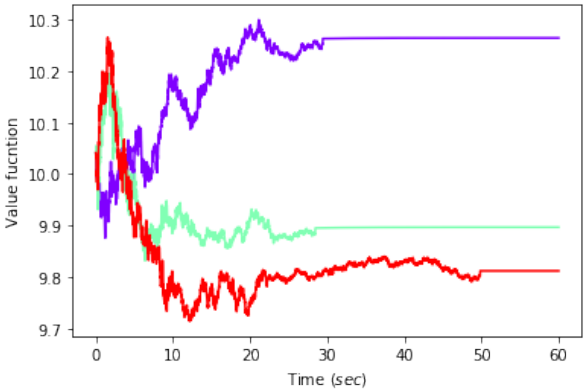
(b) Inventory Sample Paths



(c) Mid Price Sample Paths



(d) Value Function Sample Paths



execution being the same for all sample paths. The value functions for the sample paths only diverge due to the stochastic nature of the wealth process and the mid-price of the asset.

Another interesting question may be the cost savings associated with the optimal policy. Specifically, if we assume that had the agent not followed the policy given here they would have instead liquidated their inventory uniformly over the trading horizon using market orders. As such the inventory would have been liquidated at the time weighted average price (TWAP)

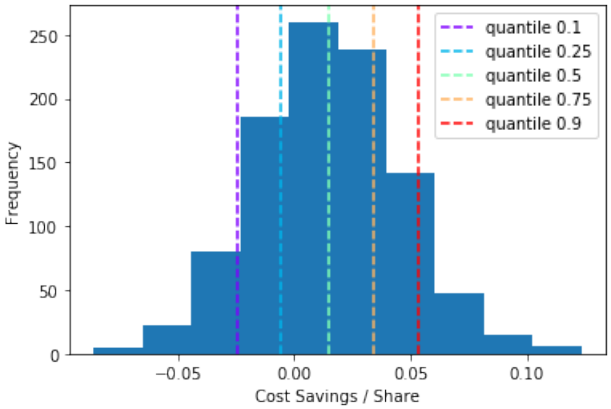


Figure 3: Cost savings by using the optimal policy

minus the mid price, ξ. Here TWAP is given by:

$$TWAP(t,T) = \int_t^T S_u \, du$$

The mean cost savings per share of following the optimal policy rather than liquidating uniformly with market orders over the 1000 sample paths is then \$0.0148. A histogram of the cost savings is then shown in Figure 3. Here the distribution of cost savings is relatively symmetric with perhaps a slight positive skew.

5.5 Limiting Properties

A natural follow up question may be what happens to $\omega(t,q)$ as the units of inventory to liquidate approaches infinity, i.e as $\eta \to \infty$. In order to explore this idea the optimal liquidation problem must be formulated in a slightly different manner. Instead of the agent liquidating a single unit of inventory they will liquidate a fixed proportion of their total inventory. The following new notation is defined:

$P_t^\delta = \frac{Q_t^{\boldsymbol{\tau},\delta}}{\eta} = 1 - \frac{N_t^\delta - F_t}{\eta}$ - The proportion of remaining inventory to be sold.

Δ_η - The proportion of inventory the agent chooses to liquidate.

Conditional on $P_{t_-}^\delta = p$ the performance criteria (5.2) can then be re scaled as follows:

$$\begin{aligned} R_\eta^{\boldsymbol{\tau}}(t,x,S,p,\delta) =& \frac{H^{\boldsymbol{\tau}}(t,x,S,q,\delta)}{\eta} \\ =& \mathbb{E}_{t,x,S,q}\left[\frac{X_\theta^{\boldsymbol{\tau},\delta}}{\eta} + \frac{Q_\theta^{\boldsymbol{\tau},\delta}}{\eta}(S_\theta - \xi) - \frac{\phi}{\eta}\int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_\mathfrak{u})^2 du\right] \\ =& \mathbb{E}_{t,x,S,p}\left[\frac{X_\theta^{\boldsymbol{\tau},\delta}}{\eta} + P_\theta^{\boldsymbol{\tau},\delta}(S_\theta - \xi) - \frac{\phi}{\eta}\int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_\mathfrak{u})^2 du\right]. \end{aligned} \tag{5.18}$$

The value function is then trivially

$$R_\eta(t,x,S,p) = \sup_{(\boldsymbol{\tau},\delta)\in\mathcal{A}_{[t,T]}} R_\eta^{\boldsymbol{\tau}}(t,x,S,p,\delta). \tag{5.19}$$

Where $\mathcal{A}_{[t,T]}$ denotes the admissible depths and set of admissible stopping times at time t.

The admissible depths are given by $\delta_t \in [0, S_t]$ and each $\tau_k \in \boldsymbol{\tau}$ is given by $\tau_k \in \{s \in [t,T] : Q_s^\delta > 0\}$.The following Theorem can then be derived:

Theorem 5.1 *Under the assumption R_k is once differentiable in t, X, Q and twice differentiable in S $\forall k \in \mathbb{N}_+$ and satisfies the form given in (5.18) and (5.19). The limit of the sequence $R_1, R_2, ..., R_k$ as $k \to \infty$ satisfies the following quasi variational inequality*

$$0 = \max\Big\{\partial_t \rho_\infty(t,p) + e^{-1}\lambda\,\rho_\infty(t,p)\}\,; \\ (\kappa\,\xi + \partial_p)\rho_\infty(t,p)\Big\}.$$

with terminal and boundary conditions

$$\rho_\infty(t,0) = \lim_{\eta\to\infty} e^{-\frac{\kappa\phi}{\eta}\int_t^T \mathfrak{q}_u^2 du} = 0$$
$$\rho_\infty(T,p) = e^{-\kappa\,p\,\xi}.$$

where

$$\rho_\infty(t,p) = \lim_{\eta\to\infty} \rho_\eta(t,p)$$
$$\rho_\eta(t,p) = \kappa e^{r_\eta(t,p)}$$
$$r_\eta(t,p) = R_\eta(t,x,S,p) - \tfrac{x}{\eta} - pS$$

Proof : Given the value function in (5.19) the following QVI can be derived by applying Theorem (4.4):

$$0 = \max\Big\{\partial_t R_\eta + \tfrac{1}{2}\sigma\partial_{SS}R_\eta - \tfrac{\phi}{\eta}(Q_t - \mathfrak{q}_t)^2 \\ + \sup_{\delta\in\mathcal{A}_{[t,T]}} \lambda\, e^{-\kappa\,\delta}\Big[R_\eta(t, x+\Delta_\eta\eta(S+\delta), S, p-\Delta_\eta) - R_\eta(t,x,S,p)\Big]; \\ \Big[R_\eta(t, x+\Delta_\eta\eta(S-\xi), S, p-\Delta_\eta) - R_\eta(t,x,S,p)\Big]\Big\} \tag{5.20}$$

Where R_η refers to $R_\eta(t,x,S,p)$ with boundary and terminal conditions

$$\begin{aligned} R_\eta(t,x,S,0) &= \frac{x}{\eta} - \frac{\phi}{\eta}\int_t^T \mathfrak{q}^2 du \\ R_\eta(T,x,S,p) &= \frac{x}{\eta} + p(S-\xi). \end{aligned} \tag{5.21}$$

The ansatz (5.6) can then be scaled and written as

$$R_\eta(t,x,S,p) = \tfrac{x}{\eta} + pS + r_\eta(t,p). \tag{5.22}$$

which when substituted into (5.20) will then give

$$\begin{aligned} 0 = \max\Bigg\{ \partial_t r_\eta - \frac{\phi}{\eta}(q-\mathfrak{q}_\mathfrak{t})^2 + \sup_{\delta\in\mathcal{A}_{[t,T]}} \lambda\, e^{-\kappa\,\delta}\Big[\Delta_\eta\delta + r_\eta(t,p-\Delta_\eta) - (t,p)\Big]; \\ - \Delta_\eta\xi + r_\eta(t,p-\Delta_\eta) - r_\eta(t,p)\Bigg\} \end{aligned} \tag{5.23}$$

where r_η refers to $r_\eta(t,x,S,p)$. Following the same procedure in section 5.3 the optimal depth and timing of market orders can be given in feedback form as

$$\delta^* = \frac{\frac{1}{\kappa} + [r_\eta(t,p) - r_\eta(t,p-\Delta_\eta)]}{\Delta_\eta} \tag{5.24}$$

and

$$r_\eta(\tau_p, p-\Delta_\eta) - r_\eta(\tau_p,p) = \Delta_\eta\xi \tag{5.25}$$

substituting (5.24) into (5.23) then gives

$$\begin{aligned} 0 = \max\Bigg\{ \partial_t r_\eta - \frac{\phi}{\eta}(q-\mathfrak{q}_\mathfrak{t})^2 + \tfrac{e^{-1}\lambda}{\kappa}\, e^{-\kappa\,[r_\eta(t,p) - r_\eta(t,p-\Delta_\eta)]}; \\ - \Delta_\eta\xi + r_\eta(t,p-\Delta_\eta) - r_\eta(t,p)\Bigg\} \end{aligned} \tag{5.26}$$

using the transformation

$$r_\eta(t,p) = \tfrac{1}{\kappa}\,\log\rho_\eta(t,p). \tag{5.27}$$

and following the same procedure in section 5.3 the QVI given in (5.26) now takes the following form:

$$0 = \max\Big\{(\partial_t - \frac{\kappa\phi}{\eta}(q-\mathfrak{q}_t)^2)\rho_\eta(t,p) + e^{-1}\lambda\,\rho_\eta(t,p-\Delta_\eta)\,; \\ e^{-\kappa\Delta_\eta\,\xi}\rho_\eta(t,p-\Delta_\eta) - \rho_\eta(t,p)\Big\} \tag{5.28}$$

with boundary and terminal conditions

$$\rho_\eta(t,0) = e^{-\frac{\kappa\phi}{\eta}\int_t^T \mathfrak{q}_u^2 du} \\ \rho_\eta(T,p) = e^{-\kappa\,p\,\xi}. \tag{5.29}$$

Using a first order Taylor expansion of $e^{-\kappa\Delta_\eta\,\xi}$ about 0 and taking the limit as $\eta\to\infty$ and $\Delta_\eta \to 0_+$ gives

$$0 = \max\Big\{\lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\{(\partial_t - \frac{\kappa\phi}{\eta}(q-\mathfrak{q}_t)^2)\rho_\eta(t,p) + e^{-1}\lambda\,\rho_\eta(t,p-\Delta_\eta)\}\,; \\ \lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\{(1-\kappa\Delta_\eta\,\xi)\rho_\eta(t,p-\Delta_\eta) - \rho_\eta(t,p)\}\Big\} \tag{5.30}$$

Noting that

$$\rho_\infty(t,p) = \lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\rho_\eta(t,p-\Delta_\eta) \\ \lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\frac{\kappa\phi}{\eta}(q-\mathfrak{q}_t)^2 = 0.$$

In addition dividing the component of equation (5.30) that corresponds to the stopping region by Δ_η and and taking the limit with respect to Δ_η and η then gives

$$0 = \lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\{(1-\kappa\Delta_\eta\,\xi)\rho_\eta(t,p-\Delta_\eta) - \rho_\eta(t,p)\} = \lim_{\substack{\Delta_\eta\to 0_+\\ \eta\to\infty}}\Big\{-\kappa\,\xi\rho_\eta(t,p-\Delta_\eta) \\ + \frac{\omega(t,p-\Delta_\eta)-\omega(t,p)}{\Delta_\eta}\Big\} = -\kappa\,\xi\rho_\infty(t,p) - \partial_p\rho_\infty(t,p).$$

As such the QVI in (5.30) becomes

$$0 = \max\Bigg\{\partial_t \rho_\infty(t,p) + e^{-1}\lambda\,\rho_\infty(t,p)\}\,; \\ (\kappa\,\xi + \partial_p)\rho_\infty(t,p)\Bigg\}. \tag{5.31}$$

The terminal and boundary conditions are then derived by noting that $\lim_{\substack{\Delta_\eta \to 0_+ \\ \eta \to \infty}} \mathfrak{q}_\mathfrak{t} = \infty$ as such

$$\begin{aligned} \rho_\infty(t,0) &= \lim_{\substack{\Delta_\eta \to 0_+ \\ \eta \to \infty}} e^{-\frac{\kappa\phi}{\eta}\int_t^T \mathfrak{q}_u^2 du} = 0 \\ \rho_\infty(T,p) &= e^{-\kappa\,p\,\xi}. \end{aligned} \tag{5.32}$$

6 Machine Learning Approach to Optimal Liquidation

Within machine learning there can be considered three main sub sections. Supervised learning which deals with labeled data, e.g. neural networks. Unsupervised learning which focuses on non labeled data, e.g. clustering algorithms. Finally, reinforcement learning which we will focus on exclusively here. Reinforcement learning refers to the group of algorithms which describe how an agent should interact with an environment in order to maximise a reward. With one of the most comprehensive collection of these algorithms being given by Barto and Sutton in their book "Reinforcement Learning: An Introduction" [20]. It is trivial to see that the task of optimal liquidation stipulated in this dissertation is indeed a reinforcement learning problem. Specifically, here our agent is an individual with a set amount of inventory to liquidate. The environment is defined by the market dynamics given in section (5.1) and the reward is given by the values within the expectation of (5.2) i.e.

$$X_\theta^{\boldsymbol{\tau},\delta} + Q_\theta^{\boldsymbol{\tau},\delta}(S_\theta - \xi) - \phi \int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_u)^2 du. \tag{6.1}$$

Recalling that $\theta = T \wedge \inf\{t : Q_t^{\boldsymbol{\tau},\delta} = 0\}$ is a stopping time that ensures that should no inventory be left no more market or limit orders will be executed.

6.1 Rational for the use of Machine Learning

In section 5 the quasi variational inequality has been solved in order to determine the value function (5.3) and optimal policy i.e. combination of depths to post limit orders and times to execute market order, δ and $\boldsymbol{\tau}$. Recall that the reduced QVI has the following form:

$$0 = \max\Big\{ (\tfrac{d}{dt} - \kappa\phi(q - \mathfrak{q}_t)^2)\omega(t,q) + e^{-1}\lambda\omega(t,q-1)\,; \\ e^{-\kappa\xi}\omega(t,q-1) - \omega(t,q) \Big\}$$

and is solved via implementing a forward Euler finite difference scheme, as explained in section 5.4. However, when implementing the forward Euler method the agent must know

both κ and λ. Where λ is a parameter determining the rate at which market buy orders occur and κ is a parameter determining how deep into the order book these same market orders will reach, see section 5.1 for more detail. In reality neither of these parameters will be known to the agent. As such at the very least the agent would have to approximate these parameters from sampled data. Furthermore, κ and λ only appear in the QVI due to the probability distributions assigned to the market dynamics in section 5.1. Specifically, recall that

> $M = (M_t)_{0 \leq t \leq T}$ - A Poisson process, with intensity λ which corresponds to the number of market buy orders that have occurred.
>
> $P(\delta) = exp(-\kappa\delta)$, $\kappa > 0$. - The probability that a limit order of depth δ is filled conditional on a market buy order arriving.

In practicality actual market dynamics will likely not be perfectly represented by this choice of stochastic process and probability distribution. At the very least the parameters λ and κ will likely be stochastic processes that vary with time. Which will greatly complicate the procedure of approximating these values. In a worst case scenario M will be an arbitrary stochastic process and $P(\delta)$ will be given by an arbitrary probability distribution. If this is the case the structure of the QVI completely breaks down. As such an alternative method to finding an optimal liquidation policy which does not require knowledge of the market dynamics will be discussed. Specifically, this approach will use a Monte Carlo method implemented within the framework of a machine learning technique.

6.2 Monte Carlo Method

The notation used to formulate the optimal liquidation problem in the context of the Monte Carlo method will now be discussed. This will build off the same notation given in section 5.1.

State Action Space:

Within the Monte Carlo problem the state space the agent occupies is defined by the follow-

ing two variables:

t - The time remaining to liquidate the inventory.

Q - The remaining inventory to liquidate.

The space of possible actions the agent can then take, with a slight abuse of notation, is written as:

$\delta_t(t, Q)$ - The depth to post a limit order at time t when there is Q units of inventory left.

$\tau(t, Q)$ - A stopping time representing the time at which an agent executes a market order.

These actions can be thought of as mapping the state space to the action space.

Policy Space:

The possible policies the agent can take are then characterised by the following process and set of stopping times.

$\delta = (\delta_t(t, Q))_{0 \leq t \leq T}$ - A sequence representing the depths at which a limit sell order is posted at each time i.e. the agent will post a sell limit order at time t at $S_t + \delta_t$.

$\boldsymbol{\tau} = \{\tau_k(t, Q) : 1 \leq k \leq K\}$ - An increasing sequence of stopping times at which the agent executes a market order, where $K \leq \eta$. Recall each stopping time represents the time to execute a single market order for a single unit of inventory. Note also that η is the total number of units of inventory to liquidate.

A policy will then be given by the tuple:

$$\pi = (\delta, \boldsymbol{\tau}) \tag{6.2}$$

Objective Function:

Now that the state action and policy spaces have been correctly defined the question of how

to evaluate a policy must be answered. The performance of any given policy is identical to that in section 5.2 bar a slight change in notation in order to be consistent with the reinforcement learning literature. Specifically,

$$H^{\pi}(t,x,S,q) = \mathbb{E}_{t,x,S,q}\left[X_{\theta}^{\boldsymbol{\tau},\delta} + Q_{\theta}^{\boldsymbol{\tau},\delta}(S_{\theta} - \xi) - \phi\int_{t}^{T}(Q_{\theta}^{\boldsymbol{\tau},\delta} - \mathfrak{q}_{\mathfrak{u}})^{2}du\right]. \tag{6.3}$$

and the agent is seeking to find an optimal policy which maximises the performance criteria such that

$$H(t,x,S,q) = \sup_{\pi\in\mathcal{A}} H^{\pi}(t,x,S,q). \tag{6.4}$$

Where $\mathcal{A}$ denotes the various parings of admissible depths and sets of admissible stopping times. The admissible depths are given by $\delta_t \in [0, S_t]$ and each $\tau_k \in \boldsymbol{\tau}$ is given by $\tau_k \in \{s \in [t,T] : Q_s^{\delta} > 0\}$. As in section 5 equation (6.4) is known as the value function.

6.2.1 Discretisation of State and Action Spaces

The policy space will be infinitely large due to the time component, t, of the state space and the depth component δ_t being continuous. In order to iteratively determine an optimal policy two approaches are typically used. Either the performance criteria given in (6.3) is considered to be differentiable with respect to the policy choice. A gradient ascent algorithm can then be implemented. Alternatively, if the differentiability assumption cannot be made then the continuous components of the state and action spaces can be discretised. This will ensure the policy space is finite. At which point a variation of a random search algorithm can be used. The key downside of this approach being the greater computational intensity. The approach used in this dissertation will be the latter.

Discretisation of State Space:

Discretising the state space is relatively simplistic. The units of inventory Q_t are by definition discrete. Discretisation of the time variable is implemented by splitting the interval $[0,T]$ using a uniform grid with k points, where k is a hyper parameter. This will then result

in the discrete set $t \in [0 = t_1..t_n..t_k = T]$ with $\Delta t = t_n - t_{n-1}$.

Discretisation of Action Space:

The discretisation of the action space is slightly more complex. The number of market sell orders the agent executes, given by $\sum_{\tau \in \boldsymbol{\tau}} \mathbb{1}_{\tau=t} \in [0..Q_t]$ is by definition discrete. However, the depth at which a limit order can be posted, δ_t, is continuous. Assuming that the state space has already been discretised δ_t can then be discretised via the following procedure:

1. A data set corresponding to how far into the limit orders book historical market buy orders reach over each time interval is generated. This is represented by

   $\tilde{\delta} = \tilde{\delta}_{t \in [1,..,k]}$ - A process representing the historical depths reached over the liquidation period

   A key assumption here being that only 1 market buy order may occur over each time interval.

2. The range of possible values the depths can reach is then split into n different sections, where n is a hyper parameter. This split is chosen such that the frequency of observed market orders at each of these depths is the same for each section. Mathematically, this can be expressed as

$$\left\{ \tilde{\delta}^i \in \delta : |\{\tilde{\delta} \in [\tilde{\delta}^i - \tilde{\delta}^{i+1}]\}| = |\{\tilde{\delta} \in [\tilde{\delta}^{i+1} - \tilde{\delta}^{i+2}]\}| \ \forall \ 1 \leq i \leq n-3 \right\}$$

   where $\tilde{\delta}^i$ marks the i^{th} division which separates $\tilde{\delta}$ into n different sections.

3. The mid point of each of these sections is then taken. This will be the depths the agent may post their limit sell orders and will be denoted by

$$\delta_t = \left\{ \delta^i : \delta^i = \frac{\tilde{\delta}^i + \tilde{\delta}^{i+1}}{2} \ \forall \ 1 \leq i \leq n-1 \right\}$$

   corresponds to the midpoint of the k^{th} section.

A graphical representation of these midpoints can be seen bellow.

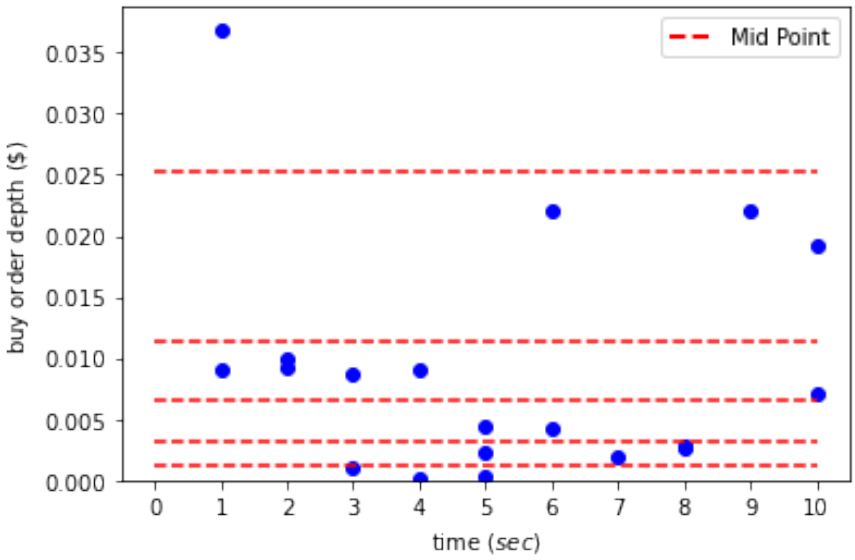


Figure 4: Simulated market buy orders

6.2.2 Policy Selection

Now the policy space has been successfully specified the question regarding how to correctly identify the optimal policy and get an accurate estimation of the value function can be tackled. Specifically, this problem revolves around balancing exploration and exploitation. This refers to the trade off of wanting to ensure that you get the most accurate estimate of the value function while simultaneously wanting to ensure the optimum policy has been correctly identified. In order to get the most accurate estimate of value function the optimum policy must be run the maximum number of times. However, in order to ensure that the policy identified as optimum is indeed optimum all possible policies must be searched a maximum number of times. This is done to ensure that the approximated value of the performance criteria for each policy is as close to the true value as possible. As such this will decrease the probability of a non optimal policy being incorrectly identified as optimal. Note both of these results follow from the central limit theorem.

The approach taken to balance this trade off is a variation of an algorithm known as ϵ-

greedy policy selection. Specifically, this operates as follows:

$$\mathbb{P}\Big(\pi^{\epsilon}(t,Q_t)=\pi(t,Q_t)\Big)=1-\epsilon+\tfrac{1}{|\delta_t|Q_t}\quad \mathbb{P}\Big(\pi^{\epsilon}(t,Q_t)\neq\pi(t,Q_t)\Big)=\epsilon-\tfrac{1}{|\delta_t|Q_t}. \tag{6.5}$$

Where π refers to an arbitrary policy, π^{ϵ} refers to the ϵ-greedy policy and $\epsilon \in [0,1]$. $\pi^{\epsilon}(t,Q_t)$ refers to the actions the ϵ-greedy policy π^{ϵ} will take when in the state (t,Q_t) and $\pi(t,Q_t)$ is the actions the policy π takes in the state (t,Q_t). Intuitively, this approach to policy selection works by following the actions prescribed by the policy π with probability $1-\epsilon$ but deviating from these actions with probability ϵ. When the ϵ-greedy policy deviates from the pre determined actions a random action for that state will be chosen. The adjustment of $\frac{1}{|\delta_t|Q_t}$ accounts for the fact that when the ϵ-greedy policy deviates from the action given by π the same action prescribed by π will be selected with probability $\frac{1}{|\delta_t|Q_t}$.

Once the time interval over which liquidation must occur has been completed the reward (6.1) will be recorded. This will then be used to update the estimate of the performance criteria for the policy (6.3). The policy with the highest estimated policy performance will then be chosen to be run again. This process will be repeated for a fixed number of iterations.

6.2.3 Modification to Policy Exploration

Two modifications are made to the ϵ-greedy approach to policy selection in order to increase the exploration of the policy space. This is done in two separate ways:

Running each policy a minimum number of times:

This is done to ensure that every policy that is explored is done so to a minimum degree of accuracy. This will lower the probability of the optimal policy being identified but incorrectly labeled as sub optimal. This would occur if simply by chance for a given run of the optimal policy a low reward was found. Due to the design of the ϵ-greedy policy this would result in the optimal policy never being being explored again.

Initialising more than one policy:

When implementing the ϵ-greedy policy an initial policy π^+ must be chosen. Other policies will then be explored by randomly branching from this initial policy. By initialising with more than one policy a lower bound will be placed on the number of policies that will be explored.

The key draw back of both these approaches is they increase the computational intensity of the algorithm. As it will increase the number of operations that must be completed.

6.3 Comparison of Monte Carlo Method with QVI Approach

A comparison of the Monte Carlo method and that given by solving the quasi variational inequality will now be given. When implementing the QVI method in order to solve for $\omega(t,q)$ the entire model must be parametised. As such the following model parameters are chosen:

$$T = 60sec, \quad \eta = 1, \quad \lambda = 50/min, \quad \kappa = 100, \quad S_0 = \$1.00, \quad \sigma = \$0.01,$$
$$\xi = 0.01, \quad \phi = 10^{-3}, \quad N\Delta t = 100000$$

recalling that $N\Delta t$ is the number of time steps used when solving for $\omega(t,q)$ using the forward Euler approach. The deterministic target schedule chosen is again the Almgren-Chriss (AC) trading schedule with temporary and permanent market impact.

$$\mathfrak{q}_t = \frac{\zeta e^{\gamma\,(T-t)} - e^{-\gamma\,(T-t)}}{\zeta e^{\gamma\,T} - e^{-\gamma\,T}}\eta,$$

where

$$\gamma = \sqrt{\frac{\tilde{\phi}}{k}} \quad and \quad \zeta = \frac{\alpha - \frac{1}{2}b + \sqrt{k\tilde{\phi}}}{\alpha - \frac{1}{2}b - \sqrt{k\tilde{\phi}}}$$

with

$$k = 0.001, \quad \tilde{\phi} = 10^{-5} \quad b = 0, \quad \alpha = +\infty.$$

The parameterisation of the Monte Carlo specific parameters is then:

$$\Delta t = 10sec, \quad |\delta_t| = 11, \quad NSIM = 5000, \quad N_{min} = 500 \quad \epsilon = 0.4.$$

Where Δt is the time interval over which a market buy order may arrive, i.e. at $t \in \{0, 10, 20, 30, 40, 50\}$ the agent may decide to change their limit order. $|\delta_t|$ represents the cardinality of δ_t and as such the number of possible depths the agent may post limit sell orders. In the case given, i.e. $|\delta_t| = 11$, the agent may post limit orders at 10 different depths or may not post a limit order at all. $NSIM$ represents the number of simulations. Where each simulation refers to a sample path of both mid point trajectory and the depth of market buy orders over the time interval from t = 0 to t = T. N_{min} refers to the minimum number of times each policy must be run.

6.3.1 Policy Space

The total policy space the Monte Carlo method searches over is extensive. Specifically, under the parametisation given in section 6.3, $\eta = 1$, the total number of possible policies are given by

$$|\mathcal{A}| = \sum_{i=0}^{\frac{T}{\Delta t}-1} (|\delta_t| - 1)^i. \tag{6.6}$$

Here each addend in the sequence corresponds to the policies available if a market order is issued at the i^{th} time step. For example $(\delta_t - 1)^0 = 1$ corresponds to the case where a market order is immediately executed. $(\delta_t - 1)^1$ corresponds to the case where 1 time step occurs and then a market order is executed e.c.t. Clearly, $(\delta_t - 1)$ gives the total number of depths limit sell orders can be posted at. Using this formula under the parametisation above $|\mathcal{A}| = 111111$.

Impact of η:

The equation given in (6.6) only holds in the case when a single unit of inventory must be liquidated, i.e. $\eta = 1$. If this is not the case the policy space becomes much larger as the

policy will not necessarily terminate if the agent issues a market order. In addition, the agent at each time step may now have the option to liquidate more than a single unit of inventory. Specifically, at any time period, t, the agent will be able to liquidate any remaining inventory, $\eta - Q_t$. The impact of both these factors is that a trivial close form solution to the number of possible policies similar to that given in equation (6.6) is not available. However a rough approximation can be given by:

$$|\mathcal{A}| = \sum_{i=0}^{\frac{T}{\Delta t}-1} \left[(|\delta_t| - 1)\eta\right]^i \tag{6.7}$$

though this will underestimate the space of possible policies.

6.3.2 Comparison of Optimal Limit Order Depths

Under this parametisation a comparison can be made between the optimal depth to place a limit sell order via solving the quasi variational inequality or using the Monte Carlo method. This can be seen in Figure 5 where the optimal depths given by the Monte Carlo method are shown by the blue dots and that given by the QVI equation by the solid line. Firstly, it is worth noting that both in the case of the Monte Carlo and QVI methods it would appear that the optimum depth is a decreasing function of time. However, there is clearly a sizable discrepancy between these methods. The reasons behind these discrepancies are discussed bellow.

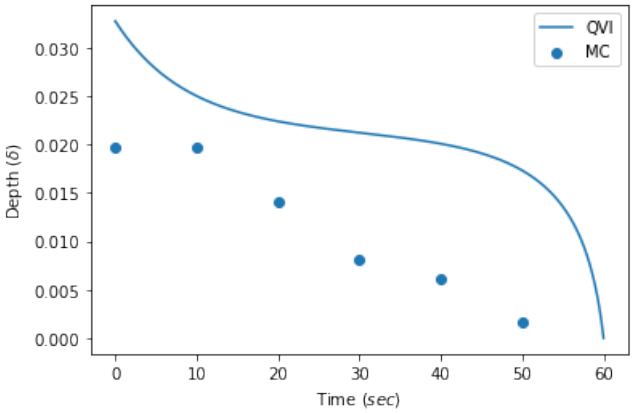


Figure 5: Optimal depth QVI vs MC approach

Discrepancies in Available Strategies:

The solution given by the QVI can be thought of as being optimal in two senses. Firstly it

is optimal in the sense that it maximises the value function given in (5.3). This property follows naturally from how the QVI is derived in section 4. However, the solution to the QVI can also be considered optimal in the sense that it gives the solution over the maximum number of strategies. This occurs due to the QVI being constructed in a continuous time context. As such δ_t can be continuously adjusted over the liquidation window. Furthermore, $\tau \in [0, T]$ i.e. a market order can be liquidated at any time during the liquidation window. In addition, δ_t is not discretised which allows for a far greater choice of possible depths to post a limit order. When looking at the Monte Carlo method due to the discretisation of t and δ_t the available strategies are far smaller than that given by the QVI. A key result of the QVI having a greater set of available strategies is the agent can be more aggressive with their choice of limit order depths. The logic behind this result being that the agent can continuously incrementally adjust their limit order if it is not filled rather than having to wait until 10 seconds has elapsed. This can be seen graphically in Figure 5 via the depths the QVI approach gives dominating those of the Monte Carlo method.

Discrepencies Caused by Errors:

The Monte Carlo method is also prone to two key types of errors. Those caused due to the discretisation of δ_t and those caused by incorrect optimal policy selection. Discretisation of δ_t means that there will always be a minimum error caused by the Monte Carlo method. Specifically, this will be given by $\arg\min_{\delta_t} |\delta_t - \delta_t^*|$. Where t represents each time at which a depths to place the limit order can be chosen and δ_t^* represents the optimal depth to place a limit sell order at time t. Even if δ_t^* did coincide with δ_t^{QVI}, where δ_t^{QVI} is the optimal depth given by the QVI method, this minimum error would still exist. Incorrect policy selection will occur if the set of possible policies is not properly explored. This can occur if simply due to chance the optimal policy is never discovered or if the optimum policy is discovered but the sample value function associated with this policy is recorded as being sub optimal. Indeed, the modifications given in section 6.2.3 are designed specifically to reduce the likelyhood of these errors occurring. Again if δ_t^* does coincide with δ_t^{QVI} but δ_t^{QVI} is not identified

as optimal then the optimal depth given by the Monte Carlo method will be different from that given by the QVI.

There is however a trade off between minimising the errors caused by discretising δ_t and the errors caused by sub optimal policy selection. Specifically, as the number of possible depths the agent can post limit orders increases so does the number of possible policies. As such in order to correctly identify the optimal policy the algorithm must search over a much larger space. In addition, the difference between the performance criteria in each of these policies will become smaller and smaller as the number of possible depths becomes larger. This is simply because the difference between possible depths become more granular and so only has a marginal impact on the performance criteria. This in turn means a greater number of simulations are required to differentiate between policies with similar depths.

Quantifying the discrepancies:

It is difficult to quantify the discrepancies between the QVI approach and Monte Carlo method. However, within the strategies uncovered by the Monte Carlo method a comparison can be made. Specifically, using a t-test a probability can be assigned that the optimal strategy is greater than any other strategy. Specifically, if one wanted to calculate the t-statistic for two policies π_1 and π_2

$$t = \frac{H^{\pi_1} - H^{\pi_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \qquad DF = \min(N_1, N_2) - 1. \tag{6.8}$$

Where H^{π_1} and H^{π_2} correspond to the sample estimates of the performance criteria under π_1 and π_2. s_1^2 and s_2^2 corresponds to the sample estimates of the variance of each policy. DF corresponds to the degrees of freedom and N_1 and N_2 the number of times each policy was run. Table 1 then lists the t-values and P-values of the 4 sub optimal policies with the highest performance criteria, i.e. closest to being optimal, and the TWAP when compared to the optimal strategy.

Policy	Performance Criteria	t-statistic	P-value
Optimal	1.00415	N/A	N/A
Sub optimal	1.00234	0.192	0.423
Sub optimal	1.00225	0.182	0.428
Sub optimal	1.00221	0.194	0.423
Sub optimal	1.00215	0.169	0.433
TWAP	0.995	1.219	0.111

Table 1: Comparison of Optimal & Sub Optimal Policies

Here the P-value has the interpretation of representing the probability that due the sub-optimal or TWAP strategy only produces a smaller performance criteria than the optimal approach due to chance. Here we see that it is impossible to differentiate between the optimal policy and any of the policies at a 5% level of significance. This in turn suggests that the we cannot differentiate between any of the policies selected and the optimal one. In order to do so additional simulations would need to be run.

6.3.3 Comparison of Effectiveness of QVI and Monte Carlo

Another natural follow up question may be to compare the effectiveness of the Monte Carlo approach relative to the QVI method. One obvious way to do so would be to compare the values of the value functions and the run times given by the Monte Carlo method and by solving the HJB.This comparison can be seen in Table 2.

Here we see that the value function of the solution given by the QVI is greater than that given by the Monte Carlo method. This makes sense as recalling the ansatz given in equation

	QVI	Monte Carlo	TWAP
S_0	1	1	1
$h(0,1)$	0.018	0.004	-.05
Value Function	1.018	1.004	0.995
Run Time	10.013	7018.906	N/A

Table 2: Comparison of QVI and Monte Carlo

(5.6),

$$H(t, x, S, q) = x + qS + h(t, q).$$

The performance criteria can be thought of as the value of the wealth process plus the book value of any remaining inventory plus the value added from being able to optimally liquidate the remaining inventory, given by $h(t, q)$. At $t = 0$, where the value function is evaluated in Table 2, note that there is only one unit of inventory which has a book value of \$1.00. As such we would expect the value function to be slightly above 1 for both methods. As noted earlier in this section, there are fewer strategies available to the Monte Carlo method. This in turn results in the added value the agent can generate by optimally liquidating the inventory to be less than that given when using the QVI approach. Unsurprisingly Table 2 also show that the Monte Carlo method is many multiples more computationally expensive than finding the optimal policy by solving the QVI equation, as shown by the much longer run time. It is also worth noting that both the QVI and Monte Carlo approach agree that a market order should only be executed at the terminal time i.e. at 60 seconds.

In order to compare the effectiveness of the Monte Carlo method to that of the QVI two approaches can be taken. A comparison between the value functions or the added value from liquidation $h(0, 1)$ can be made. When comparing the value functions the Monte Carlo method is seen to be 98.6% as effective as the QVI method. If a base case of liquidating at TWAP using market orders is taken then the Monte Carlo method is 100.9% as effective and the QVI approach is 102.3% effective. However, a large proportion of the value function is given by the book value of remaining inventory. As such it can make sense to compare the added value from optimally liquidating the portfolio, i.e. $h(0, 1)$. Here we see than the Monte Carlo method is 22.2% as effective as the QVI method.

A key problem with comparing the results from solving the QVI or the Monte Carlo technique is it is difficult to see if the discrepancies between optimal limit order depths occurs due to errors associated with the Monte Carlo technique or simply due to modeling discrepancies.

As such another sub problem which has a known analytical solution will be evaluated. This is done to show that the Monte Carlo approach can produce an optimum policy.

6.3.4 Analytical Evaluation of Loss:

Under the parametisation given in section 6.3 I now derive an analytical solution for the value functions and as such the loss associated with using the Monte Carlo method.

Specifically, recall the following from section 5.1 and 6.3:

1. The arrival of market buy orders is modelled via a Poisson process with intensity λ. As such the probability of a market buy order arriving over a time interval Δt is given by $\lambda \Delta t e^{-\lambda \Delta t}$.
2. The probability that a limit order of depth δ is filled conditional on a market buy order arriving is given by $P(\delta) = exp(-\kappa\delta),\ \kappa > 0$.
3. Under this parametisation the agent is only looking to liquidate a single unit of inventory.
4. For both the optimal liquidation schedules given by the Monte Carlo method a market order is only executed at the termination time, T.

As such the value function given by the Monte Carlo method can be derived analytically as:

$$
\begin{aligned}
&H = H^{\pi^{MC}} = \mathbb{E}_{t,x,S,q}\left[X_\theta^{\boldsymbol{\tau},\delta} + Q_\theta^{\boldsymbol{\tau},\delta}(S_\theta - \xi) - \phi \int_t^T (Q_\theta^{\boldsymbol{\tau},\delta} - \mathfrak{q}_\mathfrak{u})^2 du\right] = \\
&\sum_{i=1}^{T/\Delta t - 1} \left[\left(\prod_{j=1}^{i-1}(1 - \lambda \Delta t e^{-\lambda \Delta t} e^{-\kappa \delta_{t_j}})\right) \lambda \Delta t e^{-\lambda \Delta t} e^{-\kappa \delta_{t_i}} \left\{ S_0 + \delta_{t_i} - \phi\left(\sum_{k=1}^{i}(1 - \mathfrak{q}_{t_k})^2 + \sum_{m=i}^{T/\Delta t} \mathfrak{q}_{t_m}^2\right)\right\}\right] \\
&+ \left(\prod_{j=1}^{T/\Delta t - 1} (1 - \lambda \Delta t e^{-\lambda \Delta t} e^{-\kappa \delta_{t_j}})\right)\left\{ S_0 - \xi - \phi \sum_{i=1}^{T/\Delta t}(1 - \mathfrak{q}_{t_i})^2 \right\}
\end{aligned}
\tag{6.9}
$$

With the convention that $\prod_1^0 = 1$. Where $\theta = T \wedge \inf\{t : Q_t^{\boldsymbol{\tau},\delta} = 0\}$ is a stopping time which ensures that should no inventory be left no more market or limit orders will be executed. π^{MC}

corresponds to the optimal policy given by the Monte Carlo method and $\sum_{i=1}^{T/\Delta t-1}\big[(\prod_{j=1}^{i-1}(1-\lambda\Delta t e^{-\lambda\Delta t}e^{-\kappa\delta_{t_j}}))\lambda\Delta t e^{-\lambda\Delta t}e^{-\kappa\delta_{t_i}}\{S_0+\delta_{t_i}-\phi(\sum_{k=1}^{i}(1-\mathfrak{q}_{t_k})^2+\sum_{m=i}^{T/\Delta t}\mathfrak{q}_{t_m}^2)\}\big]$ corresponds to the impact on the value function from successfully liquidating using limit order. Unsurprisingly, $(\prod_{j=1}^{T/\Delta t-1}(1-\lambda\Delta t e^{-\lambda\Delta t}e^{-\kappa\delta_{t_j}}))\{S_0-\xi-\phi\sum_{i=1}^{T/\Delta t}(1-\mathfrak{q}_{t_i})^2$ corresponds to the impact on the value function from issuing market orders at the terminal liquidation time. Intuitively, equation (6.9) can be thought of the solution to the value function when the reward associated each possible state action pair is weighed by their corresponding probability.

The difference in the value functions can then be evaluated by evaluated by simply subtracting the result given by the optimal QVI policy from that given by equation (6.9).

7 Conclusion

Optimal portfolio liquidation is a highly studied area of quantitative finance due to both its mathematical complexity and the relevance of the problem to practitioners.

The literature so far has placed an emphasis of framing this problem as an optimal control or an optimal control and stopping problem. Under certain regulation constraints the corresponding Hamiltonian Jacobi Bellman equation or quasi variational equality can be solved for the optimal policy, see Huitema [7] or Cartea & Jaimungal [8]. This approach works well if the complete market dynamics are known. However, in practicality we will only be able to approximate the market dynamics via historical data.

Given the wealth of historical pricing data available on financial instruments it is highly probable that the topic of optimal liquidation could be tackled well by machine learning techniques.

This dissertation successfully formulates a market model and optimal liquidation problem. This problem is then solved for via formulating the corresponding QVI and solving for it numerically. The optimal depths to post limit orders and market orders are then examined in addition to the limiting properties as the the units of inventory approach infinity.

We then move on to apply machine learning to this problem. Specifically, a Monte Carlo method which uses a variation of a random search algorithm is implemented. This known as an epsilon greedy policy selection approach.

The key advantaged of this approach being that no assumptions are made regarding the dynamics of the market. However, these lack of assumptions come at a cost as the algorithm is highly computationally expensive.

Further research could still be explored on the effectiveness of this epsilon greed Monte Carlo method. Due to computational intensity the case of liquidating more than 1 unit of inventory are not examined. In order to examine if the Monte Carlo method is effective the code could either be run on a GPU or could be re written in a lower level language e.g. C++. This would be done purely to ensure that the code can run in a reasonable time frame.

References

[1] BIGIOTTI, ALESSANDRO, NAVARRA, ALFREDO, "Optimizing Automated Trading Systems", Advances in Intelligent Systems and Computing, Springer International Publishing, pp. 254–261

[2] Quantopian, https://www.quantopian.com.

[3] ALDRIDGE,High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems.

[4] ALMGREN, CHRISS, Optimal Execution of Portfolio Transactions.

[5] MARKOWITZ, Portfolio Selection (1952).

[6] ÁLVARO CARTEA, SEBASTIAN JAIMUNGAL, JOSE PENALVA. Algorithmic and High-Frequency Trading (2015). Cambridge University Press.

[7] ROBERT HUITEMA. Optimal Portfolio Execution using Market and Limit Orders (2012).

[8] ÁLVARO CARTEA, SEBASTIAN JAIMUNGAL. Optimal Execution with Limit and Market Orders (2014).

[9] YANG, CHING, GU SIU, Generalized Optimal Liquidation Problems Across Multiple Trading.

[10] CONT, KUKANOV, Optimal order placement in limit order markets

[11] LIN, "The New Financial Industry" (2014).

[12] CRISAF, MACRINA, Simultaneous Trading in 'Lit' and Dark Pools

[13] Kratz, SCHONEBORN, Optimal liquidation in dark pools.

[14] NIAM AKBARZEDEH, CEM TEKIN, MIHAELA VAN DER SCHAAR. Online Learning in Limit Order Book Trade Execution (2018).

[15] ÁLVARO CARTEA, EBASTIAN JAIMUNGAL, DAMIR KINZEBULATOVC. Algorithmic Trading with Learning (2015).

[16] Hendricks, Wilcox. A reinforcement learning extension to the Almgren-Chriss framework for optimal trade execution.

[17] WATKINS, Learning from Delayed Rewards.

[18] TOUZI, Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE.

[19] PHAM, Continuous-time Stochastic Control and Optimization with Financial Applications, Springer.

[20] BARTO, SUTTON, Reinforcement Learning: An Introduction.

[21] GROSSMAN, MILLER, Liquidity and Market Structure (1988).