The Stochastic Control of Liquidity Provision: Market Making with and without Last Look

submitted by

Kevin Olding

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

December 2022

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Candidate's signature ....

Kevin Olding

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, in collaboration with my supervisor, Alexander M. G. Cox, with the exception of the paper 'Using Echo State Networks to Approximate Value Functions for Control Problems' ([56]), which is a joint work with Hart, A.G., Cox, A.M.G., Isupova, O. and Dawes, J.H.P., and is presented in Chapter 6 along with its own statement of authorship.

Candidate's signature ....

Kevin Olding

Summary

In this thesis we study models for market making, including the case where trades are subject to a trade acceptance protocol, or ‘last look’ mechanism. We study existing models of market making, adding rigour and adapting them in new ways to the last look case. We also propose a number of novel continuous models for market making that may be considered natural extensions of their discrete counterparts, both with and without last look, which allow us to prove results via spectral theory about the long-run value of market making.

The final chapter includes the paper ‘Using Echo State Networks to Approximate Value Functions for Control Problems’ ([56]), the result of a collaborative project between the author, another PhD student and their supervisors undertaken during the PhD, which includes an application to the market making problem.

Acknowledgements

This PhD thesis was undertaken with funding from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), grant number EP/L015684/1.

I am grateful to the many friends and colleagues who have made studying in Bath a memorable and enjoyable experience, and in particular to Professor Alexander Cox for his supervision and support throughout my time here.

Contents

Chapter 1

Introduction

In this thesis we study models for market making, including the case where trades are subject to a trade acceptance protocol, or 'last look' mechanism. We study existing models of market making and adapt them to the last look case and also introduce and study new related models of our own.

Broadly speaking, this thesis sits in the field of 'market microstructure', a relatively young area of mathematical finance. Since the 1990s the proportion of orders in financial markets that are implemented in a discretionary and manual manner by human traders has decreased significantly. Financial markets are increasingly dominated instead by trading that is algorithmic and automated at high-frequency. This has led to a surge in interest in microstructure models both from practitioners and regulators and an ever increasing literature. See for example [23], [67], [34], [25] and [16] for recent texts giving a good overview of the range of techniques being employed. As computing power increases and machine learning approaches become ever more prevalent, a field of 'econopyhsics' has also emerged, combining a range of practical methodologies trained on and derived from deep insights from models in mathematics and physics.

Our approaches are mostly rooted in applications of stochastic control to market making problems. In Section 1.1 we begin by describing what is meant by market making and in Section 1.2 we describe what we mean by 'last look' or trade acceptance protocols and discuss why they are of interest. In Section 1.3 we take a brief tour of the most relevant existing literature and in Section 1.4 we set out some mathematical preliminaries. Finally, in Section 1.5 we give an overview of the rest of this thesis.

1.1 Market Making

This thesis is about market making and so we begin by attempting to define this term. As Guéant notes in [51] this is not as easy as it may first appear as a result of the electronification of markets and the emergence of high-frequency trading in many of them blurring the boundaries of different types of trading. So we begin by setting out an understanding of what we will mean by a market maker in the remainder of this thesis.

Market makers are financial agents who post limit orders in financial markets, indicating prices at which they would be willing to buy (bid) and sell (ask) specified quantities of an asset, currency or other financial product. They hope to profit by exploiting the difference between the bid and ask prices (the bid-ask spread) without looking to take any long term position in the product they are buying and selling. By posting passive limit orders they provide liquidity to a financial market, making prices available to liquidity taking market participants looking to buy or sell immediately. The profit they make from the round trip trades they are able to make may be seen as a return for taking the risk of adverse price movements, uncertain executions and the possibility of adverse selection in markets with informed traders ([64], [48]).

On some financial exchanges, the role of a market maker might be a defined one, where so called 'designated' market makers receive preferential transaction fees (or sometimes even rebates) in exchange for a promise to continuously provide a certain level of liquidity, regardless of underlying market conditions. Indeed this is certainly the traditional view of market making, but a significant and increasing proportion of liquidity in financial markets is now also provided by other agents who employ market making strategies. They provide liquidity, seeking to make a profit from their strategies without any formal obligation or agreement with the trading venue. Also, some market makers may offer one-to-one facilities to clients with buy and sell prices quoted. We will ignore this distinction and simply consider a market maker as any trader who is continuously providing limit orders on both the bid and ask side or the order book.

The market maker generally only places passive limit orders but may also have access to a stream of liquidity taking orders, which we will generally consider as originating as orders from clients of an investment bank with genuine exogenous demand for the product. In practice, however, it may be challenging or even impossible for the market maker to guarantee the type of counterparty they will trade with, and orders could also originate from other market makers or high frequency traders seeking short term profits.

The market maker's concerns may be complex, but there are two primary forces driving their decisions. Firstly, their problem is principally one of pricing the cost of holding inventory in an uncertain environment. The market maker is assumed to hold no underlying belief about the long term performance of the asset and is simply in the business of providing liquidity and profiting from the bid-ask spread. Whilst holding a long inventory the market maker fears decreases in the price of the asset, and they lose out from increases whilst they hold a short position. Their primary concern is to charge a sufficient premium in their bid-ask spread in order to cover these volatility risks.

A very significant secondary concern for the market maker is informational disadvantage. That is, they may be trading not with agents with an exogenous demand uncorrelated with future price moves, but with agents who hold insider knowledge about future price movements. In particular, in markets where high-frequency trading is common, there is a risk of trading with counterparties who have actual knowledge of price movements or demand over very short time horizons. This knowledge may potentially be legitimate, for example knowledge of their own planned future order flow, but it may also be more dubious or even illegal, for example relating to insider trading, illegitimately acquired information regarding news announcements or front-running client order flow amongst other possibilities.

The growth of high frequency trading and an increasing accessibility of the markets to independent and non-institutional traders has amplified these concerns. Without a mechanism to protect against this, market makers are vulnerable to a range of counterparties whose order flow they may consider toxic. Latency arbitrageurs are one such possible counterparty, who have the potential to significantly undermine the profitability of a market maker's business by accessing and trading on market updates at a very fast rate (perhaps only milliseconds). A market maker has two choices in how to deal with this. They must either increase their own spending on latency reducing technologies to reduce the informational disadvantage, or introduce delays and trade acceptance ('last look') protocols to minimise the impact of this disadvantage. The latter option has the advantage that it may also protect simultaneously against other forms of toxic order flow and is the choice we study in detail in Chapter 5 and so we introduce it in detail in the next section.

1.2 Last Look

A trade acceptance protocol, or 'last look' mechanism, is a process by which a market maker may decide to reject a trade after a certain hold time (typically between 20 and 200 milliseconds). This time allows the market maker to undertake various checks, which could include security checks and other internal protocols, and will typically also include a price update to check for any significant short term movements in price.

Mathematically, we will consider market making on a product whose mid-price is modelled by an arithmetic[1] Brownian motion

$$dS_t = \sigma dW_t.$$

Then our last look condition will be that if an order is submitted at time t it will be confirmed at time $t + \delta t$ (at a price quoted at time t) if

$$-\xi_1 < S_{t+\delta t} - S_t < \xi_2,$$

and cancelled otherwise. In practice the last look horizon δt and the rejection

[1]The choice of arithmetic as opposed to geometric Brownian motion is motivated by mathematical simplicity, and justified by the short time horizons we will consider.

boundaries ξ_1 and ξ_2 are to be agreed between the market maker and the client and may be included in the terms and conditions of the market making contract. In practice levels of transparency over the details of trade acceptance criteria vary between liquidity providers, as does the length of time δt between order and confirmation or rejection ([87], [69]). It may also be that either ξ_1 or ξ_2 could be set to ∞ in which case the trade is unconditional in one or both directions. The case $\xi_1 = \xi_2 = \infty$ would correspond to an entirely unconditional trade with no last look criterion. The papers of [78], [79] and [20] have studied last look models from a mathematical perspective, though last look models do not seem to have attracted a very large academic attention in general and so the approach we take to them in Chapter 5 is a novel one.

Another discrepancy between market makers is whether they offer last look facilities which are 'symmetric' or 'asymmetric'. The term 'asymmetric' here typically implies that the last look feature always acts to protect the market maker against adverse price moves, but not the liquidity taking client. That is, the trade acceptance criteria may be set to cancel trades on the ask side only when the price increases over the last look window and on the bid side when it decreases.

In a 'symmetric' last look facility, trades would be cancelled when the market moves significantly in either direction during the last look window, thus providing protection to both parties. It should be noted that many other interpretations of symmetry are possible and Oomen ([79]) considers various such interpretations. The simplest definition is to take $\xi_1 = \xi_2$ (both on the bid and ask side), and this is generally how we will proceed here. Other interpretations of symmetry might require $\xi_1 \neq \xi_2$ or different conditions on the bid and ask side, for example if the aim is to balance reject rates in each direction, or the overall cost of rejections to each party. According to a recent survey ([87]), of the top 50 liquidity providers, 32 currently offer symmetric facilities, 6 asymmetric and 3 no default position.[2] Our primary focus is the symmetric case, but we also consider more general conditions where possible.

In recent years the use of last look has found significant regulatory interest, and in one case has led to a fine of \$150 million and a subsequent high profile

[2] 9 providers did not disclose this information and some of those with default positions also offer clients an option to apply different rejection criteria on request.

employment tribunal case ([74]) for a major liquidity provider, demonstrating both the complexity and inconsistencies in the theory and practice of regulation in this area. See also [33] for a discussion of the extent to which these mechanisms might be considered abusive or legitimate.

In foreign exchange a large proportion of major liquidity providers have signed up to the FX Global Code ([30]), a voluntary code of practice setting out best practice in all aspects of dealing. Last look is addressed in Principle 17 of the code, where it is stated that

'A Market Participant should be transparent regarding its last look practices in order for the Client to understand and to be able to make an informed decision as to the manner in which last look is applied to their trading. The Market Participant should disclose, at a minimum, explanations regarding whether, and if so how, changes to price in either direction may impact the decision to accept or reject the trade, the expected or typical period of time for making that decision, and more broadly the purpose for using last look.

If utilised, last look should be a risk control mechanism used in order to verify validity and/or price. The validity check should be intended to confirm that the transaction details contained in the request to trade are appropriate from an operational perspective and there is sufficient available credit to enter into the transaction contemplated by the trade request. The price check should be intended to confirm whether the price at which the trade request was made remains consistent with the current price that would be available to the Client.'

Clearly, questions about the costs and fairness of last look are important and one of our aims in Chapter 5 is to quantify the costs and benefits of last look facilities in the context of existing market making models.

The regulatory position is moving quite fast in this area and during the period in which this thesis has been written we are aware of various moves away from asymmetric last look features, although trade acceptance protocols of some form continue to be in use with most liquidity providers. Some providers, for example XTX markets, who have also disseminated a range of useful practical research informed by the vast data and insight they hold as a major liquidity provider, offer two ‘streams’, one where last look is applied and one where it is not. Clearly there is a lot of scope for different types of facility to be matched to different clients and may be varied according to the relationship between liquidity provider and client. As XTX set out on their webpage ‘Disclosures on eFX Trading Practices’ ([70]):

'Every counterparty has its own trading style and liquidity requirements, which can vary greatly from counterparty to counterparty. XTX aims to tailor the liquidity it provides to each counterparty on a case by case basis, so as to meet each counterparty's specific needs and requirements, while taking into account the market impact and volume of their anticipated trading. In order to do this, XTX provides two different types of liquidity streams...'

No doubt this area will continue to move at pace as the advantages and potential abuses of such facilities become better understood and we hope that this work will contribute to this progress. In the following section we review the existing market making literature, with a particular focus on the papers whose work we build on in later chapters.

1.3 Existing Models for Market Making

The model of Garman (1976) [47] is often considered the earliest model of market making and is the first that attempts to work rigorously in the field of market microstructure. The model has just a single monopolistic market maker, who has full control over prices. They fix their prices at time 0 and keep them fixed throughout the whole trading period, observing demand arriving according to a Poisson process whose rate reacts linearly to the bid and ask prices they set. The market maker starts with a certain cash holding B_0 and inventory holding Q_0 and acts in a risk-neutral way, seeking to maximise the expected profit whilst avoiding either their cash or inventory holdings dropping to zero.

This model is extended by the model of Ho & Stoll (1981) [57], who allow the market maker to change their bid and ask prices continuously over time. As in [47], the demand functions are linear, but now the asset behaves as a Brownian motion and the market maker is risk-averse with a concave utility function and the authors use a stochastic optimal control technique to solve the problem.

The model of Kyle (1985) [64] was one of the first to consider the way that market makers should adjust their quotes in the presence of informed traders and has been extended by many others ([2] [3] [4] [5] [18] [28] [29] [31] [63]). Another important paper is that of Glosten and Milgrom (1985) [48] which considers adjustments to market makers' optimal bid-ask spreads that result from informed traders and adverse selection. It also considers market depth - that is the quantity of orders available at different price levels and how this is impacted by the actions of informed traders.

The Kyle model and all of these extensions assume that market makers are risk-neutral. They conclude that in equilibrium the utility of each market maker is a martingale and that their optimal strategies are to set prices to be conditional expectations of the asset's fundamental value.

Although the market makers' risk-neutrality makes the model tractable, it is not consistent with behaviour observed in markets. There is significant empirical evidence to show that most market maker's behaviour exhibits risk-aversion in such a way that causes their demand to mean revert around certain target levels ([63], [59]). Some attempts to extent the work of Kyle and others without the risk-neutral assumption have been made in both discrete and continuous time, but there are significant challenges that arise [90] [26]). In this work we take a slightly different starting point which does not face the same challenges and allows us to model market makers as risk-averse.

The model of Grossman and Miller (1988) [50] captures an important aspect of the market maker's rationale. Market makers as we define them do not hold any long view about the future price movement of an asset and so have no inherent incentive to enter the market. The risk of entering the market is a cost to them, and [50] gives a fundamental quantification of the premium a market maker would need to be paid in order to take on this risk. In particular a market maker for an asset with volatility σ^2, optimising for a utility function of the form $U(x) = -\exp(-\gamma x)$ will demand a premium of $\gamma\sigma^2$ to take the risk of holding (or shorting) a unit of asset for a unit of time, a term we will see often in the work that follows.

The models we will consider in this thesis are mostly inspired by and build upon work in a few key papers, firstly those of Avellaneda and Stoikov [1] (2008) and of Guéant, Lehalle and Fernandez-Tapia [52] (2013) that we describe in more detail in Sections 1.3.1 and 1.3.2 respectively. Some of our models also build closely on the paper of Guéant [51] (2017). We do not describe that in detail here as the model of Chapter 2 has so much in common with it.

The Avellaneda Stoikov framework [1] replaced a monopolistic market maker with one that is infinitessimally small, so that the reference stock price is effectively exogenous. They substitute optimal limit orders for optimal price quotes so that it can be considered as a model of the Limit Order Book. Turning other game theoretic models into a purely stochastic one provided a framework which researchers in mathematical finance were a lot more comfortable and as a result this became the foundation for many research papers. In particular we will take a detailed look below at the model of Guéant, Lehalle and Fernandez-Tapia (2013) [52] which we extend to the last look case in Chapter 5 and in Chapter 2 we consider a model very closely related to the model in Guéant (2017) [51] but we note also that there are many others who have built on the same work in a number of related ways ([43], [53], [42], [24], [75], [22], [44], [21], [19], [9], [10]).

Models that build directly on the papers above are certainly the most popular in the literature, though building on Avellaneda Stoikov and the work that has followed it is not the only option. We note for example the contribution of Law and Viens (2019) [66] who introduce a somewhat more complicated model with orders classified into 12 different types that they claim captures more realistically features of the Limit Order Book.

1.3.1 Avellaneda and Stoikov (2008)

In their paper [1], Avellaneda and Stoikov mathematically formalised a market making problem and used techniques of stochastic optimal control in order to characterise optimal quoting behaviour. In Section 5.1 we extend this model to include last look and so set out here the main features of their model and the results that we use in our extension.

The market maker seeks, from initial cash x and inventory q at time t, to maximise an exponential utility function

$$v(x, s, q, t) = \mathbb{E}_t[-\exp(-\gamma(X_T + q_T S_T))|S_T = s, q_t = q, X_t = x],$$

where T is a terminal time after which no further trading may take place. The utility of holding a 'frozen' inventory of quantity q and making no trades until time T can be expressed as[3]

$$v(x, s, q, t) = \mathbb{E}_t(-\exp(-\gamma(x + qS_T)) = -\exp\Big(-\gamma(x + qs) + \frac{\gamma^2 q^2 \sigma^2 (T - t)}{2}\Big).$$

The **reservation bid price** r^b and the **reservation ask price** r^a are defined implicitly as

$$v(x - r^b(s, q, t), s, q + 1, t) = v(x, s, q, t)$$

and

$$v(x + r^a(s, q, t), s, q - 1, t) = v(x, s, q, t).$$

Thus the reservation prices are the prices that make the market maker indifferent between holding their current portfolio and holding their current portfolio plus (bid) or minus (ask) one additional unit. Straightforward computations show these can be expressed as

$$r^b(s, q, t) = s + (-1 - 2q)\frac{\gamma\sigma^2(T - t)}{2},$$

and

$$r^a(s, q, t) = s + (1 - 2q)\frac{\gamma\sigma^2(T - t)}{2}.$$

The average of these two prices is referred to as the **reservation price** or **indifference price**

$$r(s, q, t) = s - q\gamma\sigma^2(T - t).$$

Avelleneda and Stoikov then go on to consider a stochastic control problem, where as control the market maker continuously chooses distances from the mid-price

$$\delta^b = s - p^b,$$

[3]Here we have applied the moment generating function of the normal random variable $S_T - s \sim N(0, \sigma^2(T - t))$

and

$$\delta^a = p^a - s,$$

to set their quotes p^a and p^b. The demand they observe depends on their quotes, and orders are modelled to arrive as a Poisson process with arrival rates

$$\lambda^b(\delta^b) = Ae^{-k\delta^b},$$

and

$$\lambda^a(\delta^a) = Ae^{-k\delta^a}.$$

These choices, which have been used often in subsequent work, including in this thesis, are given justification by Avellaneda and Stokov by reference to some empirical studies of limit order books ([46], [49], [71] [85] [93]). The basic intuition is that δ^a and δ^b represent the margin the market maker charges on each trade, and the lower this margin the greater the demand they will achieve.

They consider a market maker wishing to again maximise their exponential utility

$$u(x, s, q, t) = \max_{\delta_b, \delta_a} \mathbb{E}_t[-\exp(-\gamma(X_T + q_T S_T))],$$

where $(X_t)_{t \leq T}$ and $(q_t)_{t \leq T}$ are now stochastic processes representing their running cash and inventory. Optimally they find that the market maker should set their bid and ask quotes p^b and p^a around the indifference price above, with bid and ask quotes

$$p^b(s, q, t) = s + (-1 - 2q)\frac{\gamma\sigma^2(T-t)}{2} - \frac{1}{\gamma}\ln\Big(1 + \frac{\gamma}{k}\Big),$$

and

$$p^a(s, q, t) = s + (1 - 2q)\frac{\gamma\sigma^2(T-t)}{2} + \frac{1}{\gamma}\ln\Big(1 + \frac{\gamma}{k}\Big).$$

We note these are the indifference prices with an extra term

$$\frac{1}{\gamma}\ln\Big(1 + \frac{\gamma}{k}\Big)$$

added or subtracted.

1.3.2 Guéant, Lehalle and Fernandez-Tapia (2013)

In the paper [52], a very similar set up to [1] is considered, although they impose an inventory cap so that the market maker's inventory always stays within $q \in \{-Q, ..., Q\}$. By making a change of variables they are then able to express the system of HJB equations as a system of ordinary differential equations.

As in [1], the formulae derived for the optimal quotes depend on t. Whilst in some contexts this might be natural, for example where positions are closed at the end of the day, in other contexts it is not and [52] also considers the asymptotic behaviour of the quotes. In particular they find that the optimal quotes in the limit as $T \to \infty$ to be

$$p^b(s,q) = s - \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{1}{k}\ln\left(\frac{f_q^0}{f_{q+1}^0}\right),$$

and

$$p^a(s,q) = s + \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{1}{k}\ln\left(\frac{f_q^0}{f_{q-1}^0}\right).$$

where f^0 is the solution of a certain eigenvector problem and is given by

$$f^0 \in \underset{f\in\mathbb{R}^{2Q+1}\setminus\{0\}}{\arg\min} \frac{\sum_{q=-Q}^{Q}\alpha q^2 f_q^2 + \eta\sum_{q=-Q}^{Q-1}(f_{q+1}-f_q)^2 + \eta f_Q^2 + \eta f_{-Q}^2}{\sum_{-Q}^{Q} f_q^2},$$

where $\alpha = \frac{k}{2}\gamma\sigma^2$ and $\eta = A(1+\frac{\gamma}{k})^{-1+\frac{\gamma}{k}}$.

Further, by instead considering a related eigenvector problem in $L^2(\mathbb{R})$ approximations for the closed form bid and ask quotes are found as

$$p^b(s,q) = s - \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{2q+1}{2}\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{(1+\frac{k}{\gamma})}},$$

and

$$p^a(s,q) = s + \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) - \frac{2q-1}{2}\sqrt{\frac{\sigma^2\gamma}{2kA}\left(1+\frac{\gamma}{k}\right)^{(1+\frac{k}{\gamma})}}.$$

1.4 Mathematical Preliminaries

1.4.1 The Dynamic Programming Principle and HJB Equations

The main mathematical tools that we will consider and make use of are those of stochastic optimal control. We present an overview of the key ideas of dynamic programming and HJB equations that are essential for the thesis. Readers looking for a more thorough introduction to the subject could consult [82] or [77] or the relevant sections of [76]. The book by Cartea et al. [23] also includes a good introduction in the context required to understand algorithmic trading problems of a variety of types, including the market making problems we consider in detail later on.

We follow the notation of [82] in this section and suppose we consider a control problem in a form that will occur repeatedly through this thesis, where the state of the system is determined by a stochastic differential equation (SDE)

$$dX_s = b(X_s, \alpha_s)ds + \sigma(X_s, \alpha_s)dW_s,$$

where X may take values in $\mathbb{R}^n$ and W may be a d-dimensional Brownian motion.

Where we consider infinite horizon problems, this model should not depend on time (so as to capture stationarity) whereas in a time-dependant problem we could also consider

$$dX_s = b(X_s, \alpha_s, t)ds + \sigma(X_s, \alpha_s, t)dW_s. \tag{1.1}$$

In any case, the Brownian motion W is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions and the control $\alpha = (\alpha_s)$ is a progressively measurable process (with respect to $\mathbb{F}$), valued in a set of admissible controls $A \subset \mathbb{R}^m$.

We suppose that the measurable functions $b : \mathbb{R}^n \times A \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times A \to \mathbb{R}^{n\times d}$ satisfy a uniform Lipshitz condition in A. That is $\exists K \geq 0, \forall x, y, \in \mathbb{R}^n, \forall a \in$

A,

$$|b(x,a) - b(y,a)| + |\sigma(x,a) - \sigma(y,a)| \leq K|x-y|.$$

Then, denoting by $\mathcal{A}$ the set of control processes α such that

$$\mathbb{E}\left[\int_0^T |b(0,\alpha_t)|^2 + |\sigma(0,\alpha_t)|^2 dt\right] < \infty,$$

the existence of a strong solution $\{X_s,^{t,x}, t \leq s \leq T\}$ to the SDE (1.1) is assured for any $\alpha \in \mathcal{A}$ and any initial condition $(t,x) \in [0,T] \times \mathbb{R}^n$. When we work with a finite horizon problem, the gain function can be defined, for suitable measurable functions $f : [0,T] \times \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$, as

$$J(t,x,\alpha) = \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s)ds + g(X_T^{t,x})\right],$$

for all $(t,x) \in \mathbb{R}^n$. The objective is to maximise over control processes this gain function J to give the value function

$$v(t,x) = \sup_{\alpha \in A} J(t,x,\alpha). \tag{1.2}$$

Given initial condition $(t,x) \in [0,T) \times \mathbb{R}^n$, a control $\hat{\alpha}$ is optimal if $v(t,x) = J(t,x,\hat{\alpha})$. We are mostly interested in Markovian controls, that is a control process α of the form $\alpha_s = a(s, X_s^{t,x})$ for a measurable function $a : [0,T] \times \mathbb{R}^n \to A$. We refer the reader to [82] for a description of the infinite time horizon case as well as a fuller account of the theory in the finite time horizon case.

We can now state the dynamic programming principle (DPP) as follows. For any stopping time θ valued in $[t,T]$ we have

$$v(t,x) = \sup_{\alpha \in \mathcal{A}(t,x)} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)ds + v(\theta, X_\theta^{t,x})\right].$$

A proof of a slightly stronger DPP can be found in [82] and further descriptions of this key technique of stochastic control can also be found in many places, including [41],[92] and [15]. The DPP is also known as the Bellman principle and dates back to the work of Bellman from 1952 ([6], [7]). The interpretation is that we can split the optimisation problem into two parts. Firstly we look for an optimal control from the stopping time θ given that the process is in the state $X_\theta^{t,x}$. That is, we compute $v(\theta, X_\theta^{t,x})$. Then we maximise over controls acting on $[t,\theta]$ the quantity

$$\mathbb{E}\left[\int_t^\theta f(s,,X_s^{t,x},\alpha_s)ds + v(\theta, X_{\theta^{t,x}})\right].$$

In a great deal of the work in this thesis we will make use of an infinitesimal version of the dynamic programming principle, known as the dynamic programming equation, or Hamilton Jacobi Bellman (HJB) equation. By considering the time $\theta = t+h$ and considering carefully the limit as $h \to 0$ it can be shown (see for example [82]) that the value function v should satisfy, $\forall (t,x) \in [0,t) \times \mathbb{R}^n$, the HJB equation

$$-\frac{\partial v}{\partial t}(t,x) - \sup_{a \in A}\left[\mathcal{L}^a v(t,x) + f(t,x,a)\right] = 0,$$

where $\mathcal{L}^a$ is the operator associated to the diffusion (1.1) for the constant control a, defined by

$$\mathcal{L}^a v = b(x,a).D_x v + \frac{1}{2}\mathrm{tr}(\sigma(x,a)\sigma'(x,a)D_x^2 v).$$

The regular terminal condition associated to this PDE is

$$v(T,x) = g(x),\ \forall x \in \mathbb{R}^n,$$

which results immediately from the definition of the value function (1.2).

1.4.2 Utility functions and Risk Sensitive Control

Following [41] we let Φ denote a random variable and imagine that some values of Φ may be more significant than others so that rather than simply considering the expected value of Φ we consider a *risk sensitive criterion*

$$\mathbb{E}[F(\Phi)],$$

where F is a non-linear function. In particular in this work we are interested in the case where

$$F(\Phi) = \exp(\rho\Phi),$$

for $\rho \neq 0$. In the language of economics, thinking of F as a utility function, this choice is a CARA utility function. That is, if we consider the commonly used measure of absolute risk aversion (ARA)

$$r_F(\Phi) = \frac{|F''(\Phi)|}{|F'(\Phi)|},$$

then we have that $r_F(\Phi) = |\rho|$, a constant. That is to say that the risk/utility function F has constant absolute risk aversion (CARA). Now let us define the *certainty equivalent expectation*

$$\mathbb{E}^0(\Phi) = F^{-1}(\mathbb{E}[F(\Phi)]),$$

which in our CARA case then takes the form

$$\mathbb{E}^0(\Phi) = \rho^{-1}\log(\mathbb{E}[\exp(\rho\Phi)]). \tag{1.3}$$

We will only consider utility functions of CARA form, as is most common in the literature our work builds upon. We note however that other choices are sometimes taken. For example, Fodra and Labadie [43] also consider the risk-neutral case as well as a risk-neutral case with a penalisation on the terminal inventory, and Cartea, Jaimungal and Penalva along with various co-authors also consider an objective function that is the expected value of profit and loss less a running penalty on inventory ([23] [21] [22] [24]).

Returning to (1.3) and applying Taylor expansions we note that

$$\begin{aligned}
\lim_{\rho\to 0}\mathbb{E}^0(\Phi) &= \lim_{\rho\to 0}\rho^{-1}\log[1+\rho\mathbb{E}(\Phi)+\frac{\rho^2}{2}\mathbb{E}(\Phi^2)+O(\rho^3)] \\
&= \lim_{\rho\to 0}\rho^{-1}\Big[\rho\mathbb{E}(\Phi)+\frac{\rho^2}{2}\mathbb{E}(\Phi^2)-\tfrac{1}{2}[\rho\mathbb{E}(\Phi)]^2+O(\rho^3)\Big] \\
&= \mathbb{E}(\Phi)+\lim_{\rho\to 0}\Big[\frac{\rho}{2}\mathrm{Var}(\Phi)+O(\rho^2)\Big],
\end{aligned}$$

which means that for small $|\rho|$ that $\mathbb{E}^0(\Phi)$ is approximately a weighted combination of the mean and the variance. In our applications in Chapters 2, 3 and in particular in Chapter 4 we will work with expressions of these forms, and when considering the long run behaviour of the systems we study we will also be interested in some large deviation properties, which we will introduce in the next section.

1.4.3 Large Deviations

By way of introduction to the results of large deviations theory we consider a toy example that we have borrowed from den Hollander [58]. Let $X_1, X_2, ...$ be i.i.d. random variables defined by

$$\mathbb{P}\left(X_1 = \frac{1}{2}\right) = \mathbb{P}\left(X_1 = \frac{3}{2}\right) = \frac{1}{2}, \tag{1.4}$$

and define $S_n = \sum_{k=1}^{n} S_k$.

Suppose we are interested in the long run behaviour as $n \to \infty$ of

$$\mathbb{E}\left(\left(\frac{1}{n}S_n\right)^n\right). \tag{1.5}$$

A strong law of large numbers applies and tells us that almost surely $\frac{1}{n}S_n \to 1$ and so naively we might expect that $\frac{1}{n}\log\mathbb{E}\left(\left(\frac{1}{n}S_n\right)^n\right) \to 0$, but this intuition turns out not to be correct. The real result is a consequence of Cramér's Theorem, which goes back to [32].

Theorem 1.4.1 *Cramér's Theorem*

Let (X_i) be i.i.d. random variable satisfying

$$\varphi(t) = \mathbb{E}(e^{tX_1}) < \infty \qquad \forall t \in \mathbb{R}.$$

Let $S_n = \sum_{k=1}^{n} S_k$. Then, for all $a > \mathbb{E}(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a),$$

where

$$I(z) = \sup_{t\in\mathbb{R}}[zt - \log\varphi(t)].$$

Proof See Theorem I.4 in [58].

Remark 1.4.2 *The same statement as Theorem 1.4.1 also holds for* $\mathbb{P}(S_n \leq an)$ *and* $a < \mathbb{E}(X_1)$. *This can be checked by considering the mirror reflection* $X_1 \to -X_1$.

Remark 1.4.3 *The function* $I(z)$ *in Theorem 1.4.1 is referred to the rate function and this shows that the rate function is the Legendre-Fenchel transform of the cumulant generating function,* $\log \varphi$.

In [58] it is shown that for the example above described in (1.4) and (1.5) that we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(\left(\frac{1}{n} S_n\right)^n\right) = b,$$

where

$$b = \sup_{a>0}[\log a - J(a)], \tag{1.6}$$

and

$$J(a) = \begin{cases} \log 2 + (a - \frac{1}{2}) \log(a - \frac{1}{2}) + (\frac{3}{2} - a) \log(\frac{3}{2} - a) & a \in \left[\frac{1}{2}, \frac{3}{2}\right] \\ \infty & \textit{otherwise.} \end{cases}$$

Straightforward algebra shows in particular that the optimiser a^* of (1.6) satisfies $a^* \neq 1$ and that $b > 0$. This is we see that the expected value is not dominated by the almost sure behaviour that $\frac{1}{n} S_n \to 1$, but rather by the rare event where $\frac{1}{n} S_n$ is in the vicinity of $a^* \neq 1$.

Remark 1.4.4 *Indeed [58] also shows that if* $a > \mathbb{E}(X_1)$, *then the rate function* $I(z) \geq I(a)$ *for all* $z \geq a$ *and so we can rewrite the result of Theorem 1.4.1 as*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in A\right) = -\inf_{z\in A} I(z) \qquad \textit{with } A = [a, \infty].$$

This we see that the large deviation $\{\frac{1}{n} S_n \in A\}$ *is essentially dominated by the event that* $\frac{1}{n} S_n$ *is close to* $\bar{z}$, *the minimiser of* $I(z)$ *on* A. *In the words of den Hollander [58] this illustrates the key principle of large deviation theory, that 'any large deviation is done in the least unlikely of all the unlikely ways!'*

This is all meant to give a flavour of and a motivation for our need to consider large deviations in Chapter 4. For more details and a fuller account of the theory we refer the reader to [58] [35] [11] or [27]. In the remainder of this section we follow Pham [81] and state some large deviation results for control problems closer to those we will consider in Chapter 4.

Suppose we have a real valued process X_t^α controlled by a control process $\alpha = (\alpha_t) \in \mathcal{A}$ and suppose that a law of large numbers applies to X^α so that

$$\bar{X}_T^\alpha := \frac{X_T^\alpha}{T} \qquad \text{converges almost surely as } T \to \infty.$$

Then we expect that it also satisfies a large deviation principle

$$\mathbb{P}[\bar{X}_T^\alpha \geq c] \simeq \exp(-I(c,\alpha)T), \qquad \text{as } T \to \infty,$$

where the rate function I measures the rate of exponential convergence of the probability for $\bar{X}_T^\alpha$ to overtake a level c. The rate function is related to the moment generating function of X_T^α via the Legendre transform

$$I(c,\alpha) = \sup_{\lambda \geq 0}[\lambda c - \Lambda(\lambda,\alpha)],$$

where

$$\Lambda(\lambda,\alpha) = \limsup_{T\to\infty} \frac{1}{T} \ln \mathbb{E}[\exp(\lambda X_T^\alpha)].$$

In Chapter 4 we will notice that the system we study exhibits large deviation effects and will adapt the model we propose there in a suitable way to propose an efficient way of calculating the exponential integrals that we require there.

1.5 Overview of this Thesis

In the final section of this introduction we set out an overview of the work that follows in this thesis and highlight our main results.

In Chapter 2 we study a discrete time market making model in the style of Guéant [51]. By making suitable approximations we are able to find a linear differential operator whose spectral theory gives us a great deal of insight into the market making problem. We are able to prove rigorously some results that are not fully justified in [51] and to add some insights into the long-run value of market making per unit time. Further, when we lift the inventory cap the market maker is subject to, we are able to relate the market making problem to a quantum harmonic oscillator, giving a neat quantification of the value of market making. We find that the market making problem is essentially one of inventory control, and we are able to roughly quantify the long run value of market making and separate it from the cost of starting with non-zero inventory.

In Chapter 3 we formulate a novel continuous model that is motivated by the discrete model of Chapter 2. We work slightly less rigorously in this chapter, but find that, in suitable limits, we can recover results about the value of market making found in Chapter 2 and the optimal quotes of Guéant and his collaborators ([52] [51]). This model also sets the scene for the novel model we introduce in Chapter 4.

In Chapter 4 we reformulate the market making problem of Chapter 3 as a one-dimensional control problem. We put the view that market making is essentially a problem of inventory management centre-stage, and introduce a direct control on the drift of the inventory process, noting that such a choice implies a natural choice of bid and ask quotes. We look to apply results from the paper of Nagai [72] which allow us to relate the long-run value of market making to the principal eigenvalue of a suitable linear operator.

We propose a slight modification to the demand functions that allow us, in Theorem 4.4.1, to apply the work of [72] directly to our problem. We then make a natural conjecture that suggests the long-run value of market making should be constant and investigate the consequences of this, solving the resulting PDEs numerically. In order to account for large deviation effects we make a measure change, after which we are able to compute the relevant exponential integrals in the problem numerically. This provides a framework in which we may optimise over the various parameters in the problem efficiently, rather than having to repeatedly resolve PDEs, something that would be particularly helpful when applied to the models we consider in Chapter 5.

In Chapter 5 we consider market making with last look. We begin in Section 5.1 by extending the reservation prices proposed by Avellaneda and Stoikov [1] to a variety of last look criteria, including symmetric and asymmetric facilities. We find closed forms for these reservation prices and compare them to those without last look. We also consider the potential benefits of last look to the client as well as the market maker.

In Section 5.2 we then extend the model of Guéant, Lehalle and Fernandez-Tapia [52] to the last look case, and are able to show an asymptotic result based on the spectral theory of an appropriate matrix which allows us to capture the long-run behaviour of market making with last look and use numerical simulations to consider the overall impact on the profitability for the market maker, including in the presence of toxic order flow.

Finally, in Section 5.3 we propose a new model for last look that includes counterparties who may or may not provide toxic order flow. We then suggest a novel continuous model and use numerical simulations to show that it captures the problem well. A key advantage of this model is that it is suited to the framework of Chapter 4 and so we may be able to find efficient methods for optimising over the various parameters.

The final contribution, Chapter 6 of the thesis, is presented 'by publication' and sits slightly separately from the rest of the work in the thesis. We include the full text of the paper [56], which at the time of writing is under review, and to which the author of this thesis contributed around 20% of the work. The paper involves some novel results about Echo State Networks[4] that are primarily attributable to other authors. The paper also includes some applications of Echo State Networks, and in particular the author of this thesis contributed most significantly to the development of an application to the market making problem presented.

[4]An Echo State Network (ESN) is a type of single-layer recurrent neural network with randomly chosen internal weights and a trainable output layer.

Chapter 2

A Discrete Model for Market Making

In this chapter we consider a market making model in the style of Guéant and his collaborators in [51] and [52]. Throughout this chapter, and indeed this thesis, we assume that the market maker is trading a single asset whose price is given by an arithmetic Brownian motion

$$dS_t = \sigma dW_t. \tag{2.1}$$

This choice of *arithmetic* Brownian motion is standard in the literature[1] and is justified by the typically short time horizons under consideration.

Our models will contain two fundamental sources of risk. One arises from the price movements given by (2.1), and another from the random arrival of buy and sell orders. In Section 2.1 we begin with a simplified version of the model in which we take $\sigma = 0$ in (2.1). We will refer to this scenario as a 'riskless' world, though in fact we mean that we are temporarily switching off just this first source of randomness, allowing us to focus on how the market maker should optimally respond to the random demand.

We will find in this section optimal strategies and a long-run value of market

[1] In particular this choice is also made in all of the main models this work is based around, for example [1],[51] and [52].

making that will be a very useful reference point for the work that follows. Indeed we will find that this model gives an upper bound for the value of market making in a world where $\sigma > 0$.

In Section 2.2 we consider the case $\sigma > 0$ and construct a model in the style of [51]. The solution in this case will involve approximating the HJB equation in a suitable way and using some ideas from functional analysis and the spectral theory of a certain linear operator related to the resulting PDE. In particular we will make use of the Krein-Rutman theorem that provides us with a single leading eigenvalue/eigenvector of the linear operator that will dominate where the time horizon of the problem is far away.

Up until this point we make an assumption that the inventory levels are subject to a cap, so that there is a Q such that the inventory must always stay within the interval $[-Q, Q]$. In Section 2.3 we remove this inventory cap in order to relate the market making problem to a quantum harmonic oscillator. Whilst we leave a fully rigorous justification of this step to future work, it is a very natural extension that allows us to find a clean and interpretable form of the value function for market making. In particular we will find that there is a long run value per unit time of market making (which can be directly compared to the value in the riskless world of Section 2.1) and an additional cost if we start with a non-zero inventory level.

2.1 Market Making in a Riskless World

All the risks involved in market making arise from the possibility of adverse movements in the asset price whilst the market maker is holding inventory. Later we will model the stock price process as an (arithmetic) Brownian motion

$$dS_t = \sigma dB_t,$$

though in this first instance we consider the case $\sigma = 0$. Our intuition leads us to expect that the profitability of market making should be decreasing in σ and indeed this case will provide an upper bound for the potential value of market making in the more general case.

Although described as 'riskless' this world is not free of randomness. Unpredictability will still arise from the random arrival times of buyers and sellers, the intensity of which will be determined by the competitiveness of the market maker's quotes. The central problem in this case will be to determine the spread the market maker should quote at a given time in order to optimally profit from this demand.

2.1.1 Formulation of the Riskless Model

Since there is no randomness in the underlying asset price we may model it at all times as a constant

$$S_t = s, \qquad \forall t \geq 0.$$

The market maker's optimisation problem will involve incentivising orders by pricing as competitively as possible whilst not reducing too significantly the profit made in each trade. Throughout this chapter the market maker's control will involve setting a strategy δ comprised of two quantities that may be changed over time $\delta = \{(\delta_t^b)_{t \leq T}, (\delta_t^a)_{t \leq T}\}$ up until some terminal time T. We may usefully think of δ_t^b and δ_t^a as 'half-spreads' that the provide the market maker with a profit on each trade. These quantities will then determine the market maker's bid and ask prices S_t^b and S_t^a

$$S_t^b = s - \delta_t^b,$$

$$S_t^a = s + \delta_t^a,$$

and $\delta_t^b + \delta_t^a$ will represent the overall bid-ask spread the market maker is advertising at any given time.

Following [51] and [52] we model demand for the asset using two independent Poisson point processes $(N_t^b)_t$ and $(N_t^a)_t$ with rates Λ^b and Λ^a respectively. These intensities will be determined by the market maker's choice of δ_t^b and δ_t^a, so that $\Lambda^b = \Lambda^b(\delta^b)$ and $\Lambda^a = \Lambda^a(\delta^a)$. In particular we will consider the case where these demand functions depend on the market maker's quotes as

$$\Lambda^a(\delta^a) = Ae^{-k\delta^a}, \quad \Lambda^b(\delta^b) = Ae^{-k\delta^b}.$$

We assume that the size of each trade is a constant Δ, and so increments of the market maker's inventory level are given by

$$dq_t = \Delta dN_t^b - \Delta dN_t^a.$$

Their cash holding X_t then has dynamics

$$\begin{aligned} dX_t &= S_t^a \Delta dN_t^a - S_t^b \Delta dN_t^b \\ &= (s + \delta_t^a)\Delta dN_t^a - (s - \delta_t^b)\Delta dN_t^b. \end{aligned}$$

We assume that the market maker maximises a CARA utility function[2]

$$u^\delta(t, x, q) = \mathbb{E}_{t,x,q}[-\exp\{-\gamma(X_T^\delta + q_T^\delta s)\}],$$

where $\mathbb{E}_{t,x,q}$ is the expected value of the process where the market maker starts with cash x and inventory q at time t, and γ is a risk aversion parameter characterising the market maker.

The quantity $X_t + q_t s$ represents the marked-to-market value of the market maker's portfolio at time t, being the sum of the cash holding and the value of their currently held inventory priced at the market mid-price.

2.1.2 A Single Variable to Represent Wealth

In this riskless set-up there is no reason to prefer to hold cash instead of an equivalent valued inventory and so it is natural not to distinguish between cash and asset in our model. Thus we replace $X_t + q_t s$ with a single wealth process W_t. Whenever a trade occurs this wealth increases instantaneously by $\delta^b\Delta$ or $\delta^a\Delta$ depending on the direction of the transaction.

So we consider a process $W_t = x_t + q_t s$ with dynamics

$$dW_t = \delta^b \Delta dN_t^b + \delta^a \Delta dN_t^a,$$

[2]Note this is a standard choice in the literature as discussed further in Section 1.4.2.

and a market maker seeking at time t to maximise over controls $\delta = \{\delta^a, \delta^b\}$

$$u(t, w) = \sup_\delta (u^\delta(t, w)) = \sup_\delta \left(\mathbb{E}[-\exp\{-\gamma W_T^\delta\}]\right). \tag{2.2}$$

With the set-up simplified in this way we now use some standard techniques to find the market maker's optimal strategy and the value they derive from market making.

Proposition 2.1.1 *Choosing $\Lambda(\delta) = Ae^{-k\delta}$ and defining $\xi = \frac{k}{\gamma\Delta}$ we have that the optimal control is*

$$\delta_a = \delta_b = -\frac{1}{\gamma\Delta}\ln\left(\frac{k}{k+\gamma\Delta}\right) = -\frac{\xi}{k}\ln\left(\frac{\xi}{1+\xi}\right),$$

with associated suprema

$$\sup_{\delta^b}\left\{\frac{\Lambda(\delta^b)}{\gamma}\left(1 - e^{-\gamma\Delta\delta^b}\right)\right\} = \sup_{\delta^a}\left\{\frac{\Lambda(\delta^a)}{\gamma}\left(1 - e^{-\gamma\Delta\delta^a}\right)\right\} = \frac{A}{\xi\gamma}\left(\frac{\xi}{1+\xi}\right)^{1+\xi} := A_\xi.$$

The value function for the control problem as defined in (2.2) above is given by

$$u(t, w) = -e^{-\gamma(w+2A_\xi(T-t))}. \tag{2.3}$$

Proof By standard arguments (see e.g. [82] or [41]) we expect u to solve the HJB equation

$$0 = -\partial_t u - \sup_{\delta^b}\left\{\Lambda(\delta^b)[u(t, w + \Delta\delta^b) - u(t, w)]\right\} - \sup_{\delta^a}\left\{\Lambda(\delta^b)[u(t, w + \Delta\delta^a) - u(t, w)]\right\}.$$

Taking as ansatz $u(t, w) = -e^{-\gamma(w+\theta(t))}$, the HJB equation becomes

$$\partial_t\theta = -\sup_{\delta^b}\left\{\frac{\Lambda(\delta^b)}{\gamma}\left(1 - e^{-\gamma\Delta\delta^b}\right)\right\} - \sup_{\delta^a}\left\{\frac{\Lambda(\delta^a)}{\gamma}\left(1 - e^{-\gamma\Delta\delta^a}\right)\right\}.$$

Since the form of the two suprema is identical we write δ for δ^a and δ^b and perform a straightforward differentiation. The required optimal δ^* is the solution to

$$\frac{\partial}{\partial\delta}\Big(\frac{\Lambda(\delta)}{\gamma}(1 - e^{-\gamma\Delta\delta})\Big) = 0.$$

Hence we have

$$\begin{aligned}
-kAe^{-k\delta^*}(1-e^{-\gamma\Delta\delta^*})+\gamma\Delta Ae^{-\gamma\Delta\delta^*}e^{-k\delta^*} &= 0\\
-k(1-e^{-\gamma\Delta\delta^*})+\gamma\Delta e^{-\gamma\Delta\delta^*} &= 0\\
-k+(k+\gamma\Delta)e^{-\gamma\Delta\delta^*} &= 0\\
e^{-\gamma\Delta\delta^*} &= \frac{k}{k+\gamma\Delta},
\end{aligned}$$

which gives

$$\begin{aligned}
\delta^* &= -\frac{1}{\gamma\Delta}\ln\left(\frac{k}{k+\gamma\Delta}\right)\\
&= -\frac{\xi}{k}\ln\left(\frac{\xi}{1+\xi}\right),
\end{aligned}$$

where $\xi = \frac{k}{\gamma\Delta}$. Then to evaluate the suprema we simply substitute in the optimal control found. Noting that we have $Ae^{-k\delta^*} = A\left(\frac{k}{k+\gamma\Delta}\right)^{\frac{k}{\gamma\Delta}}$ then

$$\begin{aligned}
\sup_{\delta}\left\{\frac{\Lambda(\delta)}{\gamma}\left(1-e^{-\gamma\Delta\delta}\right)\right\} &= \sup_{\delta}\left\{\frac{Ae^{-k\delta}}{\gamma}\left(1-e^{-\gamma\Delta\delta}\right)\right\}\\
&= \frac{A}{\gamma}\left(\frac{k}{k+\gamma\Delta}\right)^{\frac{k}{\gamma\Delta}}\left(1-\frac{k}{k+\gamma\Delta}\right)\\
&= \frac{A}{\gamma}\left(\frac{\gamma\Delta}{k+\gamma\Delta}\right)\left(\frac{k}{k+\gamma\Delta}\right)^{\frac{k}{\gamma\Delta}}\\
&= \frac{A\Delta}{k}\left(\frac{k}{k+\gamma\Delta}\right)^{\frac{k}{\gamma\Delta}+1}\\
&= \frac{A}{\xi\gamma}\left(\frac{\xi}{1+\xi}\right)^{1+\xi} := A_\xi
\end{aligned}$$

After evaluating the suprema the HJB equation reduces to

$$\begin{aligned}
\partial_t\theta &= -2A_\xi\\
\Rightarrow \theta &= 2A_\xi(T-t),
\end{aligned}$$

and so the value function for this control problem defined above is given by (2.3) as required.

□

Remark 2.1.2 *In the risk-free case the HJB equation is of a very simple form, and the optimal quotes arise as the result of straightforward differentiation. The market maker only has to balance the (utility of their) profit per trade* $\Delta\delta^a$ *and* $\Delta\delta^b$ *against the impact this has on the level of demand* $\Lambda(\delta^a)$ *and* $\Lambda(\delta^b)$ *they receive.*

Remark 2.1.3 *In the limit as* $\xi \to \infty$ *the optimal quotes tend to* $\delta^a = \delta^b = \frac{1}{k}$*, since*

$$\lim_{\xi\to\infty} \delta^* = \lim_{\xi\to\infty} -\frac{\xi}{k} \ln\left(\frac{\xi}{1+\xi}\right) = \lim_{\xi\to\infty} -\frac{1}{k} \ln\left[\left(\frac{\xi}{1+\xi}\right)^{\xi}\right] = \frac{1}{k}.$$

Remark 2.1.4 *In Chapter 10 of [23], Cartea et al. consider a market making model with explicit penalties for running inventory and terminal inventory as well as an inventory cap. When in Section 10.2.1 they simplify to assume the market maker doesn't penalise running or terminal inventory and faces no inventory cap, they find that the market maker optimally seeks to maximise the probability of their limit order being filled. The model we have here in the limit finds the same solution for the optimal quotes as in that case. Indeed, one way that we could have* $\xi \to \infty$ *would be via* $\gamma \to 0$*. That is a market maker who is not risk averse in our set up is behaving in the same way as one who does not penalise the risk of holding inventory in the case considered in [23].*

Hence in this simplified case we find that the function θ is just the constant $2A_\xi$ multiplied by the remaining time available to act as a market maker. Thus we may interpret $2A_\xi$ as a profit rate from market making. The market maker should be equally happy to have an additional wealth $2A_\xi(T-t)$ or the opportunity to act as a market maker from time t until time T.

In the next section we will consider similar arguments in a related set-up where the value of the asset is no longer constant. The problem considered in this section will be an important special case. The constant $\frac{1}{k}$ will appear repeatedly in the optimal quotes the market maker chooses and the constant $2A_\xi$ will represent a reference point and upper bound for the value the market maker is able to realise in other conditions.

2.2 Market Making in a Risky World

In this section we describe the main model of this chapter, and show that the HJB equation resulting from the control problem considered can be approximated with a certain parabolic PDE. The set-up is similar to that of Guéant [51] though we take a more rigorous look at the underpinning mathematics. We will find optimal quotes that agree with the model in [51] and our approach will also allow us to focus on a neat approximation for the approximate overall value of market making.

2.2.1 Formulation of the Full Discrete Model

We now focus on a problem that can more realistically capture the situation faced by a market maker. In particular we now consider that there the asset trades around a reference price S_t that follows an arithmetic Brownian motion. That is, we have $\sigma > 0$ and

$$dS_t = \sigma dW_t.$$

The introduction of risk into the asset price drives our primary characterisation of a market maker's preferences:

All other things being equal, market makers prefer to hold as little inventory as possible. Their ideal transactions are round-trip trades where they profit from selling at a slightly higher price than they buy for, doing so quickly enough to avoid any changes in the price of the underlying asset.

Of course the real world rarely provides such opportunity, but this fundamental preference leads to the market maker's general aversion to holding inventory. So, all other things being equal, we should expect the value of market making to be a decreasing function of the size of their inventory level, $|q|$ in the sense that the mark to market value of their portfolio is also kept constant, so that any decreasing inventory is suitably balanced by increasing cash. They do not in general want to hold either long or short positions in any asset and consider the possibility that the asset price moves whilst they hold inventory to be a risk that reduces their overall profitability.

The rest of the set-up of the problem remains unchanged, and the market maker sets bid and ask prices S_t^b and S_t^a at distances δ_t^b and δ_t^a from this reference price so that as before we have

$$S_t^b = S_t - \delta_t^b,$$
$$S_t^a = S_t + \delta_t^a.$$

The market maker's problem is to optimally control δ_t^b and δ_t^a in order to maximise (a utility function of) profit. Although these quotes will still depend on a range of parameters in the model, the primary difference from the previous section is that they will now also depend on the market maker's current inventory level. Because of the market maker's general aversion to holding short or long positions we expect that they will want to offer more competitive prices as their inventory moves away from 0. In doing so they will cause their inventory process to stay as close to zero as possible whilst still making enough trades to be profitable.

As in the riskless case we model demand for the asset using two independent Poisson point processes $(N_t^b)_t$ and $(N_t^a)_t$ with rates $\Lambda^b(\delta_t^b)$ and $\Lambda^a(\delta_t^a)$. In particular we will again focus on the case where $\Lambda^a(\delta^a) = Ae^{-k\delta^a}, \Lambda^b(\delta^b) = Ae^{-k\delta^b}$.

Again we assume that the size of each trade is a constant Δ, and so the market maker's inventory level is given by $dq_t = \Delta dN_t^b - \Delta dN_t^a$ and their cash holding X_t has dynamics $dX_t = S_t^a \Delta dN_t^a - S_t^b \Delta dN_t^b = (s + \delta_t^a)\Delta dN_t^a - (s - \delta_t^b)\Delta dN_t^b$ as before.

The CARA utility function to be optimised is also unchanged from the work above aside from the fact that S_t is now random. So our market maker will optimise $u^\delta(t, x, s, q)$ over choices of this control $\delta = \{\delta_s^a, \delta_s^b\}_{t \leq s \leq T}$, and where

$$u(t, x, s, q) = \sup_{\delta} u^\delta(t, x, S, q) = \sup_{\delta^a, \delta^b} \mathbb{E}_{t,x,s,q}[-\exp\{-\gamma(X_T^\delta + Q_T^\delta S_T)\}].$$

Now $\mathbb{E}_{t,x,s,q}$ is the expected value of the process where the market maker starts with cash x and inventory q of an asset priced at $S_t = s$ at time t, and γ is a risk aversion parameter characterising the market maker as before.

We also assume that the market maker is subject to an inventory cap $\pm Q$. If their inventory ever hits this cap, they stop quoting on one side of the book so that it cannot be exceeded.

2.2.2 An Ansatz Suitable for a CARA Utility Function

In Section 2.1 we were able to simply consider the wealth of the agent $w = x + qs$ as cash and stock are easily exchangeable in a risk free world. In a risky world, this quantity will continue to play a fundamental role and will be referred to as the Marked to Market (MtM) value of the portfolio. In practice it may be necessary to use other conventions[3] but here we adopt the convention of marking to market at the asset mid-price.

Although we now consider the cash and asset holding separately, when using a CARA utility function it is possible to factor out this MtM value $x + qs$. Thus we can write the candidate value function as

$$u(t, x, q, s) = -\exp(-\gamma(x + qs + \theta(t, q, s))).$$

The guiding idea is that the value of a position will incorporate this marked to market value and the function θ will capture any aspect of the value that is determined by the riskiness of the position. Our primary interest will be in the dependence of θ on q. We expect that θ will be negative so that it will quantify the reluctance of the market market maker to hold inventory at various levels. Further, although the form of θ may include the volatility σ of the stock, due to the spatial homogeneity of S it really has no reason to depend on the value of S and so it makes sense to take as ansatz:

$$u(t, x, q, s) = -\exp(-\gamma(x + qs + \theta(t, q))). \tag{2.4}$$

[3]For example where portfolio values are being computed for regulatory risk management purposes, using a bid or ask price rather than a midprice, or including a liquidation cost may represent a more accurate liquidation value of the portfolio.

In the sections that follow we implicitly assume that θ exists and is differentiable. In Section 2.2.5 we will be a little more rigorous, but we defer this until after we have made the main approximation step that is required.

2.2.3 An HJB Equation for Market Making in a Risky World

Proposition 2.2.1 *Choosing $\Lambda(\delta) = Ae^{-k\delta}$, the function $\theta(t, q)$ in (2.4) satisfies*

$$0 = -\partial_t \theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - A_\xi \left(e^{kp^+} + e^{kp^-} \right), \tag{2.5}$$

where $p^\pm := \frac{1}{\Delta}(\theta(t, q \pm \Delta) - \theta(t, q))$ and we define $\xi = \frac{k}{\gamma\Delta}$ and $A_\xi = \frac{A}{\xi\gamma}\left(\frac{\xi}{1+\xi}\right)^{1+\xi}$ as before.

Proof By standard results we expect u to solve the HJB equation

$$\begin{aligned} 0 = -\partial_t u - \tfrac{1}{2}\sigma^2 \partial^2_{ss} u &- \mathbb{1}_{\{q \le Q\}} \sup_{\delta^b} \{\Lambda(\delta^b)[u(t, x - \Delta s + \Delta\delta^b, q + \Delta, s) - u(t, x, q, s)]\} \\ &- \mathbb{1}_{\{q \ge -Q\}} \sup_{\delta^a} \{\Lambda(\delta^a)[u(t, x + \Delta s + \Delta\delta^a, q - \Delta, s) - u(t, x, q, s)]\}. \end{aligned}$$

Upon substituting this into in the ansatz (2.4) and taking the demand function $\Lambda(\delta) = Ae^{-k\delta}$ this becomes

$$\begin{aligned} 0 = -\partial_t \theta + \tfrac{1}{2}\gamma\sigma^2 q^2 &- \mathbb{1}_{\{q \le Q\}} \sup_{\delta^b} \left\{ \tfrac{A}{\gamma} e^{-k\delta^b} \left(1 - e^{-\gamma\Delta\left(\delta^b + p^+\right)} \right) \right\} \\ &- \mathbb{1}_{\{q \ge -Q\}} \sup_{\delta^a} \left\{ \tfrac{A}{\gamma} e^{-k\delta^a} \left(1 - e^{-\gamma\Delta(\delta^b + p^-)} \right) \right\}. \end{aligned}$$

The suprema may be evaluated using a simple first order condition. Since the two suprema have essentially the same form, writing $\{\delta, p\}$ in place of $\{\delta^b, p^+\}$ and $\{\delta^a, p^-\}$ we have

$$-\tfrac{kA}{\gamma} e^{-k\delta} \left(1 - e^{-\gamma\Delta(\delta+p)} \right) + \tfrac{A}{\gamma} e^{-k\delta} \left(\gamma\Delta e^{-\gamma\Delta(\delta+p)} \right) = 0.$$

After some straightforward rearrangement this gives

$$e^{-\gamma\Delta(\delta+p)} = \frac{k}{k+\gamma\Delta}, \text{ and } e^{-k\delta} = \left(\frac{k}{k+\gamma\Delta}\right)^{\frac{k}{\gamma\Delta}} e^{kp}.$$

Writing $\xi = \frac{k}{\gamma\Delta}$ and substituting into the suprema this yields

$$0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - \tfrac{A}{\xi\gamma}\left(\tfrac{\xi}{1+\xi}\right)^{1+\xi}\left(e^{kp^+} + e^{-kp^-}\right), \tag{2.6}$$

which after writing $A_\xi = \frac{A}{\xi\gamma}\left(\frac{\xi}{1+\xi}\right)^{1+\xi}$ gives the result as stated.

□

It is not practical to solve (2.5) analytically and so we follow [51] in considering an approximate solution. In particular we will modify $p^\pm$ in the following way, defining

$$p_\epsilon^\pm := \tfrac{1}{\Delta}\left(\theta(t, q \pm \epsilon\Delta) - \theta(t,q)\right).$$

Now consider the equation

$$0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - A_\xi(e^{kp_\epsilon^+} + e^{kp_\epsilon^-}), \tag{2.7}$$

which we note is precisely equation (2.5) when we take $\epsilon = 1$.

Proposition 2.2.2 *Equation (2.7) can be expressed as*

$$0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - 2A_\xi - A_\xi\Delta k\epsilon^2\partial^2_{qq}\theta(t,q) - A_\xi k^2\epsilon^2(\partial_q\theta(t,q))^2 + o(\epsilon^2). \tag{2.8}$$

Proof Taking a series expansion of $p_\epsilon^\pm$ in q we have

$$\begin{aligned} p_\epsilon^+ &= \tfrac{1}{\Delta}(\theta(t,q+\epsilon\Delta) - \theta(t,q)) = \epsilon\partial_q\theta(t,q) + \tfrac{1}{2}\Delta\epsilon^2\partial^2_{qq}\theta(t,q) + o(\epsilon^2); \\ p_\epsilon^- &= \tfrac{1}{\Delta}(\theta(t,q-\epsilon\Delta) - \theta(t,q)) = -\epsilon\partial_q\theta(t,q) + \tfrac{1}{2}\Delta\epsilon^2\partial^2_{qq}\theta(t,q) + o(\epsilon^2). \end{aligned}$$

Then taking a series expansion of the exponential function gives

$$\begin{aligned} e^{kp_\epsilon^+} &= 1 + \epsilon k\partial_q\theta(t,q) + \tfrac{1}{2}\Delta k\epsilon^2\partial^2_{qq}\theta(t,q) + \tfrac{1}{2}k^2\epsilon^2(\partial_q\theta(t,q))^2 + o(\epsilon^2); \\ e^{kp_\epsilon^-} &= 1 - \epsilon k\partial_q\theta(t,q) + \tfrac{1}{2}\Delta k\epsilon^2\partial^2_{qq}\theta(t,q) + \tfrac{1}{2}k^2\epsilon^2(\partial_q\theta(t,q))^2 + o(\epsilon^2). \end{aligned}$$

Substituting these into equation (2.7) gives the desired result.

□

2.2.4 An Approximate PDE for the Long-Run Behaviour

Following Guéant we next discard the $o(\epsilon^2)$ terms and take $\epsilon = 1$ in (2.8) to give

$$0 = -\partial_t \theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - 2A_\xi - A_\xi \Delta k \partial^2_{qq}\theta(t,q) - A_\xi k^2 (\partial_q \theta(t,q))^2. \tag{2.9}$$

We leave a more careful consideration of this step for future work and instead focus on its consequences. In Chapters 3 and 4 we work in a continuous framework where it is possible to find these conclusions with more rigour. Indeed we note that it seems to be an inherent disadvantage of the discrete framework that these arguments are more challenging to make.

Having made this approximation, in the remainder of this section we are able to describe the long run behaviour of the system by considering the spectral theory of a suitable linear operator. First we make a transformation so that the system is in a more convenient form.

Proposition 2.2.3 *Taking $v = \exp\left(\frac{k}{\Delta}\theta\right)$, the equation (2.9) may be expressed as*

$$\partial_t v = 2\tfrac{k}{\Delta} A_\xi v - Bv, \tag{2.10}$$

where

$$Bv := \tfrac{1}{2}\tfrac{k}{\Delta}\gamma\sigma^2 q^2 v - k\Delta A_\xi \partial^2_{qq} v.$$

Proof Substituting $v = \exp(\frac{k}{\Delta}\theta)$ into (2.9), we have

$$\theta = \tfrac{\Delta}{k}\ln v, \partial_t\theta = \tfrac{\Delta}{k}\tfrac{\partial_t v}{v}, \partial_q\theta = \tfrac{\Delta}{k}\tfrac{\partial_q v}{v}, \partial^2_{qq}\theta = \tfrac{\Delta}{k}\left(\tfrac{v\partial^2_{qq}v - (\partial_q v)^2}{v^2}\right).$$

Then (2.9) becomes the linear PDE

$$0 = -\tfrac{\Delta}{k}\tfrac{\partial_t v}{v} + \tfrac{1}{2}\gamma\sigma^2 q^2 - 2A_\xi - A_\xi\Delta^2\left(\tfrac{v\partial^2_{qq}v - (\partial_q v)^2}{v^2}\right) - A_\xi\Delta^2\left(\tfrac{\partial_q v}{v}\right)^2,$$

which in turn simplifies to give

$$0 = \partial_t v - \tfrac{k}{\Delta}(\tfrac{1}{2}\gamma\sigma^2 q^2 - 2A_\xi)v + k\Delta A_\xi \partial^2_{qq} v, \tag{2.11}$$

as required.

□

2.2.5 Analysis of Long-Run Behaviour

In this section we will analyse the long run behaviour of (2.10). The main strength of making the approximations that led to this equation lies in the fact that the operator B is a positive self-adjoint operator with a compact inverse and thus we may appeal to well established spectral theory to analyse its behaviour as $T \to \infty$.

The results we are about to discuss hold for a wider class of operators than just the operator B defined above. In Section 2.3 we will use our specific knowledge of the form of B to write a more explicit form of the value function, but in this section we work in greater generality and get a flavour for the functional analysis involved.

For a more detailed understanding of the approach taken here we refer the reader to the textbooks of [14] and [17]. In particular we make use of some theorems stated in [17] to pave a path through the functional analysis required to reach our conclusions. In particular we are going to use a classical L^2 theory of the solution of PDEs to interpret the problem.

We also refer the reader to Ishii [60] which allows us to conclude that the solution we find to the PDE is a valid solution to the control problem. The normal approach in control theory would be to interpret solutions to the control problem as viscosity solutions to avoid the need to prove a priori regularity of the value function. In this example, the operator is nice enough that [60] allows us to say that if we have a solution in $H^1(\Omega)$ as we will describe below, then it will also be a viscosity solution to the control problem.[4]

[4] Ishii [60] justifies this for general elliptic PDEs. Although we are working in the parabolic case, this is included since the results in [60] do not require any uniform ellipticity conditions.

So far we have been quite loose about the domain on which we are working and so we now set a large value of Q and consider the equation (2.10), that is

$$\partial_t v = 2\tfrac{k}{\Delta} A_\xi v - Bv$$

on the domain $\Omega \times [0,T]$, where $\Omega := [-Q,Q]$. This cap can be thought of as the maximum (short or long) allowable inventory.

Note that when we talk about $v(t,q)$ there is also implicitly a dependence on the terminal time T. Later in Theorem 2.2.5 we will want to consider the long run behaviour and will make this dependence explicit, but until then we keep $T>0$ fixed.

In order to define boundary conditions we set $\theta(t,\pm Q) = -\theta^*$ for some large θ^*, thinking of θ^* representing a liquidation premium. That is, if our inventory should ever hit the boundary we are prevented from making further trades on one side of the book and have to bank a loss. By taking the value of Q large enough we may consider this to be a very unlikely event for which the liquidation premium would be significant. The consequence is that since $v = \exp\left(\frac{k}{\Delta}\theta\right)$ we would have $v \approx 0$ in this case and so we choose as a boundary condition

$$v(t,Q) = v(t,-Q) = 0 \quad \forall t \in [0,T].$$

This ensures that for every t the solutions $v(t,\cdot)$ always belong to the same linear subset of $L^2(\Omega)$ which will be necessary in applying the results of [17]. To apply these results we must first also consider the natural boundary condition

$$v(T,q) = 1 \quad \forall q \in \Omega^\circ \tag{2.12}$$

where $\Omega^\circ = (-Q,Q)$ is the interior of Ω. Note the two boundary conditions are not continuous at $v(T,Q)$, though this is not as problematic as it initially seems. We will be able to use the classical theory in Brezis [17] to find existence and uniqueness of solutions immediately for times $t \in [0, T-\epsilon]$ and we will still be able to extend this naturally to a solution for which we may take the limit $\epsilon \to 0$. That is in Proposition 2.2.4 we will find a suitable function of q for each $t \leq T-\epsilon$ and then use a continuity argument to extend to the full interval $[0,T]$.

In particular, Theorem 10.9 of [17] tells us that the solution to (2.11) exists and Theorem 10.10 of [17] further gives us that this solution v^ϵ is unique (in a weak sense) and satisfies $v^\epsilon \in C^\infty(\Omega \times [0, T - \epsilon]), \quad \forall \epsilon > 0$.[5]

Next we note that the operator B is self-adjoint and so appealing to classical spectral theory (see e.g. [14]) we can find an orthonormal basis of $L^2(\Omega)$ consisting of eigenvectors of B. In particular, the eigenvectors and eigenvalues (e_i, λ_i) satisfy $Be_i = \lambda_i e_i$, we have $\langle e_i, e_j \rangle = 0$ for $i \neq j$ and $\langle e_i, e_i \rangle = 1$ and there is a singular minimal eigenvalue λ_0 in that we have $\lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots$ The fact that the leading eigenfunction has (algebraic and geometric) multiplicity 1 (so that $\lambda_0 < \lambda_1$) is a direct consequence of the Krein-Rutman Theorem (see e.g. Theorem 6.13 in [17]).[6]

As a result we are able to determine a form for the value function and analyse its behaviour in the limit as $T \to \infty$ in the next two propositions.

Proposition 2.2.4 *Let $\Omega = [-Q, Q]$ and $(e_i, \lambda_i)_{i=0}^\infty$ be the eigenvectors and eigenvalues of the operator $B := \frac{1}{2} \frac{k}{\Delta} \gamma \sigma^2 q^2 - k \Delta A_\xi \partial^2_{qq}$. Then*

$$v(t, q; T) = \sum_{i=0}^\infty \langle 1, e_i \rangle \exp\left[\left(2\tfrac{k}{\Delta} A_\xi - \lambda_i\right)(T - t)\right] e_i(q) \tag{2.13}$$

is the unique solution to $\partial_t v = 2\frac{k}{\Delta} A_\xi v - Bv$ such that $v \in L^2([0, T]; H^1(\Omega))$, $v \in C([0, T]; L^2(\Omega))$; for each $0 \leq t < T$, $v(t, Q; T) = v(t, -Q) = 0$ and $v(T, q; T) = 1 \ \forall q \in \Omega^\circ$.

Proof Since we have a solution $v^\epsilon \in C^\infty(\Omega \times [0, T - \epsilon]), \quad \forall \epsilon > 0$ then we can take $0 < t < T - \epsilon$ and write v in the basis of eigenvectors of B as

$$v^\epsilon(t, q; T) = \sum_{i=0}^\infty \langle v^\epsilon, e_i \rangle e_i(q) =: \sum_{i=0}^\infty a_i^\epsilon(t) e_i(q). \tag{2.14}$$

[5] A proof of Theorem 10.9 of [17] can be found, for example in [68] and of Theorem 10.10 in [68], [45] or [65]. The issues with the boundary condition mean that (2.12) does not satisfy the necessary compatibility conditions to be able to immediately extend this to say that we have a solution $v \in C^\infty(\Omega \times [0, T])$ in Theorem 10.9 so we have to consider the limit as $\epsilon \to 0$ more carefully.

[6] This is analogous to the Perron-Frobenius results used in Chapter 5 where we are working with the eigenvectors of a matrix instead. Proofs of various forms of the Krein-Rutman Theorem can be found in [13] [73] [86] [88] and [91].

Then equation (2.10) becomes

$$\begin{aligned}
\partial_t v^\epsilon &= 2\tfrac{k}{\Delta} A_\xi v^\epsilon - B v^\epsilon \\
\sum_{i=0}^{\infty} \partial_t a_i^\epsilon(t) e_i(q) &= 2\tfrac{k}{\Delta} A_\xi \sum_{i=0}^{\infty} a_i^\epsilon(t) e_i(q) - B\left(\sum_{i=0}^{\infty} a_i^\epsilon(t) e_i(q)\right) \\
\sum_{i=0}^{\infty} \partial_t a_i^\epsilon(t) e_i(q) &= 2\tfrac{k}{\Delta} A_\xi \sum_{i=0}^{\infty} a_i^\epsilon(t) e_i(q) - \sum_{i=0}^{\infty} a_i^\epsilon(t) \lambda_i e_i(q) \\
\sum_{i=0}^{\infty} \partial_t a_i^\epsilon(t) e_i(q) &= \sum_{i=0}^{\infty} \left(2\tfrac{k}{\Delta} A_\xi - \lambda_i\right) a_i^\epsilon(t) e_i(q).
\end{aligned}$$

For each j we can take the inner product with $e_j(q)$ to give

$$\begin{aligned}
\partial_t a_j^\epsilon(t) &= \left(2\tfrac{k}{\Delta} A_\xi - \lambda_j\right) a_j^\epsilon(t) \\
\Rightarrow a_j^\epsilon(t) &= c_j^\epsilon \exp\left[\left(2\tfrac{k}{\Delta} A_\xi - \lambda_j\right)(T-t)\right].
\end{aligned}$$

where the c_j are constants that can be determined from the boundary conditions. In particular, if we impose a boundary condition $v^\epsilon(T-\epsilon, q) = 1$, then we have

$$\sum_{i=0}^{\infty} a_i^\epsilon(T-\epsilon) e_i(q) = 1 \quad \forall q \in \Omega.$$

Taking the inner product of each side with a particular e_j we find that

$$a_j^\epsilon(T-\epsilon) = \langle 1, e_j \rangle,$$

and so

$$c_j^\epsilon = \langle 1, e_j \rangle \exp\left(-\epsilon\left(2\tfrac{k}{\Delta} A_\xi - \lambda_j\right)\right),$$

and we note $\lim_{\epsilon\to 0} c_j^\epsilon = \langle 1, e_j \rangle$. Substituting these back into (2.14) we have that for $0 < t < T-\epsilon$,

$$v^\epsilon(t, q; T) = \sum_{i=0}^{\infty} c_i^\epsilon \exp\left[\left(2\tfrac{k}{\Delta} A_\xi - \lambda_i\right)(T-t)\right] e_i(q). \tag{2.15}$$

Now fix $\eta \in (0, T]$. Then for $\epsilon < \eta$ we have

$$v^{\epsilon}(T-\eta, q; T) = \sum_{i=0}^{\infty} c_i^{\epsilon} \exp\left[\left(2\tfrac{k}{\Delta}A_{\xi} - \lambda_i\right)\eta\right] e_i(q). \tag{2.16}$$

and we can let $\epsilon \to 0$ to find a function $v(T-\eta, q; T) \in H^1(\Omega)$ such that

$$v(T-\eta, q; T) = \sum_{i=0}^{\infty} \langle 1, e_j \rangle \exp\left[\left(2\tfrac{k}{\Delta}A_{\xi} - \lambda_i\right)\eta\right] e_i(q). \tag{2.17}$$

Now we can appeal to Theorem 10.9 of [17], which tells us that this function lies in $C([0,T] \times L^2(\Omega))$ so that we are allowed to take the limit as $\eta \to 0$ and so

$$v(t, q) = \sum_{i=0}^{\infty} \langle 1, e_j \rangle \exp\left[\left(2\tfrac{k}{\Delta}A_{\xi} - \lambda_i\right)(T-t)\right] e_i(q) \tag{2.18}$$

has all of the properties required.

□

As a consequence of the previous proposition and the Krein-Rutman Theorem telling us that operator B has a unique leading eigenvalue so that $\lambda_0 < \lambda_1$, for large T we can see that the term involving λ_0 will dominate the others and so the following theorem immediately follows.

Theorem 2.2.5 *As $T \to \infty$ we have (with convergence in the $L^2(\Omega)$ sense) that*

$$e^{-\left(2\frac{k}{\Delta}A_{\xi} - \lambda_0\right)T} v(0, q; T) \to v_{\infty}(q)$$

where $v_{\infty}(q) \in H_0^1(\Omega)$ is the solution to

$$Bv = \lambda_0 v,$$

where $H_0^1(\Omega)$ is the subset of $H^1(\Omega)$ satisfying the boundary conditions $v(\pm Q) = 0$. In particular we have that there is a constant C such that for sufficiently large T

$$\left\| e^{-\left(2\frac{k}{\Delta}A_{\xi} - \lambda_0\right)T} v^T(0, q) - v_{\infty}(q)) \right\|_{L^2} \leq C e^{-(\lambda_0 - \lambda_1)t}.$$

Moreover (see [17] Theorem 9.31 and Remark 30), v_0, λ_0 can be recovered via the Rayleigh-Ritz formula

$$\lambda_0 = \inf_{v \in H_0^1(\Omega)} \frac{\langle v, Bv \rangle}{\langle v, v \rangle} = \inf_{v \in H_0^1(\Omega), ||v||_{L^2}=1} \langle v, Bv \rangle,$$

and

$$v_0 = \underset{v \in H_0^1(\Omega), ||v||_{L^2}=1}{\arg\min} \langle v, Bv \rangle.$$

We note that this also justifies the claim of [51]

$$v_0 = \min_{f \in H^1(\mathbb{R}): ||f||_{L^2}=1} \int \tfrac{1}{2} \tfrac{k}{\Delta} \gamma \sigma^2 q^2 v^2 - \Delta k A_\xi (\partial_q v)^2, dq$$

that is given there without proof.

Further, we will have that when T is large, approximately

$$v(t, q) = \langle 1, e_0 \rangle \exp\left[\left(2 \tfrac{k}{\Delta} A_\xi - \lambda_0 \right) (T - t) \right] e_0(q). \tag{2.19}$$

2.3 The Quantum Harmonic Market Maker

In this final substantial section of this chapter we write down a natural extension of the problems considered so far in this chapter. If we lift the inventory cap by letting $Q \to \infty$ so that $\Omega = \mathbb{R}$ then we are able to recognise the operator under consideration as that of a quantum harmonic oscillator. In this case we can then easily write down the known eigenvectors and eigenvalues of the operator B and so form a very neat and interpretable explicit form for the value function. As a result we can quantify the approximate long run value of the initial market making problem. The form we find also allows us to separate clearly the long run value of market making and the cost of having a non-zero initial inventory.

In the following proposition, we use the notation $f(t) \sim g(t)$ as $T - t \to \infty$ to mean that $\frac{f(t)}{g(t)} \to 1$ as $T - t \to \infty$.

Proposition 2.3.1 *In the case* $Bv := \frac{1}{2}\frac{k}{\Delta}\gamma\sigma^2 q^2 v - k\Delta A_\xi \partial^2_{qq} v$*, the only smooth non-negative solution to*

$$\partial_t v = 2\tfrac{k}{\Delta} A_\xi v - Bv$$

satisfying the boundary conditions $\lim_{q\to\pm\infty} v(t,q) = 0 \ \ \forall t$ *and* $v(T,0) = 1$*, satisfies, as* $T - t \to \infty$*,*

$$v(t,q) \sim \exp\left(-\tfrac{1}{4\Delta}\sqrt{\frac{\gamma\sigma^2}{A_\xi}}q^2\right)\exp\left(\tfrac{k}{\Delta}\left(2A_\xi - \Delta\sqrt{\tfrac{1}{2}\gamma\sigma^2 A_\xi}\right)(T-t)\right).$$

Hence the value of market making satisfies, as $T - t \to \infty$*,*

$$u(t,x,q,S) \sim -\exp\left(-\gamma\left(x + qS - \tfrac{1}{4k}\sqrt{\frac{\gamma\sigma^2}{A_\xi}}q^2 + \left(2A_\xi - \Delta\sqrt{\tfrac{1}{2}\gamma\sigma^2 A_\xi}\right)(T-t)\right)\right).$$

Proof We can write down the eigenvectors and eigenvalues of

$$Bv := \tfrac{k}{\Delta}(\tfrac{1}{2}\gamma\sigma^2 q^2)v - k\Delta A_\xi \partial^2_{qq} v$$

by noting that the operator B is related to that of a quantum harmonic oscillator in the sense of Proposition B.0.2. In the notation used there we have $\alpha = k\Delta A_\xi$, $\beta = \frac{k}{\Delta}\frac{1}{2}\gamma\sigma^2$, $\kappa = 0$ so that we find its leading eigenpair to be

$$e_0(q) = (\tfrac{\beta}{\alpha\pi})^{\frac{1}{4}} e^{-\frac{1}{2}\sqrt{\frac{\beta}{\alpha}}q^2}, \qquad \lambda_0 = -k\sqrt{\tfrac{1}{2}\gamma\sigma^2 A_\xi}.$$

Then by a similar reasoning to Theorem 2.2.5 we find that as $T - t \to \infty$, for a suitable constant c we have,

$$v(t,q) \sim ce_0(q)\exp((2\tfrac{k}{\Delta}A_\xi + \lambda_0)(T-t)). \tag{2.20}$$

Applying the boundary condition $v(T,0) = 1$ we find that the constant c must cancel with the constant in $e_o(q)$ to give

$$v(t,q) \sim e^{-\frac{1}{2}\sqrt{\frac{\beta}{\alpha}}q^2}\exp((2\tfrac{k}{\Delta}A_\xi - \lambda_0)(T-t)).$$

Unpicking the substitution $\theta = \frac{\Delta}{k}\log v$ and the original ansatz $u(t,x,S,q) =$

$-\exp\{-\gamma(x + qS + \theta(t,q))\}$ we have

$$u(t,x,S,q) = -\exp\left\{-\gamma(x+qS) - \tfrac{\gamma\Delta}{k}\log v\right\},$$

and so substituting (2.20) in place of v and also substituting for α and β then gives the result immediately.

$\square$

We recall that in the simplified, riskless world of Section 2.1 we found that the function θ is just the constant $2A_\xi$ multiplied by the remaining time available to act as a market maker, so that the market maker should be equally happy to have an additional $2A_\xi(T-t)$ or the opportunity to act as a market maker from time t until time T. Proposition 2.3.1 allows us to make a direct comparison and we find the following two major differences in the form of the value function:

1. The profit rate per unit time is reduced from $2A_\xi$ to $2A_\xi - \Delta\sqrt{\frac{1}{2}\gamma\sigma^2 A_\xi}$; and
2. There is also a penalty of $\frac{1}{4k}\sqrt{\frac{\gamma\sigma^2}{A_\xi}}q^2$, quantifying how undesirable it is to start with a non-zero inventory.

As we expected, market making in the riskless world has given us an upper bound for the value of market making in a more general world, and this result has allowed us to approximately quantify the difference.

2.3.1 An alternative approach

In the work of this chapter we have assumed that the inventory levels should be fixed to a bounded domain $[-Q, Q]$ and a bankruptcy condition be applied at the boundaries. The compactness of the resulting operator then allowed us to approach the eigenvalue problem rigorously using the results contained in [17].

We think that it should also be possible to make the step to the quantum harmonic oscillator case in section 2.3 more rigorously too. Pinsky [83] begins by working with a general elliptic operator on a bounded domain and provides a rigorous theory for its principal eigenvalue. Under a certain uniform ellipticity condition, they are then able to extend these results to arbitrary domains $D \subset \mathbb{R}^d$.

In Chapter 5 of [83] they work in the one-dimensional case and further show that many of the calculations can be worked out explicitly. Both Chapter 5 of [83] and the paper of Elliott [37] propose integral tests to prove the existence of the principal eigenvalue that would appear to be very applicable to our problem and should allow us to justify more rigorously the step to the unbounded domain that we have made in section 2.3. A full consideration of this approach will be explored in future work.

2.4 Summary

In this chapter we considered a model for market making in the style of Guéant [51] in two cases. Firstly, in the 'riskless' case where the underlying asset has no randomness, we found optimal strategies and a long-run value of market making that provide a useful reference case for all the work that follows. Then in the more general case we have been able to approximate the resulting HJB equation in a suitable way and use spectral theory to analyse the optimal quotes and quantify the long-run value of market making per unit time. In the final section, we lifted the inventory cap and made a natural comparison of the market making problem to that of a quantum harmonic oscillator, and in doing so were able to suggest a form of the value function that neatly separates and quantifies both the value of market making and the cost of starting with a non-zero inventory. In doing all of this we have added additional rigour to the results presented in [51] as well as some additional new results about the value of market making. In the next chapter we will formulate a new continuous model for market making that may be considered a natural extension of this model.

Chapter 3

A Continuous Model for Market Making

In this chapter we will study a new continuous model for market making that retains the key features of the model of Chapter 2. We have observed that when market making with a risky asset the inventory level plays a key role. All other things being equal, the market maker would always prefer to not hold inventory and optimally they tend to adjust their quotes in a way that causes their inventory to mean-revert towards zero. Although the value of the control depends on many parameters in the model, the dynamic element rests primarily in response to changes in the inventory level q. In Chapter 4 we will take this focus on controlling the inventory even further, but in this chapter we begin by formulating a continuous model in which the market maker's bid and ask quotes are set up in similar way to Chapter 2.

In Sections 3.1 we set up the continuous model and control problem, basing the parameters and coefficients in the model carefully upon the modelling framework of Chapter 2.

In Sections 3.2 and 3.3 we consider the PDE that the value function should satisfy, and approximate it with a linear PDE that appears to give very reasonable solutions to the problem. Whilst we are not able to make fully rigorous the approximation we are able to recover results from the literature and make some interesting observations about the behaviour of the system. Indeed in Section 3.4 we do exactly this by recognising the linearised PDE as a version of a quantum harmonic oscillator.

A fully rigorous version of the operator theory required for this chapter is beyond the scope of this thesis and left for future work, but we are encouraged by closeness of the results to the literature and Chapter 2. Our real aim for this chapter though is to set the scene for Chapter 4, where we work with a slightly simpler model which we are able to treat more rigorously. The results and the intuition gained from this chapter are therefore important guides for what follows.

3.1 Formulation of the Continuous Model

The model in this section takes that of Chapter 2 as a starting point but imagines that orders may arrive continuously and be of any size. All of the terms in the driving SDEs in this model are motivated directly by rates from the discrete model so that it may naturally be considered as a continuous time equivalent of the model of Chapter 2.

3.1.1 Inventory Process

In the discrete model, demand arises as the result of orders arriving as Poisson point processes with intensities $\Lambda^b(\delta^b)$ on the bid side and $\Lambda^a(\delta^a)$ on the ask side. Since inventory increases with demand on the bid side and decreases with demand on the ask side, the bid orders alone would cause the inventory to drift upwards at rate $\Lambda^b(\delta^b)$ and the ask orders alone would cause the inventory to drift downwards at rate $\Lambda^a(\delta^a)$. Thus we would expect an overall drift in the inventory of $\Lambda^b(\delta^b) - \Lambda^a(\delta^a)$ and so we define the inventory process by the SDE

$$dq_t = \Delta\{\Lambda^b(\delta_t^b) - \Lambda^a(\delta_t^a)\}dt + \zeta(\delta_t^a, \delta_t^b)dB_t^{(1)}, \tag{3.1}$$

where $B_t^{(1)}$ is a Brownian motion and $\zeta(\delta^a, \delta^b)$ is a volatility term that will be specified below.

Remark 3.1.1 *The discrete order size Δ still appears in our continuous model.*

Although it has become somewhat arbitrary it is helpful in making direct comparisons with the results from the discrete case.

A natural choice for the square of the volatility term $\zeta(\delta^a, \delta^b)$ is to take

$$\zeta^2(\delta^a, \delta^b) = \Delta^2\{\Lambda^a(\delta^a) + \Lambda^b(\delta^b)\},$$

since in the original model the orders arrive according to two Poisson processes with rates $\Lambda^a(\delta^a)$ and $\Lambda^b(\delta^b)$. Thus the variance of total orders within a unit of time would be $\Lambda^a(\delta^a) + \Lambda^b(\delta^b)$, which when scaled by the order size Δ gives $\Delta^2\{\Lambda^a(\delta^a) + \Lambda^b(\delta^b)\}$ as the total variance of the inventory process.

3.1.2 Wealth Process

Alongside the inventory process we need to define a wealth process to keep track of the running value that the market maker has accrued. In the discrete model, when a bid order is placed at at price $S_t - \delta^b$ the market maker gains Δ units of inventory in exchange for $\Delta(S_t - \delta^b)$ in cash. If we mark to market the value of Δ units of inventory at the mid-price S_t then this transaction results in an immediate increase in wealth of $\Delta\delta^b$. Since bid orders arrive at rate $\Lambda^b(\delta^b)$ then the bid orders increase wealth at an average rate of $\Lambda^b(\delta^b)\Delta\delta^b$.

By similar reasoning, the ask orders increase wealth at an average rate of $\Lambda^a(\delta^a)\Delta\delta^a$. Adding together the bid and ask orders, the market maker accrues marked to market profits at a rate of $\Delta\{\delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b)\}$. Of course these profits are not certain to be realisable in cash, since whilst holding inventory the price of the asset may change. Since the asset price is modelled as arithmetic Brownian motion, we add a volatility term to represent these fluctuations and so arrive at the SDE

$$dW_t = \Delta\{\delta_t^a\Lambda^a(\delta_t^a) + \delta_t^b\Lambda^b(\delta_t^b)\}dt + \sigma q dB_t^{(2)}, \tag{3.2}$$

where $B_t^{(2)}$ is a second Brownian motion independent of $B_t^{(1)}$. We leave for future work consideration of the case where these are not independent.

3.1.3 Optimisation Problem and Ansatz

The utility function that we seek to optimise for also takes a very similar form to before and we write

$$v(w,q,t) = \sup_{\delta^a,\delta^b} \mathbb{E}_{w,q,t}[-\exp(-\gamma W_T)], \tag{3.3}$$

where $\mathbb{E}_{w,q,t}$ is the expected value of the process where the market maker starts with wealth w and inventory q at time t, and γ is a risk aversion parameter characterising the market maker as before.

We can deduce the existence of the value function on a finite horizon via standard methods. Moreover, the value function can be characterised in the usual manner as a viscosity solution using e.g. Theorem 4.3.1 of Pham [82]. Note that in our setting we do not need the function appearing in (4.17) of [82]. Pham's other conditions are easily verified.

As in Chapter 2, we expect that we should be able to factorise out the current wealth in the utility function, and so take as ansatz

$$v(w,q,t) = -e^{-\gamma(w+\theta(q,t))}. \tag{3.4}$$

Hence we have translated the discrete problem into one that is continuous and essentially equivalent. The rest of this chapter will be devoted to studying and solving this optimisation problem, as well as comparing these solutions to those found in the the discrete case.

3.2 Developing a PDE for the Value Function

Proposition 3.2.1 *Under the optimal choice of control, and taking*

$$\Lambda^a(\delta^a) = Ae^{-k\delta^a}, \Lambda^b(\delta^b) = Ae^{-k\delta^b},$$

the function θ defined in (3.4) satisfies

$$0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - \tfrac{2A\Delta}{ke}\exp\left[\tfrac{k\Delta}{2}(\theta'' - \gamma(\theta')^2)\right]\cosh(k\theta'). \tag{3.5}$$

Proof We consider $v(t, W_t, q_t)$ as a stochastic process and then by standard results[1]

$$\begin{aligned} 0 =& v_t + \tfrac{1}{2}\sigma^2 q^2 v_{ww} && (3.6)\\ &+ \sup_{\{\delta^a,\delta^b\}}\Big[v_w\Delta\{\delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b)\} + v_q\Delta\{\Lambda^b(\delta^b) - \Lambda^a(\delta^a)\} + \tfrac{1}{2}v_{qq}\zeta^2(\delta^a,\delta^b)\Big]. && (3.7)\end{aligned}$$

Now, applying the ansatz (3.4)gives

$$v(w,q,t) = -e^{-\gamma(w+\theta(q,t))} \qquad v_t = \gamma\dot{\theta}e^{-\gamma(w+\theta(q,t))} \qquad v_q = \gamma\theta' e^{-\gamma(w+\theta(q,t))}$$

$$v_w = \gamma e^{-\gamma(w+\theta(q,t))} \qquad v_{ww} = -\gamma^2 e^{-\gamma(w+\theta(q,t))} \qquad v_{qq} = (\gamma\theta'' - \gamma^2(\theta')^2)e^{-\gamma(w+\theta(q,t))}$$

and so, after substituting these into (3.7) and dividing by $-v_w$ we have

$$\begin{aligned} 0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - \sup_{\{\delta^a,\delta^b\}}&\Big[\Delta\{\delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b)\}\\ &+ \theta'\Delta\{\Lambda^b(\delta^b) - \Lambda^a(\delta^a)\} + \tfrac{1}{2}(\theta'' - \gamma(\theta')^2)\zeta^2(\delta^a,\delta^b)\Big]. \end{aligned} \tag{3.8}$$

Substituting $\zeta^2(\delta^a,\delta^b) = \Delta^2\{\Lambda^a(\delta^a) + \Lambda^b(\delta^b)\}$ and separating out the terms in δ^a and δ^b we have

$$\begin{aligned} &\sup_{\{\delta^a,\delta^b\}}\Big[\Delta\{\delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b)\} + \theta'\Delta\{\Lambda^b(\delta^b) - \Lambda^a(\delta^a)\} + \tfrac{1}{2}(\theta'' - \gamma(\theta')^2)\zeta^2(\delta^a,\delta^b)\Big]\\ &\quad=\Delta\sup_{\delta^a}\Big[(\delta^a + (-\theta' + \tfrac{\Delta}{2}(\theta'' - \gamma(\theta')^2))\Lambda^a(\delta^a)\Big]\\ &\qquad+ \Delta\sup_{\delta^b}\Big[(\delta^b + (\theta' + \tfrac{\Delta}{2}(\theta'' - \gamma(\theta')^2))\Lambda^b(\delta^b)\Big]. \end{aligned} \tag{3.9}$$

Further substituting the choice $\Lambda^a(\delta^a) = Ae^{-k\delta^a}, \Lambda^b(\delta^b) = Ae^{-k\delta^b}$ and noting that

[1] See e.g. [41] Theorem IV 7.1.

for any constant α

$$\sup_{\delta}\{(\delta+\alpha)\exp^{-k\delta}\} = \tfrac{1}{k}e^{(k\alpha-1)},$$

with optimising $\delta^* = \frac{1}{k} - \alpha = \frac{1}{k}(1-k\alpha)$, the suprema in (3.9) become

$$\begin{aligned}&\tfrac{A\Delta}{k}\exp\left[k(-\theta' + \tfrac{\Delta}{2}(\theta'' - \gamma(\theta')^2)) - 1\right] + \tfrac{A\Delta}{k}\exp\left[k(\theta' + \tfrac{\Delta}{2}(\theta'' - \gamma(\theta')^2)) - 1\right]\\ &= \tfrac{2A\Delta}{ke}\exp[\tfrac{k\Delta}{2}(\theta'' - \gamma(\theta')^2)]\cosh(k\theta'),\end{aligned}$$

and hence equation (3.8) becomes

$$0 = -\partial_t\theta + \tfrac{1}{2}\gamma\sigma^2 q^2 - \tfrac{2A\Delta}{ke}\exp[\tfrac{k\Delta}{2}(\theta'' - \gamma(\theta')^2)]\cosh(k\theta'),$$

as required.

$\square$

3.3 Approximation with a Linear PDE

We would like to appeal to the same spectral theory as in Chapter 2 but it is not immediately clear that we can do so, as we now have a non-linear operator for which most of the standard theory is not applicable.

Nonetheless we expect a long-run spectral interpretation to be possible for large values of $(T-t)$ and as a first step we write an ansatz representing a an average profit rate times the time remaining plus an inventory-dependent term as

$$\theta(t,q) = \lambda(T-t) - \theta_0(q).$$

Then we have $\partial_t\theta = -\lambda$ and equation (3.5) becomes an ODE for $\theta_0(q)$:

$$-\lambda = \tfrac{1}{2}\gamma\sigma^2 q^2 - \tfrac{2A\Delta}{ke}\exp[\tfrac{k\Delta}{2}(-\theta_0'' - \gamma(\theta_0')^2)]\cosh(-k\theta_0'). \tag{3.10}$$

Proposition 3.3.1 *Suppose (3.10) holds and $g(q) = \exp(\eta\theta_0(q))$, with $\eta = \frac{\gamma\Delta-k}{\Delta}$. Then the pair (λ, g) are an eigenvalue/eigenvector pair for the operator $\mathcal{L}^0$*

given by

$$\mathcal{L}^0 f := -\tfrac{1}{2}\gamma\sigma^2 q^2 f(q) + \tfrac{2A\Delta}{ke} f(q)\exp\left\{ -\tfrac{1}{2}\tfrac{k\Delta^2}{\gamma\Delta - k}\left(\frac{f''(q)}{f(q)}\right) - \tfrac{1}{2}\tfrac{k^2\Delta^2}{(\gamma\Delta-k)^2}\left(\frac{f'(q)}{f(q)}\right)^2 \right.$$
$$\left. + \ln\cosh\left(\tfrac{k\Delta}{k-\gamma\Delta}\frac{f'(q)}{f(q)}\right)\right\}. \tag{3.11}$$

That is we have $\mathcal{L}^0 g = \lambda g$.

Proof Making the substitution $g(q) = \exp(\eta\theta_0(q))$, we have $g'(q) = \eta\theta_0'(q)g(q)$ and $g''(q) = \eta\theta_0''(q)g(q) + \eta^2(\theta_0'(q))^2 g(q)$ so that

$$\theta_0' = \frac{1}{\eta}\frac{g'(q)}{g(q)},$$

$$\theta_0''(q) = \frac{1}{\eta}\frac{g''(q)}{g_0(q)} - \eta(\theta_0'(q))^2 = \frac{1}{\eta}\frac{g''(q)}{g(q)} - \frac{1}{\eta}\left(\frac{g'(q)}{g(q)}\right)^2.$$

Hence (3.10) becomes

$$\lambda = -\tfrac{1}{2}\gamma\sigma^2 q^2 + \tfrac{2A\Delta}{ke}\exp\left\{-\tfrac{k\Delta}{2\eta}\left(\frac{g''(q)}{g(q)} - \left(\frac{g'(q)}{g(q)}\right)^2 + \tfrac{\gamma}{\eta}\left(\frac{g'(q)}{g(q)}\right)^2\right)\right\}\times\cosh\left(-\frac{k}{\eta}\frac{g'(q)}{g(q)}\right).$$

Substituting $\eta = \frac{\gamma\Delta - k}{\Delta}$ we find

$$\lambda = -\tfrac{1}{2}\gamma\sigma^2 q^2 + \tfrac{2A\Delta}{ke}\exp\left\{-\tfrac{k\Delta^2}{2(\gamma\Delta-k)}\left(\frac{g''(q)}{g(q)}\right) - \tfrac{1}{2}\tfrac{k^2\Delta^2}{(\gamma\Delta-k)^2}\left(\frac{g'(q)}{g(q)}\right)^2\right\}\times\cosh\left(\tfrac{k\Delta}{k-\gamma\Delta}\frac{g'(q)}{g(q)}\right).$$

which is of the form required.

□

This non-linear operator is challenging to work with directly and so we work instead with a linear operator that approximates it. A full understanding of the analysis required for approximating the operator as we do is beyond the scope of this thesis and is left for future work. Nonetheless we give here an idea of why we think the approximation is reasonable and are encouraged that we can recover some of the conclusions of the papers [52] and [51] in doing so.

Our basic intuition arises from applying Taylor expansions to terms in the PDE (3.11). Firstly we replace by its quadratic approximation $\ln\cosh(x) \approx \frac{x^2}{2}$ to write

$$\ln\cosh\left(\frac{k\Delta}{k-\gamma\Delta}\frac{f'(q)}{f(q)}\right) \approx \tfrac{1}{2}\tfrac{k^2\Delta^2}{(k-\gamma\Delta)^2}\left(\frac{f'(q)}{f(q)}\right)^2,$$

which gives

$$\mathcal{L}^0 g \approx -\tfrac{1}{2}\gamma\sigma^2 q^2 g(q) + \tfrac{2A\Delta}{ke} g(q)\exp\left\{-\tfrac{1}{2}\tfrac{k\Delta^2}{\gamma\Delta-k}\left(\frac{g''(q)}{g(q)}\right)\right\}.$$

Then we make a linear approximation of $\exp(x) \approx 1 + x$ to give

$$\mathcal{L}^0 g \approx \mathcal{L} g := \left(\tfrac{2A\Delta}{ke} - \tfrac{1}{2}\gamma\sigma^2 q^2\right) g(q) - \tfrac{A\Delta^3}{e(\gamma\Delta-k)} g''(q). \tag{3.12}$$

In the remainder of this chapter we work on the assumption that the form of the function g makes these approximations reasonable and show that in doing so we are able to recover results from [52] and [51] and to align with Chapter 2. The intuition gained from this will also act as a guide for the more rigorous work to follow in Chapter 4.

3.4 The Quantum Harmonic Market Maker (Part 2)

Proposition 3.4.1 *The leading normalised eigenpair* (λ, g_0) *of the operator (3.12) are*

$$\lambda = \tfrac{2A\Delta}{ke} - \sqrt{\tfrac{1}{2}\gamma\sigma^2\tfrac{A\Delta}{ek}\tfrac{k\Delta^2}{(k-\gamma\Delta)}}$$

$$g_0 = \left(\frac{\gamma\sigma^2 e(k-\gamma\Delta)}{2\pi A\Delta^3}\right)^{\frac{1}{4}} e^{-\frac{1}{2}\sqrt{\frac{\gamma\sigma^2 e(k-\gamma\Delta)}{2A\Delta^3}}q^2}.$$

That is we have $\mathcal{L}g_0 = \lambda g_0$, *and further* g_0 *is the only normalised eigenvector that is positive everywhere.*

Proof The operator $\mathcal{L}$ defined in (3.12) is related to that of the quantum harmonic oscillator and is of the form of Proposition B.0.2 with $\alpha = \frac{A\Delta^3}{e(k-\gamma\Delta)}$, $\beta = -\frac{1}{2}\gamma\sigma^2$ and $\kappa = \frac{2A\Delta}{ke}$ and so the result follows immediately.

Remark 3.4.2 *Recall that* $A_\xi := \frac{A\Delta}{k}\left(\frac{\xi}{1+\xi}\right)^{1+\xi}$ *and so in the limit as* $\xi \to \infty$ *we have* $A_\xi \sim \frac{A\Delta}{ke}$*. So by writing*

$$\lambda = \tfrac{2A\Delta}{ke} - \Delta\sqrt{\tfrac{1}{2}\gamma\sigma^2\tfrac{A\Delta}{ek}\tfrac{1}{\left(1-\frac{\gamma\Delta}{k}\right)}},$$

we may note that in this limit

$$\lambda = 2A_\xi - \Delta\sqrt{\tfrac{1}{2}\gamma\sigma^2 A_\xi},$$

which is precisely the value of market making per unit time we found in the discrete case in section 2.3.

Similarly we can write

$$\frac{\gamma\sigma^2 e(k-\gamma\Delta)}{2A\Delta^3} = \frac{\gamma\sigma^2}{2\Delta^2}\frac{ek}{A\Delta}\left(1-\frac{\gamma\Delta}{k}\right) \sim \frac{\gamma\sigma^2}{2A_\xi\Delta^2},$$

and so we can observe that in this limit we can also match this with the inventory dependent term in Chapter 2.

3.5 Computing the Optimal Quotes

In this section we use the result of Proposition 3.4.1 to find approximate forms for the optimal quotes. We will see that in a suitable limit these are exactly those found in the paper of Guéant [51].

Proposition 3.5.1 *Assuming that the eigenpair of Proposition 3.4.1 are a good approximation to equivalent expressions for the operator* $\mathcal{L}_0$*, the optimal quotes for the market maker in the control problem (3.3) are*

$$\delta^a = \frac{1}{k} - \sqrt{\tfrac{\gamma\sigma^2 e}{2A\Delta(k-\gamma\Delta)}}\left(q-\frac{\Delta}{2}\right),$$

$$\delta^b = \frac{1}{k} + \sqrt{\tfrac{\gamma\sigma^2 e}{2A\Delta(k-\gamma\Delta)}}\left(q+\frac{\Delta}{2}\right).$$

Proof Write $\nu := \frac{\beta}{\alpha} = \frac{1}{2}\gamma\sigma^2\frac{e(k-\gamma\Delta)}{A\Delta^3}$. Then we have $\theta(q) = \frac{1}{\eta}\log\{g(q)\} = \frac{\Delta}{k-\gamma\Delta}\log\{g(q)\}$ and so taking g as in Proposition 3.4.1

$$g(q) = \frac{\nu^{\frac{1}{4}}}{\sqrt{2\pi}}\exp\left(-\sqrt{\tfrac{\nu}{2}}q^2\right),$$

and so

$$\theta(q) = \tfrac{\Delta}{k-\gamma\Delta}\left(\log\left(\frac{\nu^{\frac{1}{4}}}{\sqrt{2\pi}}\right) - \frac{\sqrt{\nu}}{2}q^2\right).$$

Differentiating, we find

$$\theta'(q) = -\tfrac{\Delta}{k-\gamma\Delta}\sqrt{\nu}q, \qquad \theta''(q) = -\tfrac{\Delta}{k-\gamma\Delta}\sqrt{\nu}.$$

In the proof of Proposition 3.2.1 we noted that the optimal quotes were given by

$$\delta^a = \frac{1}{k}+\theta'-\tfrac{\Delta}{2}(\theta''-\gamma(\theta')^2) = \frac{1}{k}-\frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}q-\tfrac{\Delta}{2}\left(-\frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}-\gamma\frac{\Delta^2}{(k-\gamma\Delta)^2}\nu q^2\right),$$

$$\delta^b = \frac{1}{k}-\theta'-\tfrac{\Delta}{2}(\theta''-\gamma(\theta')^2) = \frac{1}{k}+\frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}q-\tfrac{\Delta}{2}\left(-\frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}-\gamma\frac{\Delta^2}{(k-\gamma\Delta)^2}\nu q^2\right),$$

which to a linear approximation gives

$$\delta^a = \frac{1}{k} - \frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}\left(q-\frac{\Delta}{2}\right),$$

$$\delta^b = \frac{1}{k} + \frac{\Delta}{k-\gamma\Delta}\sqrt{\nu}\left(q+\frac{\Delta}{2}\right).$$

Substituting $\nu = \frac{1}{2}\gamma\sigma^2\frac{e(k-\gamma\Delta)}{A\Delta^3}$ and performing some straightforward rearrangement gives the forms required.

$\square$

Remark 3.5.2 *In the small $\frac{\gamma\Delta}{k}$ limit these quotes are closely in agreement with the model of Guéant [51]. To see this, working on the ask side (the bid side is* *identical), the quote can be written*

$$\delta^a = \frac{1}{k} - \sqrt{\frac{\gamma\sigma^2}{2A\Delta k}\frac{ek}{(k-\gamma\Delta)}}\left(q - \frac{\Delta}{2}\right).$$

In Guéant [51], the form

$$\delta^a = \frac{1}{\gamma\Delta}\ln\left(1 + \frac{\gamma\Delta}{k}\right) - \left(q - \frac{\Delta}{2}\right)\sqrt{\frac{\sigma^2\gamma}{2A\Delta k}\left(1 + \frac{\gamma\Delta}{k}\right)^{\left(1+\frac{k}{\gamma\Delta}\right)}}$$

is found, and if we take $\frac{\gamma\Delta}{k}$ to be small, then $\frac{1}{\gamma\Delta}\ln\left(1 + \frac{\gamma\Delta}{k}\right) \approx \frac{1}{k}$ and $\left(1 + \frac{\gamma\Delta}{k}\right)^{\left(1+\frac{k}{\gamma\Delta}\right)} \approx \frac{e(k+\gamma\Delta)}{k}$ so that this becomes

$$\delta^a \approx \frac{1}{k} - \left(q - \frac{\Delta}{2}\right)\sqrt{\frac{\sigma^2\gamma}{2A\Delta k}\frac{e(k+\gamma\Delta)}{k}}.$$

Now for small $\frac{\gamma\Delta}{k}$, we also have that $\frac{k}{k-\gamma\Delta} \approx 1 \approx \frac{k+\gamma\Delta}{k}$ and so we find that our solution matches very closely Guéant's approximation in this case.

Corollary 3.5.3 *With the quotes as given in Proposition 3.5.1 the bid ask spread is a constant*

$$\delta^a(q) + \delta^b(q) = \frac{2}{k} + \Delta\sqrt{\frac{\gamma\sigma^2 e}{2A\Delta(k-\gamma\Delta)}}.$$

Again we can see that as $\frac{\gamma\Delta}{k} \to 0$ this could also be written as

$$\delta^a(q) + \delta^b(q) = \frac{2}{k} + \Delta\sqrt{\frac{\gamma\sigma^2}{2kA_\xi}}.$$

3.6 The Dynamics of the Inventory Process

We can also give an approximate form of the dynamics of the inventory process under the control, showing that it behaves as an Ornstein–Uhlenbeck process.

Proposition 3.6.1 *Under the optimal control found in Proposition 3.5.1 the in-*

ventory level is an Ornstein–Uhlenbeck process.

Proof The inventory process follows

$$dq_t = \Delta\{\Lambda^b(\delta^b_t) - \Lambda^a(\delta^a_t)\}dt + \Delta^2\{\Lambda^a(\delta^a_t) + \Lambda^b(\delta^b_t)\}dB^{(1)}_t. \tag{3.13}$$

We need to show that there are constants $c_1, c_2 > 0$ such that the drift $\Lambda^b(\delta^b) - \Lambda^a(\delta^a) = -c_1 q$ and the volatility $\Lambda^a(\delta^a) + \Lambda^b(\delta^b) = c_2$ is constant.

Taking $\nu = \frac{1}{2}\gamma\sigma^2 \frac{e(k-\gamma\Delta)}{A\Delta^3}$ as above we have

$$\begin{aligned}
\Lambda^a(\delta^{a,*}) = Ae^{-k\delta^a} &= A\exp\left(-1 + \frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q - \frac{\Delta}{2}\right)\right)\\
&= Ae^{-1}\exp\left(\frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q - \frac{\Delta}{2}\right)\right)\\
&\approx Ae^{-1}\left(1 + \frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q - \frac{\Delta}{2}\right)\right),
\end{aligned}$$

$$\begin{aligned}
\Lambda^b(\delta^{b,*}) = Ae^{-k\delta^b} &= A\exp\left(-1 - \frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q + \frac{\Delta}{2}\right)\right)\\
&= Ae^{-1}\exp\left(-\frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q + \frac{\Delta}{2}\right)\right)\\
&\approx Ae^{-1}\left(1 - \frac{\Delta k}{k-\gamma\Delta}\sqrt{\nu}\left(q + \frac{\Delta}{2}\right)\right).
\end{aligned}$$

Then we find

$$\begin{aligned}
\Lambda^b(\delta^{b,*}) - \Lambda^a(\delta^{a,*}) &\approx Ae^{-1}\left(-\frac{2\Delta k}{k-\gamma\Delta}\sqrt{\nu}q\right)\\
&= -\frac{2A\Delta k}{e(k-\gamma\Delta)}\sqrt{\nu}q,
\end{aligned}$$

and

$$\Lambda^b(\delta^{b,*}) + \Lambda^a(\delta^{a,*}) \approx Ae^{-1}\left(2 - \frac{\Delta^2 k}{k-\gamma\Delta}\sqrt{\nu}\right),$$

which have the required form.

□

The results of this chapter, whilst interesting, suffer from the significant disadvantage that they are based on an approximation of an operator that we have not fully justified. The difficulties in the non-linearity arise due to the fact that our control affects the Brownian term in the inventory process, making our resulting PDEs fully non-linear (non-linear in ∂_{qq}) rather than semi-linear (linear in ∂_{qq}, non-linear in ∂_q).

Given the manner in which the results match closely those in the literature and in Chapter 2, we suspect that with further work it would be possible to give some reasonable conditions under which this approximation could be justified. Rather we have chosen in this thesis to put rigorous arguments into Chapter 4 where we will consider a slightly simpler model that is very closely related to this one, but captures neatly the essence of the market making problem.

3.7 Summary

In this chapter we have formulated and studied a new continuous model for market making that builds naturally upon and retains the key features of the model of Chapter 2. Although it was necessary to make some approximations along the way, the resulting linear PDE appears to give very natural and interpretable solutions to the market making problem and sets the scene for the model of Chapter 4, where we will work with greater rigour.

Chapter 4

A Continuous Model with a Single Control Variable

In this chapter we adapt the model of Chapter 3 to give a rigorous solution to a continuous time and space version of the original problem of Chapter 2. The model in this chapter allows us to give a clear interpretation of the long-run dynamics and optimal control and we also solve this numerically to find the optimal strategy.

Up until now we have worked with a control consisting of two variables, δ^a and δ^b. But in all of the models we have studied, in particular those of Avellaneda and Stoikov [1] and Guéant [51] [52] as well as the models of Chapters 2 and 3 we have noted the tendency of the optimal quotes to move up and down roughly in unison. We have also seen that market making is primarily a problem of inventory management. Broadly speaking, the market maker seeks to manage their inventory by incentivising or discouraging buy and sell orders with the aim of bringing their inventory levels closer to 0.

This leads us to the idea that the control in the problem may be captured effectively with just one variable instead of two and so in this chapter we modify the continuous setting of Chapter 3 and assume that the market maker may directly control the drift of the inventory process. For any choice of the drift there will be infinitely many ways in which to produce that drift with various choices of δ^a and

δ^b. Each choice will lead to a certain long run profit and the first thing we will do in Section 4.1 is to show that for a given drift there is a unique combination of δ^a and δ^b that will maximise the long run profit. So rather than choosing the quotes as two independent controls, the market maker can instead simply control the drift of the inventory process and note that this automatically implies a choice of δ^a and δ^b.

Then in Section 4.1.1 we will reformulate the control problem so that we may write it down entirely in terms of the drift and restate it in a form suitable for the risk-sensitive control framework described by Nagai in [72]. To apply these results we will need to make an assumption that the volatility of the inventory process is constant, a choice that seems reasonable at least in some sensible parameter regimes.

Since the early work of Jacobsen [61], followed by Whittle [94], there has been much work deriving a full theory of the class of stochastic control problems whose performance criteria are exponential functions of quadratic forms. This theory of risk sensitive control has received a great deal of attention because of the link it provides between stochastic and deterministic approaches to disturbances in control systems. We base our results on those of Nagai [72], but see also for example [8], [36] and [39] and the references contained therein. Since its development this approach has been applied by a number of authors to tackle a variety of problems, including many in mathematical finance, for example in [12], [40] and [62]. To the best of our knowledge it has not yet been applied to market making problems.

In fitting the original modelling assumptions of the papers of Guéant ([51] [52]) into this framework we faced a number of challenges. In particular, the tail behaviour of the demand functions $\Lambda^a(\delta^a) = Ae^{-k\delta^a}, \Lambda^b(\delta^b) = Ae^{-k\delta^b}$ mean that we are not able to satisfy a key assumption of Nagai [72]. Although we suspect that these results are adaptable, the challenge seems deep and would require the adaptation of some fundamental results relating to non-linear PDEs and so is beyond the scope of this thesis. So instead in Section 4.3 we work with modified versions of the demand functions in a way that is economically non-impactful and allows us to make progress.

Then in Section 4.4 we are able to write down a result about the asypmtotic behaviour of the system and of the long run value function. We conjecture that we can go even further and relate the long-run behaviour and the value of market making per unit time to a pair $(v(q), \chi)$ satisfying a Bellman equation of ergodic type and closely related to the Schrodinger operators of our quantum harmonic market maker of Section 2.3.

In Section 4.5 we solve the resulting PDE numerically and consider a method for finding the long-run value of market making χ numerically so that we could understand the sensitivity of the value to changes to various parameters in the problem without needing to repeatedly resolve the PDEs. Finally, we find that the exponential integral that we are trying to estimate has large deviation effects and propose a measure change that allows us to compensate for these and to find an efficient way of computing the value of χ.

4.1 Replacing δ^a and δ^b with a single control variable

We recall that in Chapter 3 we motivated a model including the following inventory and wealth processes

$$dq_t = \Delta\{\Lambda^b(\delta_t^b) - \Lambda^a(\delta_t^a)\}dt + \zeta(\delta_t^a, \delta_t^b)\, dB_t^{(1)},$$

$$dW_t = \Delta\{\delta_t^a \Lambda^a(\delta_t^a) + \delta_t^b \Lambda^b(\delta_t^b)\}dt + \sigma q_t\, dB_t^{(2)},$$

where $\zeta^2(\delta^a, \delta^b) = \Delta^2\{\Lambda^a(\delta^a) + \Lambda^b(\delta^b)\}$. As usual we work in the case $\Lambda^a(\delta^a) = Ae^{-k\delta^a}$, $\Lambda^b(\delta^b) = Ae^{-k\delta^b}$ although we also consider the more general case briefly below.

We begin by showing that if we fix a value μ for the drift of the inventory process that there is a single optimal choice for the corresponding δ^a and δ^b that will maximise the drift of the wealth process.

Proposition 4.1.1 *In order to maximise the profit rate $\Delta(A\delta^a e^{-k\delta^a} + A\delta^b e^{-k\delta^b})$ subject to a fixed drift $\mu = \Delta(Ae^{-k\delta^b} - Ae^{-k\delta^a})$ the market maker should set their* *bid and ask quotes using*

$$\delta^a = \frac{1}{k} + \frac{1}{k}\sinh^{-1}\left(\frac{e\mu}{2A}\right), \tag{4.1}$$

$$\delta^b = \frac{1}{k} - \frac{1}{k}\sinh^{-1}\left(\frac{e\mu}{2A}\right). \tag{4.2}$$

This results in an overall bid-ask spread of $\delta^a + \delta^b = \frac{2}{k}$.

Proof Introducing a Lagrange multiplier λ we set

$$f(\delta^a, \delta^b, \lambda) = \Delta(A\delta^a e^{-k\delta^a} + A\delta^b e^{-k\delta^b}) + \lambda(\Delta(Ae^{-k\delta^b} - Ae^{-k\delta^a}) - \mu).$$

We have $\frac{\partial f}{\partial \delta^a} = A\Delta e^{-k\delta^a}(1 - k\delta^a + k\lambda)$ and $\frac{\partial f}{\partial \delta^b} = A\Delta e^{-k\delta^b}(1 - k\delta^b - k\lambda)$, so that setting $\frac{\partial f}{\partial \delta^b} = 0$ and $\frac{\partial f}{\partial \delta^a} = 0$ yields

$$1 - k\delta^b - k\lambda = 0,$$

$$1 - k\delta^a + k\lambda = 0,$$

and adding these gives the optimal spread of $\delta^a + \delta^b = \frac{2}{k}$. Solving these simultaneously with the constraint $\Delta(Ae^{-k\delta^b} - Ae^{-k\delta^a}) = \mu$ leads directly to the equations

$$\sinh(k\delta^a - 1) = \frac{e\mu}{2A\Delta},$$

$$\sinh(1 - k\delta^b) = \frac{e\mu}{2A\Delta},$$

which rearrange to give (4.1) and (4.2).

Further, some straightforward algebra shows that for these choices of δ^a and δ^b we have

$$\frac{\partial^2 f}{(\partial \delta^a)^2} = \frac{\partial^2 f}{(\partial \delta^b)^2} = -kA\Delta e^{\lambda k - 1} < 0,$$

and

$$\frac{\partial^2 f}{(\partial \delta^a)^2}\frac{\partial^2 f}{(\partial \delta^b)^2} - \left(\frac{\partial^2 f}{\partial \delta^a \partial \delta^b}\right)^2 = k^2 A^2 \Delta^2 e^{2(\lambda k - 1)} > 0,$$

and so these choices do indeed give a maximum.

□

Remark 4.1.2 *The bid ask spread found here is a fixed constant that does not depend on μ. This further supports our intuition from the paper of Avellaneda and Stoikov [1] that optimal market making quotes tend to move up and down in unison.*

Remark 4.1.3 *We could also try to apply this logic with a more general choice of the control and a similar argument finds that we would take a bid-ask spread of*

$$\delta^a + \delta^b = -\frac{\Lambda^a(\delta^a)}{\Lambda'^a(\delta^a)} - \frac{\Lambda^b(\delta^b)}{\Lambda'^b(\delta^b)}.$$

To find the actual quotes we would solve this simultaneously with the constraint

$$\Delta(\Lambda^b(\delta^b) - \Lambda^a(\delta^a)) = \mu.$$

4.1.1 Reformulating the control problem in terms of the drift

Proposition 4.1.1 tells us that having fixed a drift μ, there is a single choice of δ^a and δ^b that maximises the drift of the wealth process, and so next we reformulate the SDEs for the inventory and wealth processes in terms of a control process μ_t and without any direct dependence on δ^a_t and δ^b_t.

Proposition 4.1.4 *Continuing to work in the case $\Lambda^a(\delta^a) = Ae^{-k\delta^a}$, $\Lambda^b(\delta^b) = Ae^{-k\delta^b}$, using the optimal choices of δ^a and δ^b from Proposition 4.1.1 the inventory and wealth process of the problem of Chapter 3 can be written as*

$$dq_t = \mu_t \, dt + \frac{2A\Delta^2}{ke}\sqrt{x_t^2+1}\, dB_t^{(1)}, \tag{4.3}$$

and

$$dW_t = \frac{2A\Delta}{ke}\left\{\sqrt{x_t^2+1} - x_t \sinh^{-1} x_t\right\} dt + \sigma q_t \, dB_t^{(2)}, \tag{4.4}$$

where $x_t = \frac{e\mu_t}{2A\Delta}$.

Proof Writing $x_t = \frac{e\mu_t}{2A\Delta}$, we have that the optimal choices of δ_t^a and δ_t^b from Proposition 4.1.1 satisfy

$$k\delta_t^b = 1 - \sinh^{-1} x_t, \qquad k\delta_t^a = 1 + \sinh^{-1} x_t.$$

Substituting into the profit rate gives

$$\begin{aligned}\Delta\left(A\delta_t^a e^{-k\delta_t^a} + A\delta_t^b e^{-k\delta_t^b}\right) &= \frac{A\Delta}{ke}\left\{\left(1 - \sinh^{-1} x_t\right) e^{\sinh^{-1} x_t} + \left(1 + \sinh^{-1} x_t\right) e^{-\sinh^{-1} x_t}\right\} \\ &= \frac{2A\Delta}{ke}\left\{\sqrt{x_t^2 + 1} - x_t \sinh^{-1} x_t\right\},\end{aligned}$$

where we have used in the final line the facts that $e^{\sinh^{-1} x_t} = \sqrt{x_t^2 + 1} + x_t$ and $e^{-\sinh^{-1} x_t} = \sqrt{x_t^2 + 1} - x_t$. Similarly, the volatility of the inventory process may be written as

$$\begin{aligned}\Delta^2\left(Ae^{-k\delta^a} + Ae^{-k\delta^b}\right) &= \frac{A\Delta^2}{ke}\left\{e^{\sinh^{-1} x_t} + e^{-\sinh^{-1} x_t}\right\} \\ &= \frac{2A\Delta^2}{ke}\left\{\sqrt{x_t^2 + 1}\right\},\end{aligned}$$

and so the result follows.

$\square$

Remark 4.1.5 *With this choice of control we have observed that*

$$\zeta(\delta_t^a, \delta_t^b) = \Delta^2\left(Ae^{-k\delta^a} + Ae^{-k\delta^b}\right) = \frac{2A\Delta^2}{ke}\sqrt{1 + \left(\frac{e\mu_t}{2A\Delta}\right)^2}.$$

Then $\zeta(\delta_t^a, \delta_t^b)$ can be well approximated by a constant if $\frac{e\mu_t}{2A\Delta} << 1$.

In order to apply the results of Nagai [72] we will need to make the assumption that the inventory process has constant volatility. In practice this constant could be chosen based on historical data. A possible direction for future work could be to consider the case where this volatility parameter depends on the optimal control. We could try to set up a fixed point problem where we would hope to find a choice of constant parameter ζ_0, equal to the average of $\zeta(\delta_a, \delta_b)$ over the solution when that parameter is chosen as the control. Although we expect this may be possible, without further investigation there is no guarantee that this fixed point exists so for now we proceed under the following assumption of constant volatility.

Assumption 4.1.6 *The volatility of the inventory process $\zeta(\delta^a, \delta^b) = \zeta^0$, a constant.*

4.2 Defining a New Control Problem

In this section we formulate formally the problem we will solve rigorously in the rest of the chapter. We first state this as Problem 4.2.1 in a form clearly motivated by the work above and then as Problem 4.2.2 in the form to which we can apply results from [72]. In Proposition 4.2.3 we show that Problem 4.2.1 and 4.2.2 are equivalent and then in Lemma 4.2.4 we show that Problem 4.2.1 is well-posed.

Problem 4.2.1 *We consider the problem of maximising over controls $(\mu_t)_{0\leq t\leq T}$ the following risk-sensitive expected growth rate per unit time*

$$J(w, q, \mu) = \liminf_{T\to\infty} -\tfrac{1}{\gamma T} \log \mathbb{E}_{w,q}[e^{-\gamma W_T(\mu)}].$$

Here W_T is given by the integral form of (4.4)

$$W_T = w + \int_0^T \phi(\mu_s)ds + \int_0^T \sigma q_s dB_s^{(2)},$$

where, taking $x = \frac{e\mu}{2A\Delta}$,

$$\phi(\mu_t) = \frac{2A\Delta}{ke}\{\sqrt{x_t^2+1} - x_t \sinh^{-1} x_t\}.$$

We also assume that q_t is a process as in (4.3) given by

$$dq_t = \mu_t dt + \zeta^0 dB_t^{(1)}$$

for a constant ζ^0 and that μ_t is a progressively measurable control that satisfies

$$\mathbb{E}[e^{\frac{1}{2}\int_0^T \gamma^2\sigma^2 q_t^2 dt}] < \infty. \tag{4.5}$$

Problem 4.2.2 *Choose* $(\mu_s)_{0\le s\le T}$ *to minimise*

$$\limsup_{T\to\infty}\frac{1}{\gamma T}\log\mathbb{E}[e^{-\gamma\int_0^T(\phi(\mu_s)-\frac{1}{2}\gamma\sigma^2q_s^2)ds}],$$

where, taking $x_t=\frac{e\mu_t}{2A\Delta}$ *we have*

$$\phi(\mu_s)=\frac{2A\Delta}{ke}\{\sqrt{x_s^2+1}-x_s\sinh^{-1}x_s\}$$

and

$$dq_t=\mu_t dt+\zeta^0\,dB_t^{(1)}.$$

Proposition 4.2.3 *Problem 4.2.1 and Problem 4.2.2 are equivalent.*

Proof Beginning with the value function of Problem 4.2.1

$$\begin{aligned}\mathbb{E}_{w,q}\left[e^{-\gamma W_T}\right]&=\mathbb{E}_{w,q}\left[e^{-\gamma\left(w+\int_0^T\phi(\mu_s)ds+\int_0^T\sigma q_s dB_s^{(2)}\right)}\right]\\&=\mathbb{E}_{w,q}\left[e^{-\gamma\left(w+\int_0^T\phi(\mu_s)ds\right)+\frac{1}{2}\gamma^2\sigma^2\int_0^T q_s^2ds}\times e^{-\gamma\int_0^T\sigma q_s dB_s^2-\frac{1}{2}\gamma^2\sigma^2\int_0^T q_s^2 ds}\right].\end{aligned}$$

Now, the condition (4.5) means that we can apply Novikov's condition to give that the process

$$Z_T:=e^{-\gamma\int_0^T\sigma q_s dB_s^2-\frac{1}{2}\gamma^2\sigma^2\int_0^T q_s^2ds}$$

is a martingale. So by the Girsanov Theorem we can define a change of measure by $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}|_{\mathcal{F}_T}=Z_T$ so that we have

$$\mathbb{E}_{w,q}[e^{-\gamma W_T}]=\tilde{\mathbb{E}}_{w,q}\left[e^{-\gamma\left(w+\int_0^T\phi(\mu_s)ds\right)+\frac{1}{2}\gamma^2\sigma^2\int_0^T q_s^2ds}\right],$$

and $\tilde{B}^{(2)}:=B_t^{(2)}+\gamma\sigma\int_0^t q_s ds$ is a $\tilde{\mathbb{P}}$ Brownian motion. Under $\tilde{\mathbb{P}}$ the q_t dynamics are unchanged and the dynamics for W_t are given by

$$\begin{aligned}dW_t&=\phi(\mu_t)dt+\sigma q_t dB_t^{(2)}\\&=(\phi(\mu_t)-\gamma\sigma^2q_t^2)dt+\sigma q_t d\tilde{B}_t^{(2)}.\end{aligned}$$

Our new objective is to maximise

$$J(w, q, \mu, T) = -\liminf_{T\to\infty} \frac{1}{\gamma T} \log \tilde{\mathbb{E}}_{w,q}[e^{-\gamma(\int_0^T (\phi(\mu_s)-\gamma\sigma^2 q_s^2)ds)}] + w, \tag{4.6}$$

which is clearly equivalent to minimising

$$\limsup_{T\to\infty} \frac{1}{\gamma T} \log \mathbb{E}[e^{-\gamma \int_0^T (\phi(\mu_s)-\gamma\sigma^2 q_s^2)ds}].$$

□

Lemma 4.2.4 *The control problem, Problem 4.2.1, is well-posed, that is, there exists admissible controls $(\mu_t)_{t\geq 0}$ such that $J(w, q, \mu) > -\infty$.*

Proof We consider the case where the strategy $\mu_t = -cq_t$ is used, for some $c > 0$. We will show that there exists $c > 0$ sufficiently large that (4.5) holds. Note that under this assumption, q is an Ornstein-Uhlenbeck process.

In this case, the process q_t depends only on the Brownian motion $B^{(1)}$, and since $B^{(1)}$ is independent of $B^{(2)}$, and under the assumption we can rewrite the expectation ocurring in J as

$$\mathbb{E}[e^{-\gamma \int_0^T (\phi(\mu_s)-\frac{1}{2}\gamma\sigma^2 q_s^2)\, ds}],$$

as we have shown in Proposition 4.2.3.

Since $\phi(x) \sim x \log x$ for large values of x, it will be sufficient to prove that given $\hat{\gamma} > 0$, there exists $c, \hat{\lambda} > 0$ such that

$$\frac{1}{T} \log \mathbb{E}[e^{\hat{\gamma} \int_0^T q_s^2\, ds}] \leq \hat{\lambda}$$

for all T sufficiently large.

We now fix $T > 0$, and $N \in \mathbb{N}$. Let $\delta_N := \frac{1}{N}$ and without loss of generality, we may assume that $T = K\delta_N$ for some K. Write $t_i^N := i\delta_N$ for $i = 0, \ldots, K$.

We write $I_N := \delta_N \sum_{i=0}^{K-1} q_{t_i^N}^2$ and note that $I_N \to \int_0^T q_s^2\, ds$ almost surely as $N \to \infty$ due to the continuity of the paths of q.

We aim to show that $\mathbb{E}[e^{\hat{\gamma} I_N}] < \infty$ for sufficiently large c. Note that since q is an Ornstein-Uhlenbeck process, conditional on $q_{t_0^N}, \ldots, q_{t_i^N}$, $q_{t_{i+1}^N}$ has a Gaussian distribution with mean $q_{t_i^N} e^{-c\delta_N}$ and variance $\frac{(\zeta^0)^2}{c}\left(1 - e^{-c\delta_N}\right)$.

We use the fact that if $Z \sim N(0,1)$ and $\lambda < \frac{1}{2}$, then

$$\mathbb{E}\left[e^{\lambda(Z+\mu)^2}\right] = \frac{e^{\lambda\mu^2(1-2\lambda)^{-\frac{1}{2}}}}{(1-2\lambda)^{\frac{1}{2}}}. \tag{4.7}$$

In particular, writing $\eta_N := e^{-c\delta_N}$ and $\varsigma_N^2 := \frac{\sigma^2}{2c}(1-\eta_N)$, we get:

$$\begin{aligned}
\mathbb{E}\left[e^{\tilde{\xi} q^2_{t_i^N}} | q_{t_0^N}, \ldots, q_{t_{i-1}^N}\right] &= \mathbb{E}\left[\exp\left\{\tilde{\xi}\left(q_{t_{i-1}^N}\eta_N + \varsigma_N Z\right)^2\right\} | q_{t_0^N}, \ldots, q_{t_{i-1}^N}\right] \\
&= \mathbb{E}\left[\exp\left\{\tilde{\xi}\varsigma_N^2\left(q_{t_{i-1}^N}\frac{\eta_N}{\varsigma_N} + Z\right)^2\right\} | q_{t_0^N}, \ldots, q_{t_{i-1}^N}\right] \\
&= \left(1 - 2\tilde{\xi}\varsigma_N^2\right)^{-\frac{1}{2}} \exp\left\{\tilde{\xi}\eta_N^2 q^2_{t_{i-1}^N}\left(1 - 2\tilde{\xi}\varsigma_N^2\right)^{-1}\right\},
\end{aligned}$$

where $Z \sim N(0,1)$ is a Gaussian random variable, independent of $q_{t_0^N}, \ldots, q_{t_{i-1}^N}$. It follows that we can write

$$\mathbb{E}\left[e^{\hat{\gamma} I_N} | q_{t_0^N}, \ldots, q_{t_k^N}\right] = e^{\hat{\gamma} I_{k-1}} \theta_k \exp\left\{\tilde{\xi}_k q^2_{t_k^N}\right\},$$

where $\theta_N = 1$, $\tilde{\xi}_N = \hat{\gamma}\delta_N$ and

$$\tilde{\xi}_k = \tilde{\xi}_{k+1}\left(1 - 2\theta_{k+1}\varsigma_N^2\right)^{-\frac{1}{2}}, \tag{4.8}$$

$$\theta_k = \hat{\gamma}\delta_N + \theta_{k+1}\eta_N^2\left(1 - 2\theta_{k+1}\varsigma_N^2\right)^{-1}, \tag{4.9}$$

provided that $2\theta_{k+1}\varsigma_N^2 < 1$.

We consider fixed points of equation (4.9), that is, values $\theta^{N,*}$ such that $2\theta^{N,*}\varsigma_N^2 < 1$ and

$$\theta^{N,*} = \hat{\gamma}\delta_N + \theta^{N,*}\eta_N^2\left(1 - 2\theta^{N,*}\varsigma_N^2\right)^{-1},$$

which is equivalent to

$$2(\theta^{N,*})^2\varsigma_N^2 + (\eta_N^2 - 1 - 2\hat{\gamma}\varsigma_N^2)\theta^{N,*} + \hat{\gamma}\delta_N = 0. \tag{4.10}$$

Real roots to this quadratic equation exist when

$$(\eta_N^2 - 1 - 2\hat{\gamma}\varsigma_N^2)^2 - 8\varsigma_N^2\hat{\gamma}\delta_N \geq 0.$$

Using the approximations (noting that δ_N is small) $\eta_N^2 - 1 \approx -2c\delta_N$ and $\varsigma_N^2 \approx \frac{1}{2}(\zeta^0)^2\delta_N$, we conclude that there exists N_0 depending on $\varepsilon > 0$ such that (4.9) has a fixed point whenever

$$\left(2c + \frac{1}{2}(\zeta^0)^2\right)^2 \geq 4(\zeta^0)^2\hat{\gamma} + \varepsilon,$$

or in particular if

$$c \geq \zeta^0\sqrt{\hat{\gamma} + \varepsilon'} - \frac{(\zeta^0)^2}{4},$$

for a suitable constant ε'.

Observing that

$$\theta \mapsto g(\theta) := \hat{\gamma}\delta_N + \theta\eta_N^2\left(1 - 2\theta\varsigma_N^2\right)^{-1}$$

defines an increasing function with $g'(\theta) < 1$ provided $\theta < \varsigma_N^{-2}/2$, we conclude that if $\theta_0 < \theta^{N,*}$ and $\theta^{N,*} < \varsigma_N^{-2}/2$ then θ_k given by (4.9) is an increasing sequence, converging to $\theta^{N,*}$.

Dividing (4.10) by ς_N^2, and using the same approximation for large N as above, we see that for N sufficiently large, $\theta^{N,*}$ will approximately solve the quadratic equation

$$2(\theta^*)^2 - \left(\frac{c}{(\zeta^0)^2} + 2\hat{\gamma}\right)\theta^* + \frac{2\hat{\gamma}}{(\zeta^0)^2},$$

from which we can conclude that $\theta^{N,*} < \frac{\varsigma_N^{-2}}{2}$, and hence $\theta_k < \frac{\varsigma_N^{-2}}{2}$ for all k. Moreover, it follows that there exists $\varepsilon'' > 0$ such that $\theta_k \leq \theta^* + \varepsilon''$ uniformly for all N sufficiently large, and all $k = 0, 1, \ldots, N$.

We conclude that $\xi_0 = \Pi_{k=1}^N (1 - 2\theta_k \varsigma_N^2)^{-\frac{1}{2}}$, and thus

$$\begin{aligned}
\frac{1}{T} \log \mathbb{E}\left[e^{\tilde{\xi} q^2_{t_i^N}} | q_{t_0^N}, \dots, q_{t_{i-1}^N} \right] &= \frac{1}{T} \log(\xi_0) \\
&= -\frac{1}{2} \frac{1}{T} \sum_{k=1}^N \log \left(1 - 2\theta_k \varsigma_N^2\right) \\
&\leq \frac{1}{T} \sum_{k=1}^N \left(\theta_k \varsigma_N^2 + \theta_k^2 \varsigma_N^4\right) \\
&\leq (\zeta^0)^2 (\theta^* + \varepsilon'') + o(1),
\end{aligned}$$

which is the desired result.

To conclude, we observe that sending $N \to \infty$ and using Fatou's Lemma, we have

$$\mathbb{E}\left[e^{\hat{\gamma} \int_0^T q_s^2 \, ds} \right] \leq \liminf \mathbb{E}[e^{\hat{\gamma} I_N^N}] \leq e^{T\hat{\lambda}},$$

as required.

$\square$

4.3 Modifying the demand functions $\Lambda^a(\delta^a)$ and $\Lambda^b(\delta^b)$

We are now almost in a position that we may apply the results of Nagai [72] to our problem. In particular we would like to apply Theorems 1.1 and 3.1 of Nagai [72] in order to assert that this problem has a unique solution and to infer properties of the long run behaviour. To do so we need to check that conditions

(1.4)-(1.10),(1.16) and (1.17) of [72] hold.[1] Mostly these are straightforward, however in order to satisfy (1.16) of [72] we require the function[2]

$$Q_0(p) := \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\{\mu p - \phi(\mu)\}, \qquad p \in \mathbb{R}$$

must be bounded above and below by quadratics. However this is not the case without making a slight modification to the model. It seems plausible that in fact this approach could be avoided, however the PDE theory required to adapt to the Nagai work would take us well beyond the scope of this thesis. Indeed Nagai relies on deep PDE results in [65] that make fundamental use of this assumption and so instead in the next section we modify the demand function slightly in a way that has no practical impact but avoids these difficulties.

The essential problem with fitting our case into the Nagai [72] framework arises from the exponential tails of the demand functions $\Lambda^a(\delta^a)$ and $\Lambda^b(\delta^b)$. To work around this we fix a large negative M and make the demand function linear rather than exponential beyond that point. So let us define

$$\Lambda^M(\delta) = \begin{cases} Ae^{-k\delta} & \delta > M \\ Ae^{-kM}(1 + k(M-\delta)) & \delta < M \end{cases} \tag{4.11}$$

where the linear piece of the function has been chosen to ensure that $\Lambda^M(\delta)$ is continuous with a continuous derivative.

This modification we make below has almost no impact on the practical applica-

[1] For the benefit of any reader looking to match closely our work to that of Nagai we include here a table comparing the notation used in Nagai to our own. Note that we work in the $\theta > 0$ case as written in Nagai (corresponding to the risk-averse case, and to our $\gamma > 0$).

Nagai	This Thesis	Nagai	This Thesis
Z_s	μ_s	X_s	q_s
$b(X_t)$	0	$c(X_t, Z_t)$	$\mu(q_t)$
$\phi(X_s, z_s)$	$-(\phi(\mu) - \frac{2A\Delta}{ke})$	$V(X_s)$	$\gamma\sigma^2 q_s^2 - \frac{2A\Delta}{ke}$
θ	γ	a_{ij}	ζ_0^2

[2] Note that in Nagai the function $Q_0(q,p)$ may depend on q, however since our versions do not, we just write $Q_0(p)$.

tion. Indeed, verifying the behaviour of demand in the tails in reality would be very difficult. In any case it would be very rare that the market maker would ever quote at such extreme values. In practice they would most likely simply actively transact with other market makers to reduce their position at the sorts of large inventory levels these quotes would be needed at, rather than passively offering very favourable terms. Thus our modified modelling choice is just as reasonable as the initial one and there is no cost in taking this approach.

Although of little practical importance, moving from exponential to linear demand for these large values will allow us to fit neatly into the framework of Nagai. Before we show that this is possible, we see how the change to the demand functions impacts on the optimal quotes.

Lemma 4.3.1 *The optimal choice of δ^a and δ^b are unchanged by modifying the demand functions from $\Lambda(\delta^a)$ and $\Lambda(\delta^b)$ to $\Lambda^M(\delta^a)$ and $\Lambda^M(\delta^b)$ for values of μ satisfying*

$$|\mu| < -\frac{2A}{e}\sinh(Mk - 1).$$

Proof The particular choice of the linear function $\Lambda^M(\delta)$ in (4.11) has been chosen to make the function continuous with a continuous derivative. As a result we have that $\Lambda^M(\delta) \leq \Lambda(\delta)$ for all δ. This means that to achieve a given level of demand, the market maker must offer a less profitable δ under Λ^M than under Λ. That is, there are only equal or worse options available to the market maker under δ^M compared to Λ. Thus for μ such that the resulting optimal values of δ^a and δ^b are greater than M the optimal choice will be unchanged.

Rearranging (4.1) and (4.2), the strategy is unchanged on the ask side when the optimal $\delta^a > M$. That is when

$$\mu > \frac{2A}{e}\sinh(Mk - 1).$$

On the bid side the strategy is unchanged when $\delta^b > M$, that is

$$\mu < -\frac{2A}{e}\sinh(Mk - 1).$$

Note that since M is large and negative, so is $\sinh(Mk - 1)$. When both of these conditions are satisfied so that

$$|\mu| < -\frac{2A}{e}\sinh(Mk-1),$$

then the optimal strategy will be defined as before.

$\square$

Lemma 4.3.2 *With the capped demand function $\Lambda^M(\delta)$ there is still a unique choice of δ^a and δ^b that maximises the drift in the wealth process for every value of μ.*

Proof By Lemma 4.3.1 the strategy is unchanged for $|\mu| < -\frac{2A}{e}\sinh(Mk-1)$ and so the optimal choice of δ^a and δ^b will be unchanged. So we just consider the case $|\mu| > -\frac{2A}{e}\sinh(Mk-1)$.

Let us consider first the case where μ is large and positive, so that we have

$$\mu > -\frac{2A}{e}\sinh(Mk-1). \tag{4.12}$$

In this case the market maker is trying to rapidly increase their inventory and so we would expect to find that δ^b is very large and negative and δ^a is very large and positive. In that way the market maker's quotes on the bid side are very generous (paying well over the odds for the asset in order to increase their inventory rapidly) and their ask quotes will be very unattractive and should find almost no orders.

Thus the relevant pieces of the demand functions to consider for the bid and ask quotes respectively are

$$\Lambda^b(\delta^b) = Ae^{-kM}(1+k(M-\delta^b)),$$

$$\Lambda^a(\delta^a) = Ae^{-k\delta^a}.$$

Then the constraint on the inventory $\Delta(\Lambda^b(\delta^b) - \Lambda^a(\delta^a)) = \mu$ becomes

$$\Delta(Ae^{-kM}(1+k(M-\delta^b)) - Ae^{-k\delta^a}) = \mu, \tag{4.13}$$

and the profit rate we are looking to optimise, $\Delta\{\delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b)\}$, becomes

$$\Delta(\delta^a Ae^{-k\delta^a} + \delta^b Ae^{-kM}(1 + k(M - \delta^b))).$$

Introducing the Lagrange multiplier λ as before we consider the function

$$\begin{aligned} f(\delta^a, \delta^b, \lambda) =& \Delta(\delta^a Ae^{-k\delta^a} + \delta^b Ae^{-kM}(1 + k(M - \delta^b))) \\ &+ \lambda\Delta(Ae^{-kM}(1 + k(M - \delta^b)) - Ae^{-k\delta^a}) - \mu). \end{aligned}$$

We have $\frac{\partial f}{\partial \delta^a} = A\Delta e^{-k\delta^a}(1 - k\delta^a + k\lambda)$ and $\frac{\partial f}{\partial \delta^b} = A\Delta e^{-kM}(1 + kM - 2k\delta^b - k\lambda)$, so that setting $\frac{\partial f}{\partial \delta^b} = 0$ and $\frac{\partial f}{\partial \delta^a} = 0$ yields the equations

$$\delta^a = \frac{1}{k} + \lambda, \tag{4.14}$$

$$2\delta^b = \frac{1}{k} + M - \lambda. \tag{4.15}$$

Upon substituting these into the constraint (4.13) we find the equation

$$\tfrac{1}{2}k\lambda - e^{kM-1}e^{-\lambda k} = \tfrac{\mu}{A\Delta}e^{kM} - \tfrac{1}{2} - \tfrac{1}{2}kM. \tag{4.16}$$

Now if we let $x = -\lambda k$ then we see that this is of the form $be^x = ax - c$ with $a < 0$ and $b > 0$. Since a straight line of negative slope will intersect the exponential graph exactly once we can deduce that there is a unique solution for λ and hence that the problem of finding the optimal δ^a and δ^b also has a unique solution.

Further, some straightforward algebra shows that for these choices of δ^a and δ^b we have

$$\frac{\partial^2 f}{(\partial \delta^a)^2} = -kA\Delta e^{\lambda k - 1} < 0,$$

$$\frac{\partial^2 f}{(\partial \delta^b)^2} = -2kA\Delta e^{-kM} < 0,$$

and

$$\frac{\partial^2 f}{(\partial \delta^a)^2}\frac{\partial^2 f}{(\partial \delta^b)^2} - \left(\frac{\partial^2 f}{\partial \delta^a \partial \delta^b}\right)^2 = 2k^2A^2\Delta^2 e^{1+\lambda k + kM} > 0,$$

and so these choices do indeed give a maximum.

The case where μ is large and negative is almost identical, with the roles of bid and ask reversed. By similar computations we find that

$$2\delta^a = \frac{1}{k} + M - \lambda, \tag{4.17}$$

$$\delta^b = \frac{1}{k} + \lambda. \tag{4.18}$$

where λ is the solution to

$$\tfrac{1}{2}k\lambda + e^{kM-1}e^{\lambda k} = \tfrac{\mu}{A\Delta}e^{kM} + \tfrac{1}{2} + \tfrac{1}{2}kM. \tag{4.19}$$

□

Next we will show that we can satisfy condition (1.16) in Nagai [72]. To do so we are hoping to find k_1 and k_2 such that

$$-\frac{k_1}{2}\zeta_0^2 p^2 \leq Q_0(p) \leq -\frac{k_2}{2}\zeta_0^2 p^2$$

where

$$Q_0(p) = \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\{\mu p - \phi(\mu)\}, \qquad p \in \mathbb{R}$$

Our strategy is to show first that we can find $m_1, m_2 > 0$ such that

$$-m_1\mu^2 \leq \phi(\mu) \leq -m_2\mu^2. \tag{4.20}$$

As a consequence of Lemma 4.3.1 we note that for $|\mu| < -\frac{2A}{e}\sinh(Mk-1)$ the profit rate $\phi(\mu)$ is also unchanged and so in this region we have

$$\phi(\mu) = \frac{2A\Delta}{ke}\{\sqrt{x^2+1} - x\sinh^{-1}x\},$$

where $x = \frac{e\mu}{2A\Delta}$, and in particular $\phi(0) = \frac{2A\Delta}{ke}$ so it is clear that we cannot satisfy (4.20) in its current form. So we subtract $\frac{2A\Delta}{ke}$ from $\phi(\mu)$ to make sure it passes through the origin and instead look to find $m_1, m_2 > 0$ such that

$$-m_1\mu^2 \leq \phi(\mu) - \frac{2A\Delta}{ke} \leq -m_2\mu^2. \tag{4.21}$$

We also redefine

$$Q_0(p) := \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\left\{\mu p - \left(\phi(\mu) - \frac{2A\Delta}{ke}\right)\right\}, \qquad p \in \mathbb{R}$$

accordingly and so we will plan to apply the results of Nagai to $\phi(\mu) - \frac{2A\Delta}{ke}$ instead.

Remark 4.3.3 *Note the quantity that has been subtracted from $\phi(\mu)$, that is $\frac{2A\Delta}{ke}$, also played an important role in Chapter 3, where we noted that it converged in the limit as $\frac{\gamma\Delta}{k} \to 0$ to the quantity $2A_\xi$, the long-run value of market making found in the riskless world of Section 2.1.*

Although we cannot find a closed form solution for (4.16) we can compute asymptotic results that will allow us to find suitable m_1 and m_2. We note that some of these asymptotic results relate to cases that are not likely to occur in practice, but are nonetheless important to establish the applicability of the results we wish to apply.

Proposition 4.3.4 *With demand modelled by the capped function, asymptotically we find that the optimal quotes are given by*

$$\delta^a \sim \frac{1}{k} + M + \frac{e^{kM}}{\Delta A k}\mu, \qquad \text{as } \mu \to \infty \tag{4.22}$$

$$\delta^b \sim \frac{1}{k} + M - \frac{e^{kM}}{\Delta A k}\mu, \qquad \text{as } \mu \to -\infty, \tag{4.23}$$

and asymptotically as $|\mu| \to \infty$ the profit rate satisfies

$$\phi(\mu) \sim -\frac{e^{kM}}{Ak}\mu^2.$$

Proof By Lemma 4.3.1 the optimal strategy is unchanged for $|\mu| < -\frac{2A}{e}\sinh(Mk-1)$ and so $\phi(\mu)$ will also be unchanged for these values. So we are considering the case $|\mu| > -\frac{2A}{e}\sinh(Mk-1)$.

Let us consider the case where μ is large and positive, so that we have $\mu > -\frac{2A}{e}\sinh(Mk-1)$. As we noted heuristically and is confirmed by (4.14) and (4.15), in this case where we are looking to achieve a very large positive drift we will want to choose δ^a to be very large and positive and the corresponding δ^b to be very large and negative.

The fact that δ^a will be governed by the usual demand function and the quotes in this regime are so wildly uncompetitive that demand on that side of the book will be approximately 0 and the drift will be almost exclusively achieved by orders on the bid side. In the limiting case as $\mu \to \infty$ the market maker's quotes on one side of the book become are filled so rarely that their one-sided quote can be found by finding the δ^b satisfying $\mu = \Delta\Lambda^M(\delta^b)$. That is

$$\mu = \Delta(Ae^{-kM}(1 + k(M - \delta^b)).$$

Straightforward rearrangement gives

$$\delta^b = \frac{1}{k} + M - \frac{e^{kM}}{\Delta Ak}\mu,$$

and the resulting drift in the profit will be

$$\phi^M(\mu) := \delta^b\Delta\Lambda^M(\delta^b) = \left(\frac{1}{k} + M - \frac{e^{kM}}{\Delta Ak}\mu\right)\Delta\mu = \left(M + \frac{1}{k}\right)\mu - \frac{e^{kM}}{Ak}\mu^2.$$

The case of looking for a large negative drift $\mu < 0$ is almost identical. We would look for $-\mu = \Delta\Lambda^M(\delta^b)$ so that

$$-\mu = \Delta(Ae^{-kM}(1 + k(M - \delta^a)),$$

giving

$$\delta^a = \frac{1}{k} + M + \frac{e^{kM}}{\Delta Ak}\mu,$$

and

$$\phi^M(\mu) := \delta^a\Delta\Lambda^M(\delta^a) = -\left(\frac{1}{k} + M + \frac{e^{kM}}{\Delta Ak}\mu\right)\Delta\mu = -\left(M + \frac{1}{k}\right)\mu - \frac{e^{kM}}{Ak}\mu^2.$$

Then we have that $\phi(\mu) \sim -\frac{e^{kM}}{Ak}\mu^2$ as $|\mu| \to \infty$ as required.

$\square$

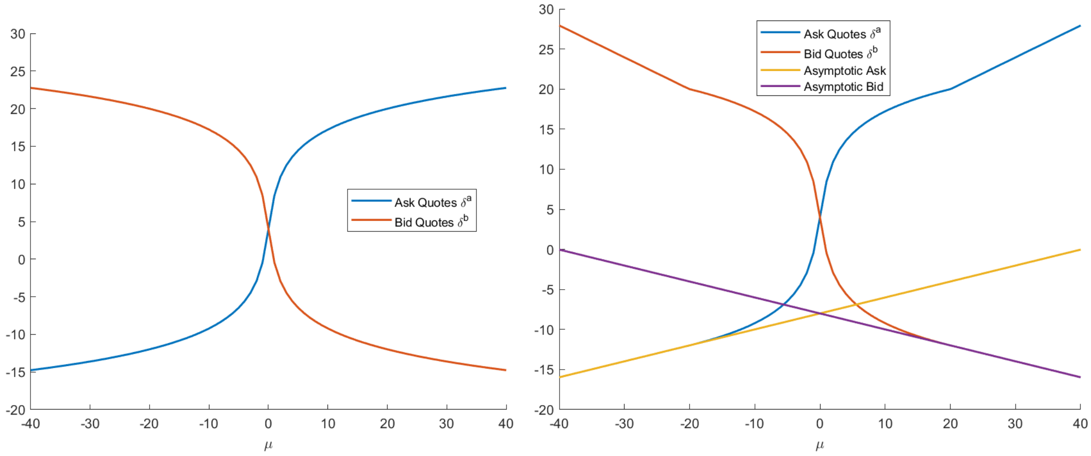


Figure 4-1: The left panel shows the optimal bid and ask quotes with the original demand functions as found in Proposition 4.1.1. The right panel shows the optimal quotes for the capped demand function, computed numerically in the case $|\mu| < -\frac{2A}{e}\sinh(Mk-1)$ (where they differ from the original functions), and using the original quotes otherwise. The straight lines show the asymptotic behaviour. These were computed with $k = 0.25$, $M = -12$, $A = 1$, $\Delta = 1$.

In Figure 4-1, to illustrate the asymptotic behaviour of the quotes we have repeatedly solved equations (4.16) and (4.19) to compute the optimal quotes for $|\mu| < -\frac{2A}{e}\sinh(Mk-1)$ and we have also plotted straight lines to show the asymptotic behaviour given in (4.22) and (4.23).

We are now in a position to show that we can satisfy the necessary conditions in Nagai [72]. In order to do so we will justify the following Proposition.

Proposition 4.3.5 *Let $M < 0$ be a large negative value. With the demand function defined by (4.11), the function*

$$Q_0(p) = \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\left\{\mu p - \left(\phi(\mu) - \frac{2A\Delta}{ke}\right)\right\}, \qquad p \in \mathbb{R} \tag{4.24}$$

may be bounded above and below by quadratics in the sense of condition (1.16) in Nagai [72], where $\phi(\mu)$ is the drift of the resulting wealth process.

We note that we know the following about the profit rate $\phi(\mu)$ in the capped

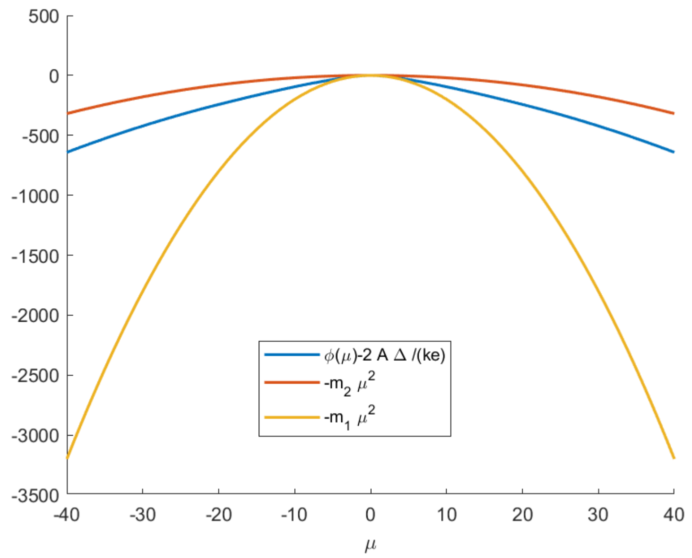


Figure 4-2: A plot of the function $\phi(\mu) - \frac{2A\Delta}{ke}$, computed numerically and sandwiched between $-m_1\mu^2$ and $-m_2\mu^2$ for $m_1 = \frac{1}{k}$ and $\mu_2 = \frac{e^{kM}}{Ak}$.

case

$$\phi(\mu) = \begin{cases} \frac{2A\Delta}{ke}\{\sqrt{x^2+1} - x\sinh^{-1} x\} & |\mu| < -\frac{2A}{e}\sinh(Mk-1) \\ \phi^M(\mu) & |\mu| > -\frac{2A}{e}\sinh(Mk-1) \end{cases} \tag{4.25}$$

where $x = \frac{e\mu}{2A\Delta}$ and we have as $|\mu| \to \infty$

$$\phi^M(\mu) \sim -\frac{e^{kM}}{Ak}\mu^2.$$

Although we cannot give a closed form expression for ϕ^M for all values, we may use the computed optimal quotes to find it numerically via

$$\phi(\mu) = \delta^a\Lambda^a(\delta^a) + \delta^b\Lambda^b(\delta^b).$$

Further, the proof of Proposition 4.3.4 makes it clear that in fact $\phi^M(\mu) \leq -\frac{e^{kM}}{Ak}\mu^2$ for $|\mu| > -\frac{2A}{e}\sinh(Mk-1)$ and it is easy to verify that this is in fact true for all μ and also for $\phi(\mu) - \frac{2A\Delta}{ke}$. We illustrate this by plotting both of these functions in Figure 4-2 along with a quadratic that lies entirely above $\phi(\mu) - \frac{2A\Delta}{ke}$. Finding the quadratic that lies below $\phi(\mu) - \frac{2A\Delta}{ke}$ is much easier. We just need to make sure that the function stays below near zero, but we have a closed form for $\phi(\mu)$

in this region, so to make sure that the quadratic function $-m_1\mu^2$ lies below it we just need to compare their second derivatives. We can easily compute that

$$\frac{d^2\phi}{d\mu^2}\Big|_{\mu=0} = -\frac{1}{k}$$

and so the relevant condition is that we should take

$$m_1 \geq \frac{1}{2k}.$$

So we have found $m_1, m_2 > 0$ such that

$$-m_1\mu^2 \leq \phi(\mu) - \frac{2A\Delta}{ke} \leq -m_2\mu^2,$$

where $m_1 = \frac{1}{2k}$ and $m_2 = \frac{e^{kM}}{Ak}$.

Next, we consider putting these quadratics in functions of a similar form to the Legendre-Fenchel style Q_0 that we need to consider to apply Nagai's results. We can readily compute

$$\begin{aligned} Q_1(p) :=& \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\{\mu p + m_1\mu^2\}, \\ &= \frac{\gamma}{2}\zeta_0^2 p^2 + m_1 \inf_{\mu}\left\{\left(\mu + \frac{p}{2m_1}\right)^2 - \frac{p^2}{4m_1^2}\right\} \\ &= \left(\frac{\gamma}{2}\zeta_0^2 - \frac{1}{4m_1}\right)p^2, \end{aligned}$$

and similarly

$$Q_2(p) := \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\{\mu p + m_2\mu^2\} = \left(\frac{\gamma}{2}\zeta_0^2 - \frac{1}{4m_2}\right)p^2.$$

Thus by properties of Legendre-Fenchel transforms we may conclude that

$$\left(\frac{\gamma}{2}\zeta_0^2 - \frac{1}{4m_2}\right)p^2 \leq Q_0(p) \leq \left(\frac{\gamma}{2}\zeta_0^2 - \frac{1}{4m_1}\right)p^2,$$

where we recall that

$$Q_0(p) = \frac{\gamma}{2}\zeta_0^2 p^2 + \inf_{\mu}\left\{\mu p - \left(\phi(\mu) - \frac{2A\Delta}{ke}\right)\right\}, \qquad p \in \mathbb{R}.$$

For Nagai we require that these coefficients be negative so that

$$\gamma\zeta_0^2 < \frac{1}{2m_1}, \qquad \gamma\zeta_0^2 < \frac{1}{2m_2},$$

with the former condition being the more restrictive. With such a choice it is clear that we have justified Proposition 4.3.5 and so we have satisfied (1.16) of Nagai and may proceed to use their results.

Remark 4.3.6 *There is a further condition (1.17) required in Nagai, and whilst it is clear that the required gradient conditions are satisfied it also requires smoothness of the function Q_0. If we further approximate the demand function Λ^M of (4.11) with a smooth function, for example by taking the values given by (4.11) for $\delta < M$ and $\delta > M+1$ and choosing a smooth interpolation for $M < \delta < M+1$ then we can satisfy this condition whilst not impacting any of our other conclusions.*

4.4 Long-Run Behaviour

We are now in a position to apply the results of Nagai via the following theorem.

Theorem 4.4.1 *Writing*

$$u(t,q) = \frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma\int_0^t(\phi(\mu_s)-\gamma\sigma^2 q_s^2)ds}],$$

and taking Q_0 as defined in (4.24) and ϕ as given by (4.25), then there exists $0 < \alpha < 1$ such that the equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\zeta_0^2 u''(q) + Q_0(u') + \gamma\sigma^2 q^2 - \frac{2A\Delta}{ke} \tag{4.26}$$

has a unique non-negative solution $u \in C^{1+\frac{\alpha}{2},2+\alpha}((0,\infty)\times\mathbb{R}) \cap C([0,\infty)\times\mathbb{R})$. Further there exists an increasing sequence $T_i \subset \mathbb{R}_+$ with $T_i \to \infty$ such that

$u(T_i, q) - u(T_i, 0)$ converges to a function $v \in C^2(\mathbb{R})$ uniformly on each compact set and strongly in $W^1_{2,loc}$ and $\frac{\partial u}{\partial t}(T_i, q)$ to $\chi(q) \in C(\mathbb{R})$ uniformly on each compact set. Moreover $(v(q), \chi(q))$ satisfies

$$\chi(q) + \frac{2A\Delta}{ke} = \frac{1}{2}\zeta_0^2 v''(q) + Q_0(v') + \gamma\sigma^2 q^2.$$

Proof The theorem follows from a direct application of Theorem 1.1 and Lemma 3.1 of [72]. Of the conditions required to apply the result, (1.4)-(1.10) and (1.17) are easily satisfied (noting Remark 4.3.6), and (1.16) follows as a result of the capped demand function we have imposed and Proposition 4.3.5.

□

We would like to take this one step further and assert that the function $\chi(q)$ is in fact a constant χ[3] and so make the following conjecture:

Conjecture 4.4.2 *As $T \to \infty$ the solution $u(t, x)$ to Problem 4.2.2 satisfies:*

1. *$u(t, q) - u(t, 0)$ converges to a function $v(q)$ in $W^1_{2,loc}$ uniformly on each compact set.*
2. *$\frac{\partial u}{\partial t}(t, q) \to \chi$ on each compact set, where $\chi \in \mathbb{R}$.*
3. *The pair (v, χ) is the unique solution of*

$$\chi + \frac{2A\Delta}{ke} = \frac{1}{2}\zeta_0^2 v''(q) + Q_0(v') + \gamma\sigma^2 q^2, \tag{4.27}$$

   where Q_0 is as defined in (4.24).

A full justification of this result, along with our belief that the form of equation (4.27) is very closely related to the principal eigenpair of our quantum harmonic market maker of Section 2.3 and that $(v(q), \chi)$ can be recovered as the principal eigenpair of a Schrödinger operator, is just out of reach. Although our problem is very close to examples given in Section 3.2 of Nagai [72] with these properties which would allow us to use their Theorem 3.4 to prove our Conjecture 4.4.2, the functional analysis required to adapt rather than borrow their results is deep and

[3]In this case equation (4.26) is a 'Bellman equation of ergodic type'.

challenging enough that it must be left for future work. Nonetheless our intuition is that this constant $\chi + \frac{2A\Delta}{ke}$ can be shown rigorously to represent the long-run value of market making per unit time. We expect χ to be negative so that this represents a reduction from the maximum theoretical profit per unit of time $\frac{2A\Delta}{ke}$ as found in Section 2.1.

In the following section we assume this conjecture is true and explore the consequences via a numerical solution of equation (4.27).

4.5 Numerical Solution

We now try looking for a fixed point solution via iteration. That is, we guess a (zero) solution for the equation

$$\frac{1}{2}\zeta_0^2 u''(q) + Q_0(u') + \gamma\sigma^2 q^2 = \chi + \frac{2A\Delta}{ke},$$

where

$$Q_0(p) = \frac{1}{2}\zeta_0^2 p^2 + \inf_{\mu}\left[\mu p - (\phi(\mu) - frac2A\Delta ke)\right].$$

and update u by computing the error and moving a small distance in the direction of the error.

Remark 4.5.1 *We note that the shift by $\frac{2A\Delta}{ke}$ does not significantly impact the underlying analysis. Indeed if we go back to the original definition of*

$$Q_0(p) = \frac{1}{2}\zeta_0^2 p^2 + \inf_{\mu}\left[\mu p - \phi(\mu)\right],$$

then the Q_0 function just shifts by $\frac{2A\Delta}{ke}$ and so we can equivalently use this definition along with the equation

$$\frac{1}{2}\zeta_0^2 u''(q) + Q_0(u') + \gamma\sigma^2 q^2 = \chi,$$

which is how we have set out the work in this section.

In the numerical work of this section we have taken $\gamma = 0.4$, $\zeta_0 = 0.5$, $\sigma = 0.5$, $\Delta = 1$, $A = 1$ and $k = 2$. In Figure 4-3 we plot the optimal drift as a function of

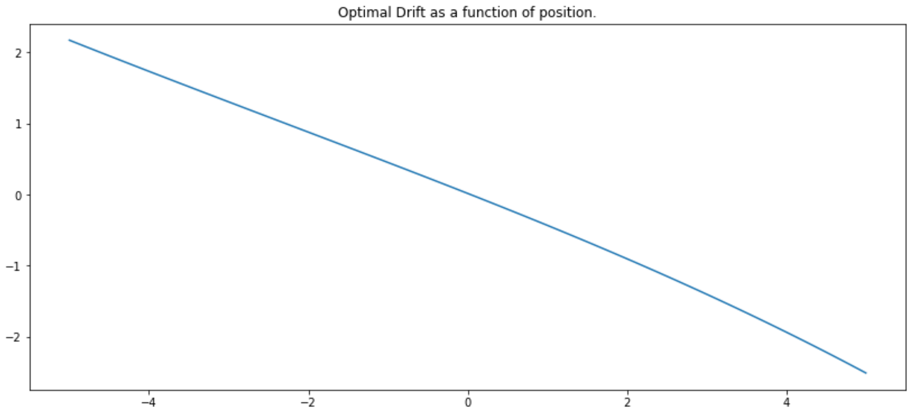


Figure 4-3: A plot of the optimal drift based on numerical work.

q that has been computed in this way.

Once we have identified the optimal strategy, we are able to find the optimal stationary distribution, and the corresponding cost. In particular, we know from above that there is an optimal strategy $u(q_t)$ for given current position q. Then the law of the corresponding process q_t solves

$$dq_t = \zeta_0 \, dB_t^{\mathbb{Q}} + u_t \, dt.$$

In particular, the generator of the stationary measure is given by

$$\mathcal{L}^q f(q) := \frac{d}{dt} \left\{ \mathbb{E}_q^{\mathbb{Q}} \left[f(q_t) \right] \right\} = \frac{1}{2} \zeta_0^2 f''(q) + u(q) f'(q).$$

If we suppose the stationary measure of q has density g, then it follows that we should have

$$\int g(q) \left\{ \frac{1}{2} \zeta_0^2 f''(q) + u(q) f'(q) \right\} dq = 0,$$

and hence (formally, by integration by parts)

$$\frac{1}{2} \zeta_0^2 g''(q) - (u(q) g(q))' = 0.$$

This has solution

$$g(q) = A \exp \left\{ \frac{2}{\zeta_0^2} \int_0^q u(\tilde{q}) \, d\tilde{q} \right\}. \tag{4.28}$$

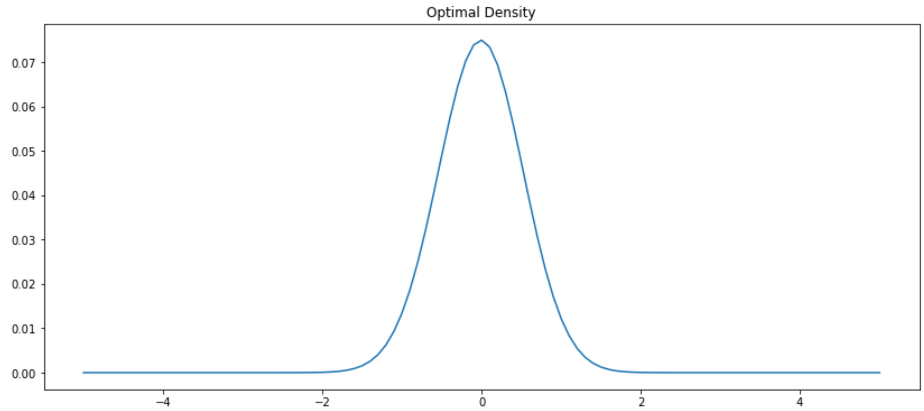


Figure 4-4: A plot of the optimal density $g(q)$ of (4.28) based on numerical work.

In Figure 4-4 we plot this optimal density from our numerical work. Once we have solved for this density, then recalling that our aim is to maximise a function of the form

$$\frac{1}{\gamma}\log\mathbb{E}^{\mathbb{Q}}\left[-\exp\left\{\int_0^T\left(\hat{\gamma}q_t^2+\phi(u_t,q_t)\right)\,dt\right\}\right], \tag{4.29}$$

we can hope to estimate the long-run value per unit time by integrating the density against the value of the integral in (4.29) at each time. If this were possible then we would have an easy way of numerically varying various parameters and recomputing the eigenvalue without having to resolve the PDE at each step. This would be particularly useful in the model of Section 5.3 later where there are many parameters involving the last look feature.

Although we have not experimented extensively with different parameter regimes, we have found with some reasonable choices of parameters that the eigenvalue estimated in this manner differs from the one found directly as the solution of the PDE by a percent or two. The reason for this discrepancy is that this method does not take into account the large deviation effects of the form discussed in Section 1.4.3. In the final section of this chapter we address the large deviation effects to find a much more accurate estimate of the long-run value. This would also give an interpretation of the long run 'typical' behaviour of the system (see Remark 1.4.4).

4.6 A Measure Change to Incorporate Large Deviation Effects

Next we try to improve our numerical estimate of the growth rate χ of Problem 4.2.2, which is given by

$$\chi := \lim_{T\to\infty} \frac{1}{T} \log\left\{\mathbb{E}^{\mathbb{Q}}\left[\exp\left\{\int_0^T \left(\hat{\gamma} q_t^2 - \gamma\phi(u^*(q_t), q_t)\right)\, dt\right\}\right]\right\}, \tag{4.30}$$

where $\hat{\gamma} = \frac{1}{2}\gamma^2\sigma^2$, $\mathbb{Q}$ is the measure corresponding to the optimal control u^*, and the dynamics of q follow

$$dq_t = \zeta_0\, dW_t + u_t^*\, dt. \tag{4.31}$$

By Jensen's inequality we can bound this by

$$\chi \leq \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^{\mathbb{Q}}\left[\int_0^T \left(\hat{\gamma} q_t^2 - \gamma\phi(u_t, q_t)\right)\, dt\right] = \int g^*(q)\left(\hat{\gamma} q^2 - \gamma\phi(u^*(q), q)\right)\, dq, \tag{4.32}$$

where $g^*(q) = A\exp\left\{\frac{2}{\zeta_0^2}\int_0^q u^*(\tilde{q})\, d\tilde{q}\right\}$ is the stationary distribution of the inventory process under the optimal control. In general, this inequality will be strict due to the fact that there is randomness in the integral term, and hence we may potentially see large deviation effects. We will introduce a change of measure to the original dynamics and correct the process under these dynamics to try and get equality to hold in Jensen's inequality. We introduce a strictly positive function h, and write

$$\kappa(q) := \hat{\gamma} q^2 - \gamma\phi(q, u^*(q)),$$
$$\mathcal{L}^* f(q) := \frac{1}{2}\zeta_0^2 f''(q) + u^*(q) f'(q).$$

Note that the process

$$M_t^h := \exp\left\{-\int_0^t \frac{\mathcal{L}^* h}{h}(q_s)\, ds\right\}\frac{h(q_t)}{h(q_0)},$$

has $M_0^h = 1$ and M_t^h is a strictly positive $\mathbb{Q}$-martingale. In particular, we can define a new probability measure $\mathbb{Q}^h$ by

$$\left.\frac{d\mathbb{Q}^h}{d\mathbb{Q}}\right|_t = M_t^h,$$

under which we can compute the dynamics of the process q using

$$\mathbb{E}^{\mathbb{Q}^h}\left[f(q_t)\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left\{-\int_0^t \frac{\mathcal{L}^* h}{h}(q_s)\, ds\right\} \frac{h(q_t)}{h(q_0)} f(q_t)\right].$$

From this we can deduce that under $\mathbb{Q}^h$, the generator of the process q is given by

$$\mathcal{L}^h f(q) := \frac{1}{2}\zeta_0^2 f''(q) + \left(u^*(q) + \frac{\zeta_0^2 h'(q)}{h(q)}\right) f'(q) = \mathcal{L}^* f(q) + \frac{\zeta_0^2 h'(q)}{h(q)} f'(q).$$

Now consider our expression of interest under $\mathbb{Q}^h$. Then we get

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}\left[\exp\left\{\int_0^T \left(\hat{\gamma} q_t^2 - \gamma\phi(u^*(q_t), q_t)\right)\, dt\right\}\right] \\
&\quad = \mathbb{E}^{\mathbb{Q}^h}\left[\exp\left\{\int_0^T \left(\frac{\mathcal{L}^* h}{h}(q_t) + \hat{\gamma} q_t^2 - \gamma\phi(u^*(q_t), q_t)\right)\, dt\right\} \cdot \frac{h(q_0)}{h(q_T)}\right].
\end{aligned}$$

If we choose h such that

$$\chi^h = \frac{\mathcal{L}^* h}{h}(q_t) + \hat{\gamma} q_t^2 - \gamma\phi(u^*(q_t), q_t),$$

for some constant χ^h, then (assuming $h(q_T)^{-1}$ is well behaved), we expect 'nice' long-run behaviour of the term in the integral, and that we will get equality in Jensen's inequality. Specifically, that we would expect

$$\chi = \chi^h = \int g^h(q)\left(\frac{\mathcal{L}^* h}{h}(q_t) + \hat{\gamma} q^2 - \gamma\phi(u^*(q), q)\right)\, dq,$$

where g^h is the stationary law of q under $\mathbb{Q}^h$. I.e. we expect g^h to be the adjoint solution to

$$\frac{1}{2}\zeta_0^2 f''(q) + \left(u^*(q) + \frac{\zeta_0^2 h'(q)}{h(q)}\right) f'(q),$$

that is, the function g^h such that

$$\int g^h(q)\left(\frac{1}{2}\zeta_0^2 f''(q) + \left(u^*(q) + \frac{\zeta_0^2 h'(q)}{h(q)}\right) f'(q)\right) dq = 0$$

for all (nice) f, or as above,

$$g^h(x) = A^h \exp\left\{\int_0^q \left(\frac{2}{\zeta_0^2} u(\tilde{q}) + 2\frac{h'(\tilde{q})}{h(\tilde{q})}\right) d\tilde{q}\right\} = A^h h(q)^2 g(q).$$

Note that in the trivial case where $h \equiv 1$, g and g^h are the same and χ^h is (possibly non-constant) equal to the estimate we made above. Following the discussion above, we exect the optimal choice of h to solve

$$\chi = \frac{\mathcal{L}^* h}{h}(q_t) + \hat{\gamma} q_t^2 - \gamma\phi(u^*(q_t), q_t).$$

Recalling that the function w, which is essentially the (log) value function is the solution to the non-linear equation

$$\frac{1}{2}\zeta_0^2 w''(q) + Q_0(q, w') + \hat{\gamma}\sigma^2 q^2 = \chi.$$

Then we see that we need to choose h to solve

$$\frac{\mathcal{L}^* h}{h}(q_t) = \frac{1}{2}\zeta_0^2 w''(q) + \frac{1}{2}\zeta_0^2 (w'(q))^2 + w'(q) u^*(q).$$

Recalling that $\mathcal{L}^* f(q) = \frac{1}{2}\zeta_0^2 f''(q) + u^*(q) f'(q)$ we observe that if we take $h(q) = e^{w(q)}$ then this holds. In particular, we observe that we get the modified expression:

$$\begin{aligned}\chi &= \int g^h(q)\left(\frac{\mathcal{L}^* h}{h}(q_t) + \hat{\gamma} q^2 - \gamma\phi(u^*(q), q)\right) dq \\ &= \int g^h(q)\left(\frac{1}{2}\zeta_0^2 w''(q) + \frac{1}{2}\zeta_0^2 (w'(q))^2 + w'(q) u^*(q) + \hat{\gamma} q^2 - \gamma\phi(u^*(q), q)\right) dq.\end{aligned}$$

This form is useful, since we can compute this numerically easily. Indeed, whereas our previous estimate without taking into effect the large deviations had an error of a few percent, using this integral we see almost no difference between the value computed this way and the accurate value from the original PDE. In Figure 4-5 we show plots of the optimal density and the function being integrated against before and after incorporating the large deviation effects. We note that the new density spends a little more time in the tails but the function being integrated against is much nicer. The power of this approach is that we are now able to compute

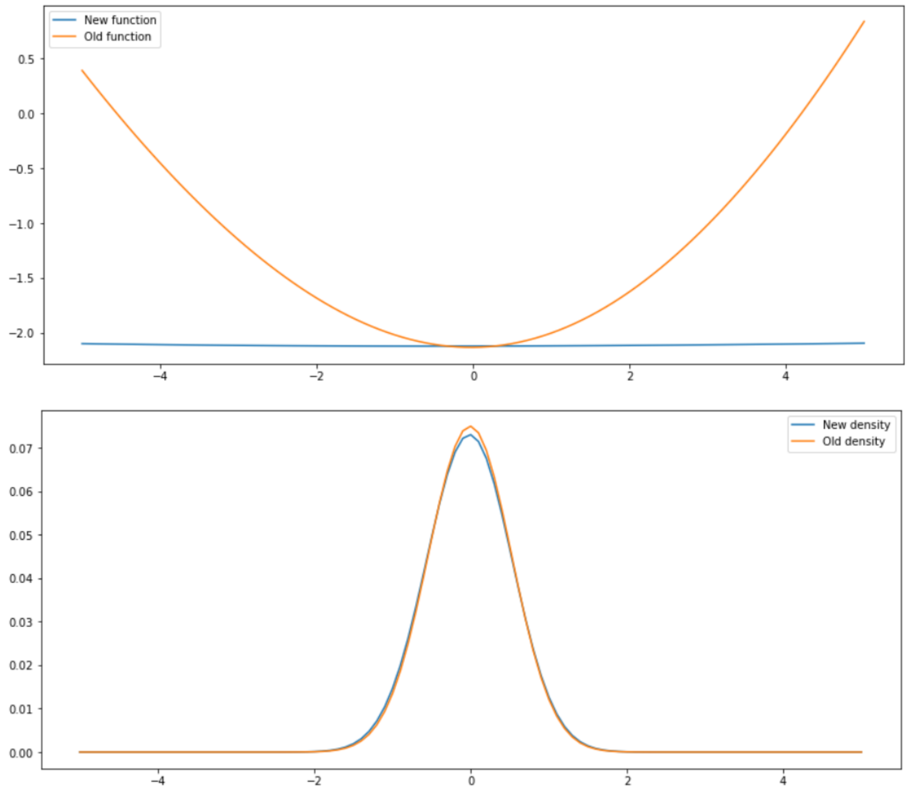


Figure 4-5: Plots of density and function being integrated against before and after accounting for large deviation effects.

an estimate of the long run growth under a range of parameter choices without having to resolve the solution to the PDE each time and we could optimise over parameter choices by a gradient descent algorithm or similar procedure. In particular in Section 5.3 we will suggest a way that we could set up a similar model that incorporates a last look feature. Although we do not directly investigate optimising over the various parameters in the present work, this framework will give us a way we may be able to to so in the last look case where there are a larger number of relevant parameters to optimise over.

4.7 Summary

In this chapter, by putting the inventory process centre-stage and moving to a problem with a single control variable we have given a rigorous solution to a continuous time and space version of the original problem of Chapter 2. This has allowed us to focus clearly on the long-run dynamics and optimal control and we have also been able to solve for this optimal strategy numerically. We have also found a way to compute the required exponential integrals for the value function numerically so that we have a framework in which we can optimise for value over various parameters in this problem as well as in the similar model we will consider in Chapter 5.

Chapter 5

Market Making with Last Look

In this chapter we begin by returning to the discrete world and consider the optimal market making problem in the case where transactions are subject to a conditional execution, or 'last look' criterion. We consider results from two existing related models from the market making literature, and extend them naturally to the last look case. We will also propose a continuous model which could be suitable for a similar analysis as in Chapter 4.

Recall that (as discussed further in the introductory Section 1.2) a 'trade acceptance protocol' or 'last look' mechanism is a term, usually written into a market maker's terms of service, that sets out certain conditions under which a trade may be cancelled after it has been agreed. We will consider last look mechanisms that cancel the trade if the asset price moves quickly after the trade has been agreed. As discussed in Section 1.2, there are a range of reasons market makers may wish to include such mechanisms, but the main one we have in mind is to protect against informed traders who may have a short-term informational advantage. We consider both one and two-sided mechanisms (that may protect just one party or both), including two sided mechanisms that may act either symmetrically or asymmetrically.

Firstly, in Section 5.1, we consider the market making model proposed by Avellaneda and Stoikov [1], and derive adjustments to their reservation prices under a range of possible last look criteria. The main adaptation with last look is that the inventory change and wealth changes are now conditional on movements of the price of the underlying asset, and so where the previous argument had Normally distributed variables, ours follow truncated Normal distributions.

Since our results make extensive use of these truncated Normal distributions and their moment generating functions, some useful results about these are collected for reference in Appendix A. We are able to find closed forms for the reservation prices in each last look case and to see how these compare naturally with the case without last look. We also consider the perspective of the client in balancing reduced spread costs against less certain execution.

In Section 5.2 we turn to the stochastic control problem posed in [1], but take as a starting point the model of Guéant, Lehalle and Fernandez-Tapia [52] that builds on these results. In Section 5.2.2 we consider an asymptotic result based on the spectral theory of an appropriate matrix. We note that the arguments here are somewhat similar to those in Section 2.2.5 - in that section the Krien-Rutman Theorem of functional analysis allowed us to identify the leading eigenvalue and in this case the Perron-Frobenius Theorem plays the same role. As a result we are able to derive natural adjustments to the optimal quotes presented in [52] for the last look case.

In Section 5.2.3 we present the results of simulating numerically a market maker using the optimal quoting strategies to continuously quote and trade with a random demand. We find empirically that in the absence of toxic order flow, that is when orders arrive naturally as a result of uninformed random demand, a symmetric last look feature reduces the overall profitability of the market maker as the last look feature effectively thins out the order flow with no particular protective benefit. However, when the market maker is faced with some clients who have inside information about future price moves of the underlying product, a last look feature becomes vital to protect profits.

In Section 5.3 we propose a continuous model that can capture the last look problem in a similar way to that of Chapter 4. We begin by setting up a model that contains a last look mechanism and also allows for orders of a number of different types. This is important as last look is primarily of use when trading with counterparties with informational advantages. The model we propose is able to capture this, but is too complicated to use directly to optimise over the various parameters. So we propose a continuous model that captures this well and is suited to a similar analysis to that of Chapter 4. We conduct numerical simulations to show that our model captures the problem well and that it fits into this framework.

5.1 A Discrete Model for Market Making with Last Look

We consider a scenario equivalent to the 'frozen inventory' case of Avellaneda and Stoikov [1] which we described in some detain in Section 1.3.1 and consider a market maker taking a single trade at time t that is subject to a last look condition at time $t + \delta t$. Aside from this trade, the market maker makes no further trades until some terminal time $T > t + \delta t$. We assume that the market maker starts at time t with inventory $q_t = q$ and cash $x_t = x$ and that the asset $S_t = s$ behaves as in previous chapters as an arithmetic Brownian motion satisfying $dS_t = \sigma dW_t$.

5.1.1 Reservation Prices for Two-Sided Symmetric Last Look

We begin by considering a two sided last look feature which operates symmetrically towards the market maker and the client. In particular, there is a level ξ such that if the price is further than ξ from the price when the order was submitted at the end of the last look window of length δt the trade is cancelled, and otherwise it is fulfilled.

As in the model of Avellaneda and Stoikov [1] we let r^b and r^a denote the reservation bid and ask prices respectively. That is, these are the prices at which the market maker will transact if they are not cancelled by the last look feature. In this case we may write the market maker's terminal inventory and wealth on the bid side by

$$q_T = q_t + \mathbb{1}_{\{|S_{t+\delta t} - S_t| < \xi\}},$$

$$x_T = x_t - r^b \mathbb{1}_{\{|S_{t+\delta t} - S_t| < \xi\}},$$

and on the ask side by

$$q_T = q_t - \mathbb{1}_{\{|S_{t+\delta t} - S_t| < \xi\}},$$
$$x_T = x_t + r^a \mathbb{1}_{\{|S_{t+\delta t} - S_t| < \xi\}}.$$

Reservation bid price

Consider first the market maker's problem of setting their reservation bid price $r_b(s, q, t)$ in such a way as to be indifferent between accepting the order with last look and holding their current inventory without making any trade.

We assume as in previous chapters that the market maker optimises a CARA utility function

$$v(x, s, q, t) = \mathbb{E}_{t,x,q,s}(-\exp(-\gamma(x_T + q_T S_T))),$$

where $\mathbb{E}_{t,x,s,q}$ is the expected value of the process where the market maker starts with cash x and inventory q of an asset priced at $S_t = s$ at time t, and γ is a risk aversion parameter characterising the market maker as before. For simplicity we will write $\mathbb{E}$ to represent the expectation $\mathbb{E}_{t,x,s,q}$.

Proposition 5.1.1 *Suppose a market maker holding q units of inventory and x in cash accepts a bid order at time t at the price r_b subject to a last look mechanism that will cancel the trade if $|S_{t+\delta t} - S_t| > \xi$. Then we have*

$$\begin{aligned}\mathbb{E}(-e^{-\gamma(x_T + q_T S_T)}) = &- e^{-\gamma(x+qs) + \frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}} g(q) \\ &- e^{-\gamma(x-r^b) - \gamma(q+1)s + \frac{\gamma^2 (q+1)^2 \sigma^2 (T-t)}{2}} g(q+1),\end{aligned}$$

where we define $g(q) := \Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right) - \Phi\left(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)$ and Φ is the normal CDF.

In order to set the reservation bit price r^b at a level so that they are indifferent *between taking the trade or not, they should set the price as*

$$r^b = s - (1+2q)\frac{\sigma^2\gamma(T-t)}{2} + \frac{1}{\gamma}\ln\frac{g(q)}{g(q+1)}.$$

Proof We define an indicator function that is 1 in the case that the transaction is not impacted by the last look feature and 0 if the last look feature cancels the trade. That is

$$\chi := \begin{cases} 1 & , |S_{t+\delta t} - S_t| < \xi \\ 0 & , |S_{t+\delta t} - S_t| \geq \xi. \end{cases}$$

Then the expected utility of the market maker's terminal position having taken the order is given by

$$\begin{aligned} \mathbb{E}(-e^{-\gamma(x_T+q_TS_T)}) = \mathbb{E}(-e^{-\gamma(x+qS_T)}|\chi=0)\mathbb{P}(\chi=0) \\ + \mathbb{E}(-e^{-\gamma(x-r^b+(q+1)S_T)}|\chi=1)\mathbb{P}(\chi=1). \end{aligned}$$

Since each of the two expectations on the right hand side is the moment generating function of a truncated normal distribution, we may rewrite this in terms of the normal CDF.

$$\begin{aligned} \mathbb{E}(-e^{-\gamma(x_T+q_TS_T)}) =& \mathbb{E}(-e^{-\gamma(x+qS_T)}|\chi=0)\mathbb{P}(\chi=0) \\ &+ \mathbb{E}(-e^{-\gamma(x-r^b+(q+1)S_T)}|\chi=1)\mathbb{P}(\chi=1) \\ =& -e^{-\gamma x}\mathbb{E}_t(e^{-\gamma q(S_T-S_{t+\delta t})})\mathbb{E}(e^{-\gamma q(S_{t+\delta t})}|\chi=0)\mathbb{P}(\chi=0) \\ &- e^{-\gamma(x-r^b)}\mathbb{E}_t(e^{-\gamma(q+1)(S_T-S_{t+\delta t})})\mathbb{E}(e^{-\gamma(q+1)(S_{t+\delta t})}|\chi=1)\mathbb{P}(\chi=1). \end{aligned}$$

We now note that $S_T - S_{t+\delta t} \sim N(0, \sigma^2(T-(t+\delta t)))$ and that starting from $S_t = s$, $S_{t+\delta t} \sim N(s, \sigma^2\delta t)$. Applying the moment generating function results from Appendix A we have

$$\begin{aligned} \mathbb{E}(-e^{-\gamma(x_T+q_TS_T)}) =& -e^{-\gamma x}e^{\frac{\gamma^2\sigma^2q^2}{2}(T-t-\delta t)}e^{-\gamma qs+\frac{\gamma^2\sigma^2q^2}{2}\delta t}\Big(\frac{1-g(q)}{1-\mathbb{P}(\chi=1)}\Big)(1-\mathbb{P}(\chi=1)) \\ &- e^{-\gamma(x-r^b)}e^{\frac{\gamma^2\sigma^2}{2}(T-t-\delta t)}e^{-\gamma(q+1)s+\frac{\gamma^2\sigma^2(q+1)^2}{2}\delta t}\Big(\frac{g(q+1)}{\mathbb{P}(\chi=1)}\Big)\mathbb{P}(\chi=1) \\ =& -e^{-\gamma(x+qs)}e^{\frac{\gamma^2\sigma^2q^2}{2}(T-t)}(1-g(q)) \\ &- e^{-\gamma(x-r^b+(q+1)s)}e^{\frac{\gamma^2\sigma^2(q+1)^2}{2}(T-t)}g(q+1). \end{aligned}$$

Thus to find the reservation bid price we equate this to the utility of simply holding the frozen inventory and not entering into any transaction, that is to

$$v(x,s,q,t) = -\exp\left(-\gamma(x+qs) + \frac{\gamma^2 q^2 \sigma^2 (T-t)}{2}\right).$$

So we set

$$\mathbb{E}\left(-e^{-\gamma(x_T + q_T S_T)}\right) = e^{-\gamma(x+qs)+\frac{\gamma^2\sigma^2 q^2 (T-t)}{2}}$$

which rearranges to give

$$(1-g(q)) - e^{-\gamma(s-r^b)} e^{\frac{\gamma^2\sigma^2(2q+1)}{2}(T-t)} g(q+1) = 1.$$

Finally this can be rearranged to give the reservation bid price as stated:

$$r^b = s - (1+2q)\frac{\sigma^2\gamma(T-t)}{2} + \frac{1}{\gamma}\ln\frac{g(q)}{g(q+1)}.$$

$\square$

Remark 5.1.2 *If we take approximations* $g(q+1) \approx g(q) + g'(q)$ *and* $\ln(1+x) \approx x$ *to give* $\ln\left(\frac{g(q+1)}{g(q)}\right) \approx \ln\left(1 + \frac{g'(q)}{g(q)}\right) \approx \frac{g'(q)}{g(q)}$ *then we can write the reservation bid price approximately as*

$$r^b \approx s - (1+2q)\frac{\sigma^2\gamma(T-t)}{2} + \frac{1}{\gamma}\frac{g'(q)}{g(q)}$$

which is equivalent to

$$r^b \approx s - (1+2q)\frac{\sigma^2\gamma(T-t)}{2} - \sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\right) - \phi\left(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right) - \Phi\left(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}.$$

where Φ *and* ϕ *are the standard Normal c.d.f. and p.d.f. respectively.*

Remark 5.1.3 *Although we have chosen to start with a two-sided symmetric last look mechanism, there is nothing particularly special about this and in Sections 5.1.2 and 5.1.3 we will write down very similar forms in the one-sided and asym-*

metric cases whose proofs would follow exactly the same argument as here, with a suitably modified χ.

Reservation ask price

A near identical argument on the ask side yields a reservation ask price of

$$r^a = s + (1-2q)\frac{\sigma^2\gamma(T-t)}{2} - \frac{1}{\gamma}\ln\frac{g(q)}{g(q-1)},$$

and an approximate reservation ask price of

$$r^a \approx s + (1-2q)\frac{\sigma^2\gamma(T-t)}{2} - \frac{1}{\gamma}\frac{g'(q)}{g(q)},$$

or equivalently

$$r^a \approx s + (1-2q)\frac{\sigma^2\gamma(T-t)}{2} - \sigma\sqrt{\delta t}\frac{\phi\Big(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big) - \phi\Big(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big)}{\Phi\Big(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big) - \Phi\Big(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big)}.$$

Reservation mid-price and spread

Averaging the reservation bid and ask prices yields a reservation mid-price

$$r = s - q\sigma^2\gamma(T-t) - \sigma\sqrt{\delta t}\frac{1}{\gamma}\frac{g'(q)}{g(q)},$$

and the reservation spread is

$$r^a - r^b = \sigma^2\gamma(T-t).$$

We see that the spread is exactly as in the model of [1] and the impact of the last look feature in this case is a simple translation of the bid and ask prices.

We also note that the term

$$\sigma\sqrt{T-t}\frac{\phi\Big(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big) - \phi\Big(\frac{-\xi}{\sigma\sqrt{T-t}} + q\gamma\sigma\sqrt{\delta t})\Big)}{\Phi\Big(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big) - \Phi\Big(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big)},$$

is the adjustment needed to be made to the reservation price of [1] in the last look case. Note that its denominator is always positive and its numerator is positive if and only if the inventory q is negative, and so this term has the opposite sign to the inventory. As illustrated in Figure 5-1, we see that the Avellaneda and Stoikov [1] prices translate the $q = 0$ prices according to the inventory level, and with last look this translation is partially reversed by this new term.

We note that this does not account for possible behavioural effects. That is, we assume that the agents will still make the same trades with the last look feature in place as they would have done without it, which in practice may not be the case.

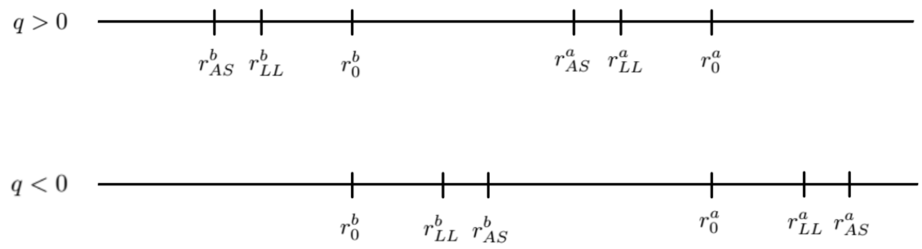


Figure 5-1: Illustration of three sets of reservation prices (in each of the cases $q > 0$ and $q < 0$); (i) r^b_0 and r^a_0 the reservation prices with $q = 0$, (ii) r^b_{AS} and r^a_{AS}, the Avellaneda and Stoikov [1] prices, and (iii) r^b_{LL} and r^a_{LL} the prices with our Last Look adjustment.

5.1.2 Reservation Prices for More General Two-Sided Last Look

In fact, whilst the previous case was presented as a quite natural choice of symmetric last look condition, the conclusions above may easily be stated more generally. The last look feature may be specified using four values $\xi_1, \xi_2, \xi_3, \xi_4$ so that the market maker's terminal inventory and wealth are given on the bid side by

$$q_T = q_t + \mathbb{1}_{\{-\xi_1 < S_{t+\delta t} - S_t < \xi_2\}},$$

$$x_T = x_t - r^b \mathbb{1}_{\{-\xi_1 < s_{t+\delta t} - s < \xi_2\}},$$

and on the ask side by

$$q_T = q_t - \mathbb{1}_{\{-\xi_3 < S_{t+\delta t} - S_t < \xi_4\}},$$

$$x_T = x_t + r^a \mathbb{1}_{\{-\xi_3 < S_{t+\delta t} - S_t < \xi_4\}}.$$

We will then have an approximate reservation bid price of

$$r^b \approx s - (1+2q)\frac{\sigma^2\gamma(T-t)}{2} - \sigma\sqrt{\delta t}\frac{\phi\Big(\frac{\xi_2}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big) - \phi\Big(\frac{-\xi_1}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big)}{\Phi\Big(\frac{\xi_2}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big) - \Phi\Big(\frac{-\xi_1}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big)},$$

and an approximate reservation ask price of

$$r^a \approx s + (1-2q)\frac{\sigma^2\gamma(T-t)}{2} - \sigma\sqrt{\delta t}\frac{\phi\Big(\frac{\xi_4}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big) - \phi\Big(\frac{-\xi_3}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t})\Big)}{\Phi\Big(\frac{\xi_4}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big) - \Phi\Big(\frac{-\xi_3}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\Big)}.$$

The proofs of these results are near identical to that given above for the reservation bid price.

5.1.3 Reservation Prices for Asymmetric Last Look

The previous result can be interpreted naturally in the case that any of the $\xi_1, \xi_2, \xi_3, \xi_4 \to \infty$, and in particular setting $\xi_2 = \xi_3 = \infty$ above gives a case where the last look feature provides protection against adverse price moves for the market maker but not for the client. Let us also take $\xi_1 = \xi_4 = \xi$ so that the market maker's terminal inventory and wealth is given on the bid side by

$$q_T = q_t + \mathbb{1}_{\{S_{t+\delta t} - S_t > -\xi\}},$$

$$x_T = x_t - r^b \mathbb{1}_{\{S_{t+\delta t} - S_t > -\xi\}},$$

and on the ask side as

$$q_T = q_t - \mathbb{1}_{\{S_{t+\delta t} - S_t < \xi\}},$$

$$x_T = x_t + r^a \mathbb{1}_{\{S_{t+\delta t} - S_t < \xi\}}.$$

Then the reservation bid price will be given by

$$
\begin{aligned}
r^b &\approx s + (1+2q)\frac{\sigma^2\gamma(T-t)}{2} + \sigma\sqrt{\delta t}\frac{\phi\left(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}{1-\Phi\left(\frac{-\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)} \\
&= s + (1+2q)\frac{\sigma^2\gamma(T-t)}{2} + \sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} - q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} - q\gamma\sigma\sqrt{\delta t}\right)},
\end{aligned}
$$

and the reservation ask price will be given by

$$
r^a \approx s + (1-2q)\frac{\sigma^2\gamma(T-t)}{2} - \sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}.
$$

Then the mid price becomes slightly skewed, and the spread is reduced from $\sigma^2\gamma(T-t)$ to

$$
\sigma^2\gamma(T-t) - \sigma\sqrt{\delta t}\left(\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} - q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} - q\gamma\sigma\sqrt{\delta t}\right)} + \frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}} + q\gamma\sigma\sqrt{\delta t}\right)}\right).
$$

5.1.4 Liquidity Taker's Perspective

If we only consider the impact of the last look facility on the bid-ask spread, it would seem that the asymmetric facility is beneficial to the liquidity taking client, since the spread is reduced and so they will be paying less on each trade. However, the client must also factor in the 'slippage' cost involved in the case that the last look mechanism cancels the trade when the price moves against them. A measure used in practice to capture this is the 'effective spread', defined in [69] as

Effective spread = Spread paid on Fill + {Reject Ratio * Reject Cost}

and we use the term *slippage costs* to refer to the part of the effective spread that arises from trades being rejected, that is the term {Reject Ratio * Reject Cost}.

In the case of a symmetric facility as defined above the overall slippage costs are zero, since the cost is just as likely to be a benefit when the last look activates in the client's favour rather than the market maker's. In the asymmetric case as set up in Section 5.1.3 we can easily compute that the effective half spread on the ask side[1] will be reduced from $\frac{\sigma^2\gamma(T-t)}{2}$ by

$$\begin{aligned}
&\sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}-\mathbb{P}(S_{t+\delta t}-S_t>\xi)\mathbb{E}[S_{t+\delta t}-S_t|S_{t+\delta t}-S_t>\xi]\\
=&\sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}-\left(1-\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)\right)\sigma\sqrt{\delta t}\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)}{1-\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)}\\
=&\ \sigma\sqrt{\delta t}\left(\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}-\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)\right).
\end{aligned}$$

Similarly the reduction on the bid side would be

$$\sigma\sqrt{\delta t}\left(\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}-q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}-q\gamma\sigma\sqrt{\delta t}\right)}-\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)\right),$$

leading to an overall reduction in the effective spread of

$$\sigma\sqrt{\delta t}\left(\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}-q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}-q\gamma\sigma\sqrt{\delta t}\right)}+\frac{\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}{\Phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}+q\gamma\sigma\sqrt{\delta t}\right)}-2\phi\left(\frac{\xi}{\sigma\sqrt{\delta t}}\right)\right).$$

In Figure 5-2 we plot this overall reduction in the effective spread with sensible parameter choices and find that the asymmetric last look facility appears to always benefit the liquidity taker by this measure.

The apparent benefit to the liquidity taker here masks an important subtlety, which is revealed by plotting the bid and ask components of the spread separately. In figure 5-3 we present equivalent plots for the bid and ask half-spreads.

[1]By this we mean the ask component of the spread plus the cost of rejects on the ask side only.

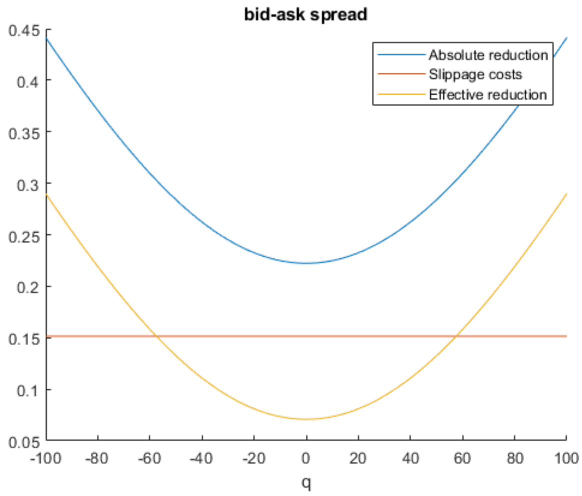
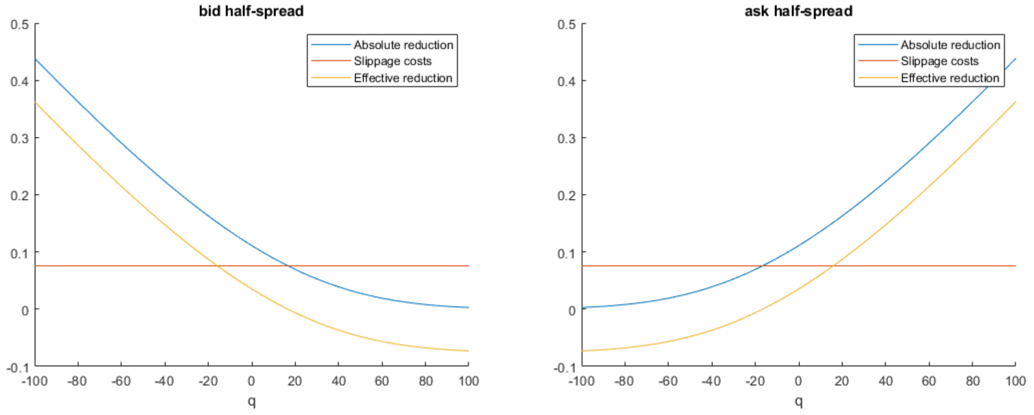


Figure 5-2: Plot showing the absolute/overall reduction in spread compared to the non-last look case, the slippage costs and the overall effective reduction (the absolute reduction less the slippage costs) to the bid-ask spread in an asymmetric facility. We have taken $\sigma = 0.3$, $\gamma = 0.1$, $\xi = 0.4$, $\delta t = 0.5$.

Now we see that there are values of the market maker's inventory q on both the bid and ask sides for which the effective reduction in costs is negative, that is where the last look facility disadvantages the liquidity taker. Since the slippage costs from the liquidity taker's perspective are modelled independently of the market maker's inventory levels, the last look facility is now only offering an effective discount when the market maker does not have too short a position on the ask side or a too long position on the bid side.

Nonetheless, it would appear that the last look facility is in general beneficial to the liquidity taker. On the ask side, for example, when the market maker's inventory is longer, they will be offering more competitive prices, and it is in this case that the effective discount is positive. In the case when the market maker's inventory is very short, their ask quotes will most likely become uncompetitive, and so the liquidity taker will be more likely to trade with an alternative market maker. Thus practically it would appear that the last look facility is providing a benefit to the liquidity taker, in terms both of the absolute and effective spreads, for the scenarios where they are most likely to be trading with the market maker.

These comments only apply to liquidity takers not trading on informational advantage. Of course the last look feature will be harmful to a liquidity taker whose



(a) Values for the bid half-spread (b) Values for the ask half-spread

Figure 5-3: Plots showing the absolute reduction, the slippage costs and the overall effective reduction to the bid and ask half-spreads in an asymmetric facility. We have taken $\sigma = 0.3$, $\gamma = 0.1$, $\xi = 0.4$, $\delta t = 0.5$.

order flow is toxic due to their knowledge of upcoming price moves. This type of liquidity taker will experience a very high reject rate for their trades and be obstructed from monetising their informational advantage.

Further work in this area could be undertaken to consider in more detail subtleties that might arise. In particular it would be natural to consider whether any of these conclusions would change if the market maker's inventory levels or the liquidity taker's demand are correlated with future price moves, or in situations where the market maker holds informational advantage or disadvantage.

5.2 Optimal Bid and Ask Quotes with Last Look

We now return to a consideration of the market making problem as a stochastic control problem, and take a very similar set up to that of Guéant, Lehalle and Fernandez-Tapia [52] (see Section 1.3.2), and adapt their results to the case where transactions are subject to a last look feature. So we consider a market maker maximising a CARA utility function

$$u^{\delta}(t,x,S,q) = \mathbb{E}_{t,x,S,q}[-\exp\{-\gamma(X_T^{\delta} + Q_T^{\delta} S_T)\}],$$

where as before γ is a risk aversion parameter characterising the market maker and $\delta = \{(\delta^b)_{t\leq T}, (\delta^a)_{t\leq T}\}$ is the control process by which the market maker posts limit bid and ask quotes. In particular they set bid and ask prices S_t^b and S_t^a at distances δ_t^b and δ_t^a around a reference price S_t that follows an arithmetic Brownian motion. That is we have

$$dS_t = \sigma dW_t,$$

$$\delta_t^b = S_t - S_t^b,$$

$$\delta_t^a = S_t^a - S_t,$$

and the market maker attempts to optimise $u^{\delta}(t,x,S,q)$ over choices of this control δ.

To simplify notation we write $\Delta S_t = S_{t+\delta_t} - S_t$ and denote by χ an indicator function taking value 1 when the order is filled without the last look feature being used, as in Proposition 5.1.1. We initially take a last look feature which cancels orders symmetrically, so that

$$\chi := \begin{cases} 1 & , |\Delta S_t| < \xi \\ 0 & , |\Delta S_t| \geq \xi. \end{cases}$$

As in [52] and in previous chapters we assume that the market maker acts according to inventory restrictions so that their inventory will always lie in the range $q \in \{-Q, ..., Q\}$. As in [52] we suppose the market maker's inventory is given by $q_t = N_t^b - N_t^a$, where N^b and N^a are point processes representing the number of units bought and sold respectively. The intensities Λ^b and Λ^a associated to the processes N^b and N^a are supposed to be functions of the distance the market maker places quotes from the reference price, so that $\Lambda^b = \Lambda^b(\delta^b)$ and $\Lambda^a = \Lambda^a(\delta^a)$.

Then it is straightforward to show that the dynamic programming equation (DPE) associated with this control problem is

$$\begin{aligned}
0 &= (\partial_t + \frac{1}{2}\sigma^2 \partial_{SS})u \\
&+ \sup_{\delta^a} \{\Lambda^a(\delta^a)\mathbb{E}[u(t, x + \chi(S_t + \delta^a), q - \chi, S_{t+\delta t}) - u(t, x, q, S_t)]\} \mathbb{1}_{\{q>-Q\}} \\
&+ \sup_{\delta^b} \{\Lambda^b(\delta^b)\mathbb{E}[u(t, x - \chi(S_t - \delta^b), q + \chi, S_{t+\delta t}) - u(t, x, q, S_t)]\} \mathbb{1}_{\{q<Q\}},
\end{aligned}$$

with associated terminal condition

$$u(T, x, s, q) = -e^{-\gamma(x+qs)}.$$

As in [52] we choose the demand functions $\Lambda^b(\delta^b) = Ae^{-k\delta^b}$ and $\Lambda^a(\delta^a) = Ae^{-k\delta^a}$ and then we have the following proposition.

Proposition 5.2.1 *Writing $u(t, x, q, s) = -e^{-\gamma(x+qs)} v_q(t)^{-\frac{\gamma}{k}}$, the above DPE can be reformulated in terms of $v_q(t)$ as*

$$\begin{aligned}
0 = -\frac{\gamma}{k}\frac{\dot{v}_q(t)}{v_q(t)}u + \frac{1}{2}\gamma^2\sigma^2 q^2 u + u \sup_{\delta^a} \left\{ Ae^{-k\delta^a} \left[B(q-1) \left(\frac{v_{q-1}(t)}{v_q(t)} \right)^{-\frac{\gamma}{k}} e^{-\gamma\delta^a} - 1 \right] \right\} \mathbb{1}_{\{q>-Q\}} \\
+ u \sup_{\delta^b} \left\{ Ae^{-k\delta^b} \left[B(q+1) \left(\frac{v_{q+1}(t)}{v_q(t)} \right)^{-\frac{\gamma}{k}} e^{-\gamma\delta^b} - 1 \right] \right\} \mathbb{1}_{\{q<Q\}},
\end{aligned}$$

with terminal condition $v_q(T) = 1$, where

$$B(q) := e^{\frac{1}{2}\sigma^2\delta t\gamma^2 q^2} \left[\Phi\left(\frac{\xi + \gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}} \right) - \Phi\left(\frac{-\xi + \gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}} \right) \right].$$

Proof Applying the definition of χ, the DPE can be rewritten as

$$
\begin{aligned}
0 = &\left(\partial_t + \tfrac{1}{2}\sigma^2\partial_{SS}\right)u \\
&+ \sup_{\delta^a}\left\{\Lambda^a(\delta^a)\mathbb{E}[\chi u(t, x+(S_t+\delta^a), q-1, S_{t+\delta t}) + (1-\chi)u(t,x,q,S_t) - u(t,x,q,S_t)]\right\}\mathbb{1}_{\{q>-Q\}} \\
&+ \sup_{\delta^b}\left\{\Lambda^b(\delta^b)\mathbb{E}[\chi u(t, x-(S_t-\delta^b), q+1, S_{t+\delta t}) + (1-\chi)u(t,x,q,S_t) - u(t,x,q,S_t)]\right\}\mathbb{1}_{\{q<Q\}}
\end{aligned}
$$

and then as

$$
\begin{aligned}
0 = &\left(\partial_t + \tfrac{1}{2}\sigma^2\partial_{SS}\right)u \\
&+ \sup_{\delta^a}\left\{\Lambda^a(\delta^a)\mathbb{E}[\chi u(t, x+(S_t+\delta^a), q-1, S_{t+\delta t}) - \chi u(t,x,q,S_t)]\right\}\mathbb{1}_{\{q>-Q\}} \\
&+ \sup_{\delta^b}\left\{\Lambda^b(\delta^b)\mathbb{E}[\chi u(t, x-(S_t-\delta^b), q+1, S_{t+\delta t}) - \chi u(t,x,q,S_t)]\right\}\mathbb{1}_{\{q<Q\}}.
\end{aligned}
$$

Then since $\chi = 1$, when $|\Delta S_t| < \xi$ and 0 otherwise, we can write

$$
\begin{aligned}
0 = &\left(\partial_t + \tfrac{1}{2}\sigma^2\partial_{SS}\right)u \\
&+ \sup_{\delta^a}\left\{\Lambda^a(\delta^a)\mathbb{P}(\chi)\mathbb{E}[u(t, x+(S_t+\delta^+), q-1, S_{t+\delta t}) - u(t,x,q,S)\Big||\Delta S_t| < \xi]\right\}\mathbb{1}_{\{q>-Q\}} \\
&+ \sup_{\delta^b}\left\{\Lambda^b(\delta^b)\mathbb{P}(\chi)\mathbb{E}[u(t, x-(S_t-\delta^b), q+1, S_{t+\delta t}) - u(t,x,q,S)\Big||\Delta S_t| < \xi]\right\}\mathbb{1}_{\{q<Q\}}.
\end{aligned}
$$

Now applying $u(t,x,q,s) = -e^{-\gamma(x+qs)}v_q(t)^{-\frac{\gamma}{k}}$ we note that

$$
\begin{aligned}
u(t, x+(S_t+\delta^a), q-1, S_{t+\delta t}) &- u(t,x,q,S) \\
&= u(t,x,s,q)\left[\left(\frac{v_{q-1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}} e^{-\gamma\delta^a}e^{-\gamma(q-1)\Delta S_t} - 1\right],
\end{aligned}
$$

and

$$
\begin{aligned}
u(t, x-(S_t-\delta^b), q+1, S_{t+\delta t}) &- u(t,x,q,S) \\
&= u(t,x,s,q)\left[\left(\frac{v_{q+1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}} e^{-\gamma\delta^b}e^{-\gamma(q+1)\Delta S_t} - 1\right].
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\mathbb{E}\left[u(t,x+(S_t+\delta^a),q-1,S_{t+\delta t})-u(t,x,q,S)\Big||\Delta S_t|<\xi\right]\\
=&u(t,x,q,S)\left[\left(\frac{v_{q-1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}}e^{-\gamma\delta^a}\mathbb{E}_{t,x,S,q}\left[e^{-\gamma(q-1)\Delta S_t}\Big||\Delta S_t|<\xi\right]-1\right],
\end{aligned}
$$

and so

$$
\begin{aligned}
&\mathbb{E}(u(t,x-(S_t-\delta^b),q+1,S_{t+\delta t})-u(t,x,q,S)\Big||\Delta S_t|<\xi)\\
=&u(t,x,q,S)\left[\left(\frac{v_{q+1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}}e^{-\gamma\delta^b}\mathbb{E}_{t,x,S,q}\left[e^{-\gamma(q+1)\Delta S_t}\Big||\Delta S_t|<\xi\right]-1\right].
\end{aligned}
$$

Applying the moment generating function of a truncated normal distribution (see Appendix A) we can rewrite the conditional expectation as

$$
\mathbb{E}\left(e^{-\gamma q\Delta S_t}\Big||\Delta S_t|<\xi\right)=e^{\frac{1}{2}\sigma^2\delta t\gamma^2q^2}\left[\frac{\Phi\left(\frac{\xi+\gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}}\right)-\Phi\left(\frac{-\xi+\gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}}\right)}{\mathbb{P}(\chi=1)}\right]=:\frac{B(q)}{\mathbb{P}(\chi=1)}.
$$

Substituting this as well as the choice of $\Lambda^b(\delta^b)=Ae^{-k\delta^b}$ and $\Lambda^a(\delta^a)=Ae^{-k\delta^a}$ leads to the equation for $v_q(t)$

$$
\begin{aligned}
0=&-\frac{\gamma}{k}\frac{\dot v_q(t)}{v_q(t)}u+\tfrac{1}{2}\gamma^2\sigma^2q^2u\\
&+u\sup_{\delta^a}\left\{\mathbb{P}(\chi=1)Ae^{-k\delta^a}\left[\frac{B(q-1)}{\mathbb{P}(\chi=1)}\left(\frac{v_{q-1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}}e^{-\gamma\delta^a}-1\right]\right\}\mathbb{1}_{\{q>-Q\}}\\
&+u\sup_{\delta^b}\left\{\mathbb{P}(\chi=1)Ae^{-k\delta^b}\left[\frac{B(q+1)}{\mathbb{P}(\chi=1)}\left(\frac{v_{q+1}(t)}{v_q(t)}\right)^{-\frac{\gamma}{k}}e^{-\gamma\delta^b}-1\right]\right\}\mathbb{1}_{\{q<Q\}},
\end{aligned}
$$

which is clearly equivalent to that stated.

□

5.2.1 Optimal Quotes and Solution of Control Problem

We next give a solution of the control problem and provide the optimal quotes in the last look case, extending the work of [52].

Using a first order derivative condition we can easily find that the values of δ^a and δ^b optimising the suprema from Proposition 5.2.1 are

$$\delta^{*,a} = \frac{1}{k}\ln\Big(\frac{v_q(t)}{v_{q-1}(t)}\Big) + \frac{1}{\gamma}\ln\Big(1+\frac{\gamma}{k}\Big) + \frac{1}{\gamma}\ln\frac{B(q-1)}{\mathbb{P}(\chi=1)},$$

and

$$\delta^{*,b} = \frac{1}{k}\ln\Big(\frac{v_q(t)}{v_{q+1}(t)}\Big) + \frac{1}{\gamma}\ln\Big(1+\frac{\gamma}{k}\Big) + \frac{1}{\gamma}\ln\frac{B(q+1)}{\mathbb{P}(\chi=1)}.$$

After some straightforward substitution and algebra the equation for $v_t(q)$ becomes

$$\begin{aligned}0 = -\frac{\gamma}{k}\frac{u}{v_q(t)}\Bigg[&\dot v_q(t) - \frac{1}{2}k\gamma\sigma^2 q^2 v_q(t) + \Big\{A\Big(\frac{1+\frac{\gamma}{k}}{\mathbb{P}(\chi=1)}\Big)^{-(1+\frac{k}{\gamma})} B(q-1)^{\frac{-k}{\gamma}} v_{q-1}(t)\Big\}\mathbb{1}_{\{q>-Q\}} \\ &+ \Big\{A\Big(\frac{1+\frac{\gamma}{k}}{\mathbb{P}(\chi=1)}\Big)^{-(1+\frac{k}{\gamma})} B(q+1)^{\frac{-k}{\gamma}} v_{q+1}(t)\Big\}\mathbb{1}_{\{q<Q\}}\Bigg].\end{aligned}$$

This can all be written as a matrix ODE

$$\partial_t \mathbf{v}(t) + \mathbf{M}\mathbf{v}(t) = \mathbf{0},$$

where we define the vector $\mathbf{v}(t)$ and matrix $\mathbf{M}$ as

$$\mathbf{v}(t) = (v_{-Q}(t), v_{-Q+1}(t), ..., v_0(t), , , , v_{Q-1}(t), v_Q(t))'$$

and

$$\mathbf{M} = \begin{pmatrix} \alpha Q^2 & -\eta B(-Q+1)^{\frac{-k}{\gamma}} & 0 & \cdots & \cdots & 0 \\ -\eta B(-Q)^{\frac{-k}{\gamma}} & \alpha(1-Q)^2 & -\eta B(-Q+2)^{\frac{-k}{\gamma}} & \ddots & \ddots & \vdots \\ 0 & -\eta B(-Q+1)^{\frac{-k}{\gamma}} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\eta B(Q-1)^{\frac{-k}{\gamma}} & 0 \\ \vdots & \ddots & \ddots & -\eta B(Q-2)^{\frac{-k}{\gamma}} & \alpha(Q-1)^2 & -\eta B(Q)^{\frac{-k}{\gamma}} \\ 0 & \cdots & \cdots & 0 & -\eta B(Q-1)^{\frac{-k}{\gamma}} & \alpha Q^2 \end{pmatrix},$$

where $\alpha = \frac{k}{2}\gamma\sigma^2$ and $\eta = A\Big(\frac{1+\frac{\gamma}{k}}{\mathbb{P}(\chi)}\Big)^{-(1+\frac{k}{\gamma})}$. Hence the solution of the control problem is given by

$$u(t,x,q,s) = -\exp(-\gamma(x+qs))v_q(t)^{-\frac{\gamma}{k}},$$

where

$$\mathbf{v}(t) = \exp(-\mathbf{M}(T-t)) \times (1,...,1)'.$$

5.2.2 Asymptotic behaviour of the optimal quotes

We next prove the following proposition giving the optimal quotes in the limit as $T \to \infty$. We follow a similar reasoning to [52], adding in some extra details and adapting the results for the last look case.

Proposition 5.2.2 *In the limit as $T \to \infty$ the optimal quotes become*

$$\delta^{*,a} = \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{1}{\gamma}\ln\frac{B(q-1)}{\mathbb{P}(\chi)} + \frac{1}{k}\ln\left(\frac{f_q^0}{f_{q-1}^0}\left(\frac{B(q-1)}{B(q)}\right)^{\frac{-k}{2\gamma}}\right)$$

$$\delta^{*,b} = \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{k}\right) + \frac{1}{\gamma}\ln\frac{B(q+1)}{\mathbb{P}(\chi)} + \frac{1}{k}\ln\left(\frac{f_q^0}{f_{q+1}^0}\left(\frac{B(q+1)}{B(q)}\right)^{\frac{-k}{2\gamma}}\right)$$

where f^0 satisfies

$$f_0 \in \underset{f\in\mathbb{R}^{2Q+1},||f||=1}{\arg\min} \sum_{q=-Q}^{Q} \alpha q^2 f_q^2 + \sum_{q=-Q}^{Q} 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}} f_q^2 - 2\eta \sum_{q=-Q}^{Q-1} (B(q)B(q+1))^{\frac{-k}{2\gamma}} f_q f_{q+1}.$$

Equivalently, we may write

$$f_0 \in \underset{f\in\mathbb{R}^{2Q+1},||f||=1}{\arg\min} f'(\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I})f,$$

where $\mathbf{J}$ is a matrix, similar to $\mathbf{M}$, defined below. This condition is equivalent to choosing f_0 to be an eigenvector corresponding to the smallest eigenvalue of $\mathbf{J}$.

Proof We work with a symmetric matrix $\mathbf{J}$, similar to $\mathbf{M}$ and defined by the similarity transform $\mathbf{J} := \mathbf{D}^{-1}\mathbf{M}\mathbf{D}$ where the transformation matrix $\mathbf{D}$ is defined as

$$\mathbf{D} := \mathrm{diag}(\delta_{-Q}, \ldots, \delta_Q) \quad \text{for} \quad \delta_i := \left(\frac{B(-Q)}{B(i)}\right)^{\frac{-k}{2\gamma}}.$$

Writing $B_q := B(q)$ this yields the symmetric tridiagonal matrix $\mathbf{J} =$

$$\begin{pmatrix} \alpha Q^2 & -\eta\sqrt{B(-Q)B(-Q+1)}^{\frac{-k}{\gamma}} & & & \\ -\eta\sqrt{B(-Q)B(1-Q)}^{\frac{-k}{\gamma}} & \alpha(1-Q)^2 & -\eta\sqrt{B(1-Q)B(2-Q)}^{\frac{-k}{\gamma}} & & \\ & -\eta\sqrt{B(1-Q)B(2-Q)}^{\frac{-k}{\gamma}} & \ddots & \ddots & \\ & & \ddots & \ddots & -\eta\sqrt{B(Q-1)B(Q)}^{\frac{-k}{\gamma}} \\ & & & -\eta\sqrt{B(Q-1)B(Q)}^{\frac{-k}{\gamma}} & \alpha Q^2 \end{pmatrix}.$$

We note that $\mathbf{M}$ and $\mathbf{J}$ have the same eigenvalues and $\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I}$ is positive definite. To see this, firstly recall the definition of

$$B(q) = e^{\frac{1}{2}\sigma^2\delta t\gamma^2 q^2}\left[\Phi\left(\frac{\xi + \gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}}\right) - \Phi\left(\frac{-\xi + \gamma q\sigma^2\delta t}{\sigma\sqrt{\delta t}}\right)\right].$$

Then

$$\min_{q\in\{-Q,\ldots Q\}} B(q) = B(0) = \mathbb{P}(\chi),$$

and hence

$$\max_{q\in\{-Q,\ldots Q\}} B(q)^{\frac{-k}{\gamma}} = B(0)^{\frac{-k}{\gamma}} = \mathbb{P}(\chi)^{\frac{-k}{\gamma}}.$$

Then, labelling a $2Q+1$ vector $\mathbf{x} = (x_{-Q}, \ldots, x_Q)'$ we have

$$\begin{aligned}
&\mathbf{x}'(\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I})\mathbf{x} \\
&\quad = \sum_{q=-Q}^{Q}\alpha q^2 x_q^2 + \sum_{q=-Q}^{Q} 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}} x_q^2 - 2\eta\sum_{q=-Q}^{Q-1}(B(q)B(q+1))^{\frac{-k}{2\gamma}} x_q x_{q+1} \\
&\quad > \sum_{q=-Q}^{Q}\alpha q^2 x_q^2 + \sum_{q=-Q}^{Q} 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}} x_q^2 - 2\eta\sum_{q=-Q}^{Q-1}\mathbb{P}(\chi)^{\frac{-k}{\gamma}} x_q x_{q+1} \\
&\quad = \sum_{q=-Q}^{Q}\alpha q^2 x_q^2 + \eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\sum_{q=-Q}^{Q-1}(x_{q+1} - x_q)^2 + \eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}} x_Q^2 + \eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}} x_{-Q}^2 \\
&\quad \geq 0.
\end{aligned}$$

In particular, $\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I}$ is an invertible 'M-matrix' and by statement F15 in

[84] it is inverse positive. That is, $(\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I})^{-1}$ has all positive entries. By the Perron-Frobenius theorem $(\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I})^{-1}$ has a largest eigenvalue whose corresponding eigenvector has strictly positive entries.

It is also a well known fact that any real symmetric matrix has real eigenvalues, and if moreover the matrix is tridiagonal, then all eigenvalues are distinct if all off-diagonal entries are non-zero (see e.g. [80]). Hence $\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I}$ has a smallest eigenvalue, and this eigenvalue is simple. Since a matrix and its inverse share eigenvectors and all of the eigenvalues of $\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I}$ are reciprocals of those of $(\mathbf{J} + 2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I})^{-1}$ and are all positive, this simple eigenvalue must also have a corresponding eigenvector with strictly positive entries.

Since $\mathbf{J}$ has the same eigenvectors as $\mathbf{J}+2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}\mathbf{I}$, but with each corresponding eigenvalue reduced by $2\eta\mathbb{P}(\chi)^{\frac{-k}{\gamma}}$, then $\mathbf{J}$ also has a smallest simple eigenvalue with eigenvector f_0 whose entries are strictly positive.

Then we can write

$$v(0) = \exp(-\mathbf{M}T) \times (1, \dots, 1)' = \mathbf{D}\exp(-\mathbf{J}T)\mathbf{D}^{-1} \times (1, \dots, 1)',$$

and in particular we can write (after diagonalising $\mathbf{J}$ and performing some matrix computations)

$$v_q(0) = \delta_q \sum_{i=0}^{2Q} \exp(-\lambda_i T) \left\langle g^i, \Big(\frac{1}{\delta_{-Q}}, \dots, \frac{1}{\delta_Q}\Big) \right\rangle g_q^i,$$

where $\lambda^0 < \lambda^1 \dots \leq \lambda^{2Q}$ are the eigenvalues of $\mathbf{J}$ (equivalently $\mathbf{M}$) in increasing order (possibly repeated) and g^i is an associated orthonormal basis of eigenvectors of $\mathbf{J}$.

We can take $g^0 = f^0$ and then we get that[2], as $T \to \infty$

$$v_q(0) \sim \delta_q \exp(-\lambda_0 T) \left\langle f^0, \left(\frac{1}{\delta_{-Q}}, \dots, \frac{1}{\delta_Q}\right) \right\rangle f_q^0.$$

[2]The positivity of all entries of the dominant eigenpair guarantees that the inner product for this term is not zero.

Then we have that the optimal quotes in the limit as $T \to \infty$ are

$$\delta^{*,a} = \frac{1}{\gamma} \ln \left(1 + \frac{\gamma}{k}\right) + \frac{1}{\gamma} \ln \frac{B(q-1)}{\mathbb{P}(\chi)} + \frac{1}{k} \ln \left(\frac{f_q^0}{f_{q-1}^0} \frac{\delta_q}{\delta_{q-1}} \right),$$

and

$$\delta^{*,b} = \frac{1}{\gamma} \ln \left(1 + \frac{\gamma}{k}\right) + \frac{1}{\gamma} \ln \frac{B(q+1)}{\mathbb{P}(\chi)} + \frac{1}{k} \ln \left(\frac{f_q^0}{f_{q+1}^0} \frac{\delta_q}{\delta_{q+1}} \right),$$

which upon recalling the definition of δ_i leads directly to the optimal quotes as stated.

□

In the next section we present the result of some numerical simulations of this model which help us understand the impact and significance of last look.

5.2.3 Simulations

Impact of inventory limits

In Figure 5-4 we plot the optimal quotes from Section 5.2.1, taking as parameter choices $Q = 5$, $k = 0.3$, $\sigma = 0.3$, $T = 50$, $\gamma = 0.1$, $A = 0.9$, $\xi = 0.4$, $dt = 0.5$.

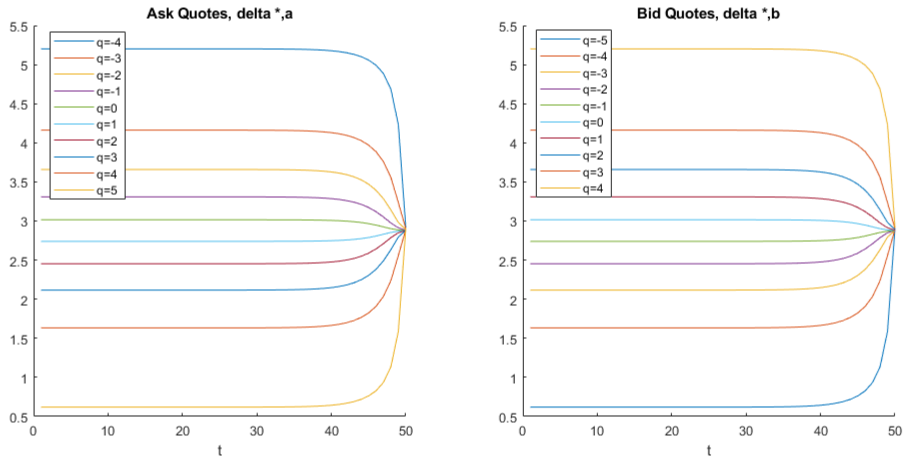


Figure 5-4: The uppermost line for ask quotes corresponds to $q = -4$ and for bid quotes to $q = +4$.

In Figure 5-5 we plot again with the same parameter choices except to extend the inventory cap to $Q = 10$.

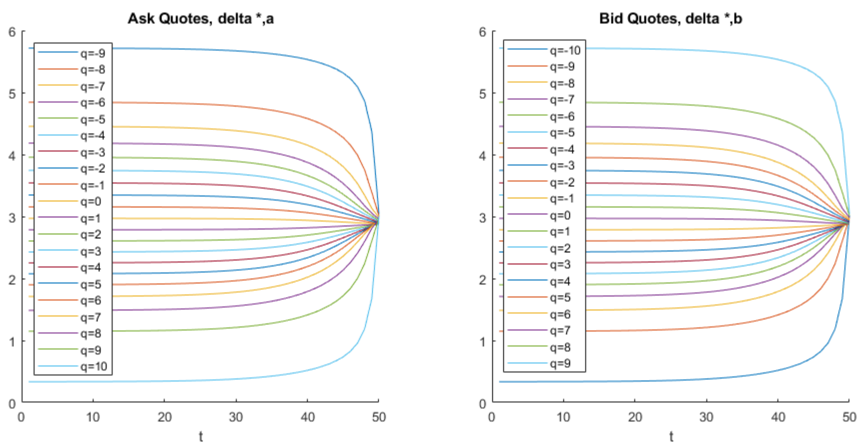


Figure 5-5: The uppermost line for ask quotes corresponds to $q = -9$ and for bid quotes to $q = +9$.

We observe in both figures that the optimal quotes converge as expected for times far from the terminal time. We also observe an interesting feature whereby quotes closest to the inventory cap are further spaced than those closer to $q = 0$, and it appears that all of the quotes depend on the inventory limit Q. In Figure 5-6 we also plot the bid quotes corresponding to $q = 5$ for various values of Q and we observe that the quotes do settle down as Q increases, so this effect is of little impact so long as the inventory cap is significantly larger than typical inventory levels.

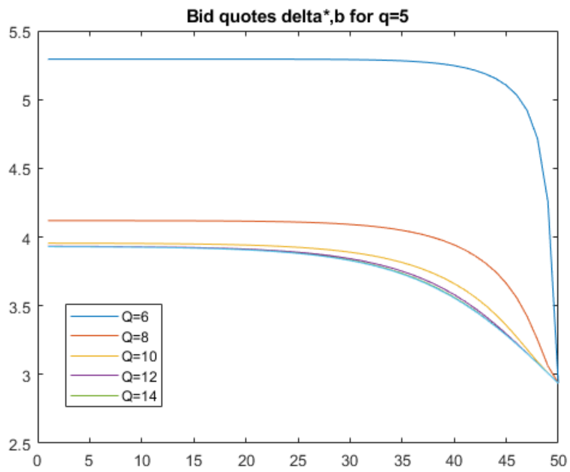


Figure 5-6: Plot of bid quotes corresponding to $q = 5$ for $Q \in \{6, 8, 10, 12, 14\}$.

Optimal Quotes with and without Last Look

We next keep all of the parameters fixed from the previous section, except we now take $T = 100$ and fix an inventory cap $Q = 10$. In Figure 5-7 we compute the bid quotes with last look (setting $\xi = 0.4$, $\delta t = 0.5$ for an acceptance probability $\mathbb{P}(\chi)$ of around 0.94) and also with the last look feature effectively removed (by setting $\xi = 4$ and $\delta t = 0.5$ for $\mathbb{P}(\chi) \approx 1$). The plots in Figure 5-7 results from plotting the difference between the optimal bid quotes with and without last look across different values of q and t.

We observe a similar effect to that seen in Section 5.1.1 for the reservation prices. In particular we see in Figure 5-7 that there is an upward adjustment with last look for $q < 0$ and a downward one for $q > 0$, effectively undoing a little of the inventory adjustment in the quotes without Last Look. We also observe

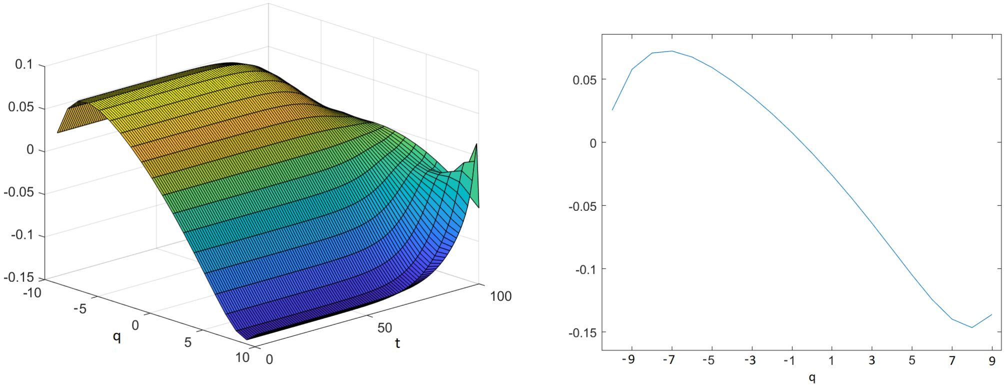


Figure 5-7: Difference between optimal bid quotes with and without Last Look. The left panel displays the difference across times from $t = 0$ to the terminal $T = 100$. The right panel shows a 2-dimensional snapshot of the difference at $t = 0$ where the quotes have settled down as we are far from the terminal time. There is no bid price at $q = 10$ due to the inventory cap.

some interesting behaviour at the extremal quote levels that we have yet to fully explain.

Empirical Comparison of Utility with and without Last Look

In order to consider whether the last look feature is valuable to the market maker we have also created a simulation of the market making problem with a symmetric last look condition. In figure 5-8 we show the output of one run of this simulator, which tracks the running quantities and estimated current profit $(x_t + q_t S_t)$ to the market maker. We have computed the profit made by the market maker over $T = 100$ with the last look feature turned on (reject probability as above around 5 or 6 percent). The plots show the market maker's inventory, running profit, quotes and the times the last look feature was applied.

In Figure 5-9 we show the results of running the simulation 1000 times, with and without a Last Look feature. We observe that market making is more profitable for the market maker under these conditions when there is no last look feature. This should perhaps not be too surprising, since nothing in the model as yet incorporates any sort of adverse selection or insider knowledge effect, so the market maker need not be worried about traders with any form of informational advantage. So the last look feature just has the effect of thinning out demand

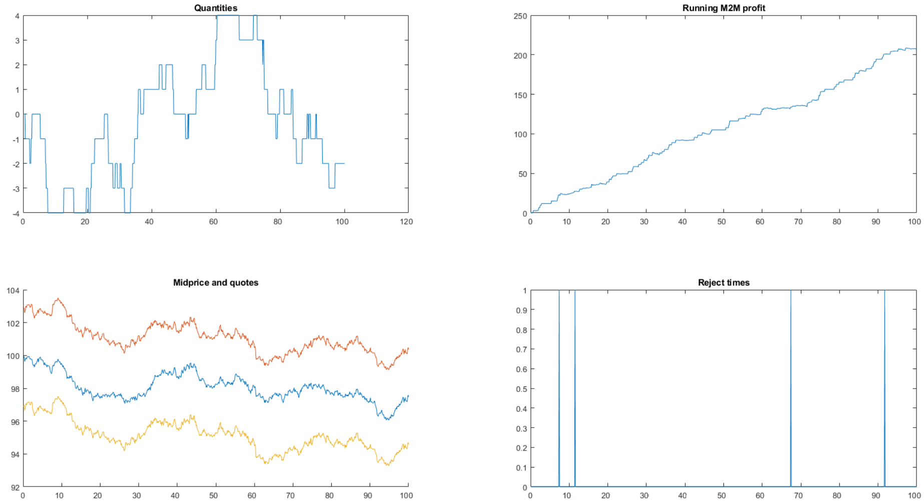


Figure 5-8: Output from a run of the market making simulator. The panels show (anticlockwise from top-left) the running inventory levels of the market maker; the asset midprice and resulting bid and ask quotes; the times at which a last look reject occurs; and the running marked to market profit to the market maker. The profit is 'marked to market' in the sense that as well as observing the running cash levels, the running inventory is valued at the current asset mid-price.

at random when the price moves extremely in either direction. In these circumstances such price moves are equally likely to favour or disadvantage the market maker, and so overall the market maker simply loses out a little as a result of fulfilling fewer orders overall.

In Figure 5-10 we show the results of a similar situation but under significantly adverse market conditions for the market maker. We introduce occasional 10σ jumps to the asset price that another market participant is able to detect and act upon by placing the appropriate order before the market maker is able to update their prices. In this scenario we see that the Last Look feature is strongly protective for the market maker and their profit distribution is significantly enhanced when they have this protection.

This supports the view that last look is predominantly a cost for the market maker in ordinary conditions, but one that is a necessary protection against informed traders who might otherwise drive the market maker out of business.

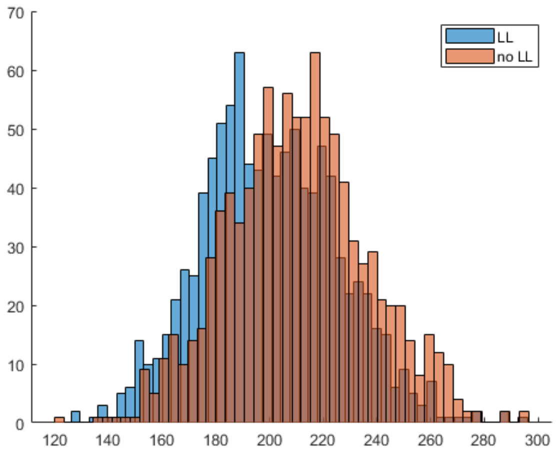


Figure 5-9: Market maker profit distribution over 1000 runs with and without Last Look.

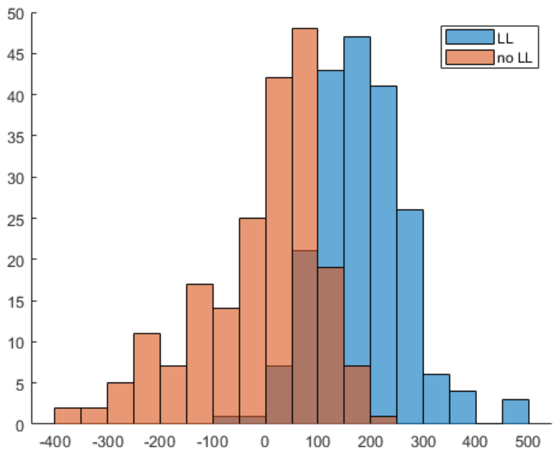


Figure 5-10: Market maker profit distribution over 200 runs with and without last look in the presence of jumps known in advance to a liquidity taker.

5.3 A Continuous Model for Market Making with Last Look

So far in this chapter we have worked to extend the discrete model of [52] to the case with last look. The work we have done, in particular the numerical solution, leads us to the conclusion that last look is most interesting in a world where some traders have informational advantage and this will form a part of the model in this section. Here we set out a continuous framework will usefully capture a last look mechanism in a model like that of Chapter 4 and which is suitable for numerical analysis in a similar manner as in Section 4.5.

5.3.1 Setting Up the model

We first describe slightly informally our model. A more careful description of a slight generalisation will be given in Section 5.3.4.

As before, at any given time the market maker chooses bid and ask spreads via δ^b and δ^a. Sell and buy orders arrive as Poisson Processes with rate $\Lambda^b(\delta^b), \Lambda^a(\delta^a)$ respectively. In particular, if the asset price at time t is S_t, the market maker will fill a sell order (buy from the client) with price $S_t - \delta^b$, and fill a buy order (sell to the client) with price $S_t + \delta^a$. Typical orders are of size Δ.

Unlike in the previous chapters, we consider two types of price change that we may think of as orders from two distinct groups of investors. We will think of the total rate of arrival of orders in the market μ_0 being decomposed into $\mu_0 = \mu_1 + \mu_2$, corresponding to two types of order:

1. Smaller market moves at a rate μ_1. None of these orders are traded with the market maker, which correspond to market moves with mean zero and variance σ_1^2. (Type 1)

2. Larger market moves. Arrivals at rate μ_2 of orders that are more likely traded with the market maker, which correspond to market moves with mean zero and variance σ_2^2. (Type 2)

Arrivals of Type 1 see a change in the price of the underlying asset, S_t but no change to the market maker's inventory q_t as these are not trades involving the market maker. Arrivals of Type 2 see a change in the price of the underlying asset, and if they are trades with the market maker then they will also see a change in the quantity of inventory, which moves up or down by Δ, depending on whether the order is a buy or sell order.

Suppose the asset price dynamics are determined so that price changes happen at rate μ_0, and have a fixed variance σ_0^2. The probability that a given Type 2 order is a trade depends on the trading strategy, given so that the total rate matches that the previous chapters. We expect $\mu_2 \geq \Lambda^b(\delta^b)+\Lambda^a(\delta^a)$ and $\mu_2 \ll \mu_0$, and so $\mu_1 = \mu_0 - \mu_2 \approx \mu_0$. That is to say, the orders traded with the market maker are always a relatively small proportion of the overall rate of orders in the market.

The total variance of a typical order can then be computed as

$$\sigma_0^2 = \frac{\mu_1\sigma_1^2 + \mu_2\sigma_2^2}{\mu_0}.$$

5.3.2 Introducing Traders with Informational Advantage

Since the primary value of last look is in protecting against traders who may be looking to exploit predictive informational advantages about future market moves we now introduce these into the model. We set a parameter $\pi \in [0, 1]$ depending on how prevalent these traders are. The greater π is the more often we expect to see sell orders just before price decreases and buy orders just before price increases. Precisely, given δS_t, the change in the market price, orders arrive at the following rates:

1. When $\delta S_t < 0$:
   (a) Sell orders arrive at rate $\frac{(1+\pi)}{2}\Lambda^b(\delta^b)$; and
   (b) Buy orders arrive at rate $\frac{(1-\pi)}{2}\Lambda^a(\delta^a)$.
2. When $\delta S_t > 0$:
   (a) Sell orders arrive at rate $\frac{(1-\pi)}{2}\Lambda^b(\delta^b)$; and
   (b) Buy orders arrive at rate $\frac{(1+\pi)}{2}\Lambda^a(\delta^a)$.

In order to get the rates to compute, we suppose δS_t is equally likely to be positive or negative. Then Type 2 orders with positive price moves arrive at rate $\mu_2/2$, and conditional on such an order arriving

- with probability $\frac{(1+\pi)\Lambda^a(\delta^a)}{\mu_2}$ it will be a buy order with the market maker;
- with probability $\frac{(1-\pi)\Lambda^b(\delta^b)}{\mu_2}$ it will be a sell order with the market maker; and
- with probability $1 - \frac{(1+\pi)\Lambda^a(\delta^a)+(1-\pi)\Lambda^b(\delta^b)}{\mu_2}$ it will not be an order with the market maker.

Similarly conditional on the price move being negative and a Type 2 order arriving then

- with probability $\frac{(1-\pi)\Lambda^a(\delta^a)}{\mu_2}$ it will be a buy order with the market maker;
- with probability $\frac{(1+\pi)\Lambda^b(\delta^b)}{\mu_2}$ it will be a sell order with the market maker; and
- with probability $1 - \frac{(1-\pi)\Lambda^a(\delta^a)+(1+\pi)\Lambda^b(\delta^b)}{\mu_2}$ it will not be an order with the market maker.

Note that in order for all of the terms above to be probabilities between 0 and 1 we need

$$(1+\pi)\Lambda^a(\delta^a)+(1-\pi)\Lambda^b(\delta^b) \leq \mu_2 \iff \pi(\Lambda^a(\delta^a)-\Lambda^b(\delta^b)) \leq \mu_2-(\Lambda^b(\delta^b)+\Lambda^a(\delta^a)),$$

and

$$(1-\pi)\Lambda^a(\delta^a)+(1+\pi)\Lambda^b(\delta^b) \leq \mu_2 \iff \pi(\Lambda^b(\delta^b)-\Lambda^a(\delta^a)) \leq \mu_2-(\Lambda^b(\delta^b)+\Lambda^a(\delta^a)),$$

or equivalently

$$\pi|\Lambda^b(\delta^b) - \Lambda^a(\delta^a)| \leq \mu_2 - (\Lambda^b(\delta^b) + \Lambda^a(\delta^a)). \tag{5.1}$$

We want to see the impact of the parameter π on the market maker's profitability and the interaction of their wealth and corresponding holdings and also the effect of a last look feature. We aim to consider limiting models that are continuous in time and space as in Chapter 4.

5.3.3 Modelling Last Look

We now introduce the last look feature, and define the way in which trades will be cancelled. We assume that the Last Look feature may cancel a trade if the price move is too large. That is, we choose $\xi_-^a, \xi_-^b < 0 < \xi_+^a, \xi_+^b$ such that a buy order is cancelled if

$$\delta S > \xi_+^a \text{ or } \delta S < \xi_-^a,$$

and similarly a sell order is cancelled if

$$\delta S > \xi_+^b \text{ or } \delta S < \xi_-^b.$$

In Figure 5-11 we show the result of simulating this model for a fixed control $\delta_a = 0.05$, $\delta_b = 0.1$, taking $T = 10$, $k = 20$, $A = 500$, $\sigma_0 = 0.9$, $\mu_0 = 15$, $\Delta = 5$. We note that the choice to simulate using a fixed control rather than an optimal one means that we see in our plots decreasing inventory. Were we to simulate with an optimal control we would expect to see mean-reverting inventory levels instead, but the simulation with a fixed control is sufficient to show the reasonableness of the model choice. The last look parameters ξ_a and ξ_b are chosen in such a way that they cancel approximately 2% of expected return trades in total ξ and do so symmetrically. In Figure 5-12 we also plot the correlation between moves in the asset price and changes to the market maker's inventory levels with and without last look.

Figures 5-11 and 5-12 demonstrate that implementing a LL feature seems to decrease the correlation between the asset price move and q, decrease the profit from trades via the bid-ask spread, and decrease the 'cost' associated with the anticipating trades.

Although it may not be entirely clear how each of these contributions affects the overall behaviour, we would expect that decreasing the correlation is desirable as is decreasing the costs associated with anticipating trades. We are interested to study further whether the decrease in the profit from cancelling trades is small enough to justify this trade off, and understanding this completely is non-trivial.

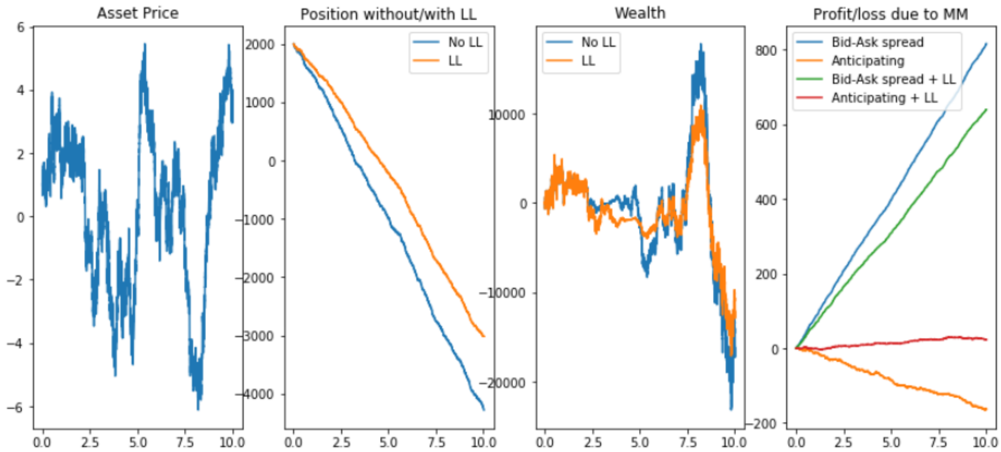


Figure 5-11: The first panel plots the random movements of the asset price. The second panel shows the inventory (which is declining because we are simulating a constant control rather than an optimal one) with and without last look. In the third panel we plot the resulting wealth process $(x_t + q_t S_t)$ with and without last look. In the fourth panel we see in blue and green the profit purely from the spread with and without last look. In orange we see the losses due to anticipating trades, and in red the profit from such trades when the last look mechanism is included.

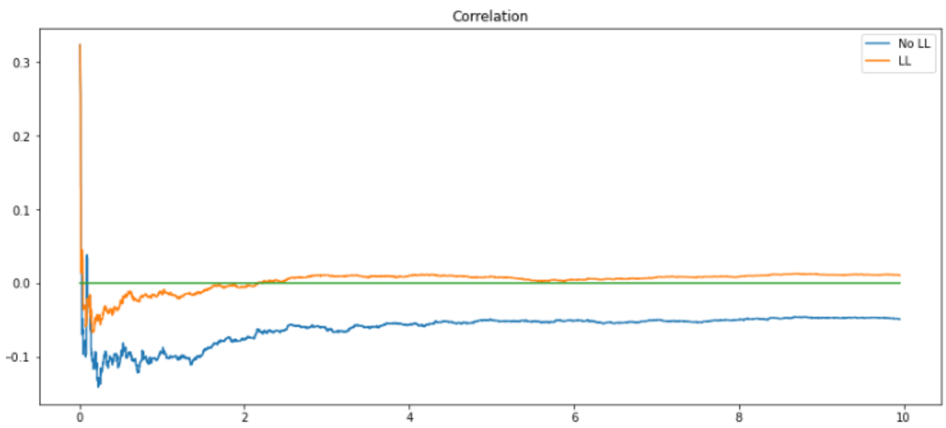


Figure 5-12: A plot showing the correlation between moves in the asset price and changes to the market maker's inventory levels with and without last look.

To consider this in more detail let us now generalise slightly.

5.3.4 Generalising the Model

We now suppose price moves are made up from J different types of price change. We fix N and suppose each price move occurs as a Poisson Point Process with rate $N\mu_i$, and that there is a corresponding sequence of times, $0 = t_0 \leq t_1 \leq \dots$, where $t_{i+1} - t_i$ is exponential with mean $(N\mu_0)^{-1}$, and $\mu_0 = \sum_{j=1}^{J} \mu_j$. Write $\mathcal{T} = \{t_1, t_2, \dots\}$. We let $\mathcal{D}_j$ be the set of times t_i for which there is a price move of type j in (t_i, t_{i+1}) and suppose at each time t_i, $t_i \in \mathcal{D}_j$ with probability $\frac{\mu_j}{\mu_0}$. Given $t_i \in \mathcal{D}_j$, our price move $\delta S_{t_i}^N$ has mean 0 and variance $N^{-1}\sigma_j^2$. Note that we do not necessarily assume Gaussian price moves, but we will assume δS is equally likely to be > 0 and < 0, with zero probability of being equal to zero.

We write $S_t^N = \sum_{i:t_i \leq t} \delta S_{t_i}^N$.

Proposition 5.3.1 *Letting $N \to \infty$, $(S_t^N)_{t\geq 0}$ converges in law to a scaled Brownian motion $(S_t)_{t\geq 0}$, i.e. $S_t =_{\mathcal{L}} \sigma_0 W_t$, for W_t a Brownian motion and*

$$\sigma_0^2 = \sum_{j=1}^{J} \frac{\mu_j}{\mu_0} \sigma_j^2.$$

Sketch Proof S_1 is the sum of (approximately) $\mu_0 N$ i.i.d. samples of a random variable with mean zero and variance

$$\frac{1}{N} \sum_{j=1}^{J} \frac{\mu_j}{\mu_0} \sigma_j^2,$$

by conditioning on which type we observe. (Convergence to Brownian motion follows from Donsker's Invariance Principle.)

□

We now consider the limiting behaviour of the inventory and wealth processes in N. Additionally, we suppose that there is the usual rate of buy and sell orders, which are $\Lambda^a(\delta^a)$ and $\Lambda^b(\delta^b)$ as above (both of which, we assume scale in N, through changes to the parameter A).

We assume that we are given $(\alpha_j)_{j=1,\dots,J}$, with $\alpha_j \geq 0$ and $\sum_j \alpha_j = 1$, $(\beta_j)_{j=1,\dots,J}$, with $\beta_j \geq 0$ and $\sum_j \beta_j = 1$, and we also suppose that the rate of orders $\Lambda^a(\delta^a) + \Lambda^b(\delta^b) < \mu_0$. Then α_j indicates that a proportion α_j of incoming price moves correspond to price moves of type j, so $\alpha_j\mu_0 = \mu_j$. Also β_j is the probability that an order is an order of type j. We also suppose we are given $(\pi_j)_{j=1,\dots,J}$, so that $0 \leq \pi_j \leq 1$, and then we suppose that π_j encodes the rate at which orders of type j are anticipative (from traders with informational advantage).

Write $t_i \in \mathcal{U}$ if $\delta S_{t_i} > 0$, and write $t_i \in \mathcal{B}$ if time t_i corresponds to a buy order, write $t_i \in \mathcal{S}$ if time t_i corresponds to a sell order and $\mathcal{N} = \mathcal{T} \setminus (\mathcal{S} \cup \mathcal{B})$ if there is no order. Then we have

- $t_i \in \mathcal{U} \cap \mathcal{D}_j \cap \mathcal{B}$ with probability $\frac{(1+\pi_j)\beta_j\Lambda^a(\delta^a)}{\mu_0}$,
- $t_i \in \mathcal{U} \cap \mathcal{D}_j \cap \mathcal{S}$ with probability $\frac{(1-\pi_j)\beta_j\Lambda^b(\delta^b)}{\mu_0}$,
- $t_i \in \mathcal{U} \cap \mathcal{D}_j \cap \mathcal{N}$ with probability

$$\begin{aligned} &\frac{\alpha_j}{2} - \frac{(1+\pi_j)\beta_j\Lambda^a(\delta^a) + (1-\pi_j)\beta_j\Lambda^b(\delta^b))}{\mu_0} \\ &= \frac{\alpha_j}{2} + \frac{\pi_j\beta_j(\Lambda^b(\delta^b) - \Lambda^a(\delta^a)) - \beta_j(\Lambda^a(\delta^a) + \Lambda^b(\delta^b))}{\mu_0}. \end{aligned}$$

- $t_i \in \mathcal{U}^C \cap \mathcal{D}_j \cap \mathcal{B}$ with probability $\frac{(1-\pi_j)\beta_j\Lambda^a(\delta^a)}{\mu_0}$,
- $t_i \in \mathcal{U}^C \cap \mathcal{D}_j \cap \mathcal{S}$ with probability $\frac{(1+\pi_j)\beta_j\Lambda^b(\delta^b)}{\mu_0}$,
- $t_i \in \mathcal{U}^C \cap \mathcal{D}_j \cap \mathcal{N}$ with probability

$$\begin{aligned} &\frac{\alpha_j}{2} - \frac{(1-\pi_j)\beta_j\Lambda^a(\delta^a) + (1+\pi_j)\beta_j\Lambda^b(\delta^b)}{\mu_0} \\ &= \frac{\alpha_j}{2} + \frac{\pi_j\beta_j(\Lambda^a(\delta^a) - \Lambda^b(\delta^b)) - \beta_j(\Lambda^a(\delta^a) + \Lambda^b(\delta^b))}{\mu_0}. \end{aligned}$$

Observe that in a similar way to (5.1), in order for all of the terms above to be probabilities between 0 and 1 we require

$$\pi_j|\Lambda^a(\delta^a) - \Lambda^b(\delta^b)| \leq \frac{\mu_j}{\beta_j} - (\Lambda^a(\delta^a) + \Lambda^b(\delta^b)).$$

We further introduce the Last Look feature. Fix $\xi^a_-, \xi^b_- < 0 < \xi^a_+, \xi^b_+$ such that a buy order is cancelled if

$$\delta S > \xi^a_+ \text{ or } \delta S < \xi^a_-,$$

and similarly a sell order is cancelled if

$$\delta S > \xi^b_+ \text{ or } \delta S < \xi^b_-.$$

Write χ_i to be an indicator of the event that a trade on the i^{th} move is not cancelled.

Note that we can compute the expectation and variance of price change for a typical observation by conditioning on the type of move as

$$\mathbb{E}[\delta S] = 0, \quad \mathbb{V}(\delta S) = \sum_j \frac{\mu_j}{\mu_0}\sigma_j^2.$$

Similarly, we can compute

$$\begin{aligned}\mathbb{E}[\delta q] = \frac{\Delta}{2\mu_0}\sum_j \Big[&(1-\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,+}(\xi^b_+) + (1+\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,-}(\xi^b_-) \\ &- (1+\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,+}(\xi^a_+) - (1-\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,-}(\xi^a_-)\Big],\end{aligned}$$

where $p_{j,+}(\xi)$ is the probability of an up move of type j being below ξ, and $p_{j,-}(\xi)$ is the probability of a down move of type j being above ξ.

Write

$$\bar{p}_+ = \frac{1}{2\mu_0}\sum_j \left[(1-\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,+}(\xi^b_+) + (1+\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,-}(\xi^b_-)\right]$$

$$\bar{p}_- = \frac{1}{2\mu_0}\sum_j \left[(1+\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,+}(\xi^b_+) + (1-\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,-}(\xi^a_-)\right].$$

Note that $\bar{p}_+$ is the probability that a sell trade happens, and $\bar{p}_-$ is the probability that a buy trade happens. Then we can write

$$\begin{aligned}\mathbb{V}[\delta q] &= \bar{p}_+(\Delta - \mathbb{E}[\delta q])^2 + \bar{p}_-(\Delta + \mathbb{E}[\delta q])^2 + (1 - \bar{p}_+ - \bar{p}_-)(\mathbb{E}[\delta q])^2 \\ &= (\bar{p}_+ + \bar{p}_-)\Delta^2 + (\bar{p}_+ - \bar{p}_-)\mathbb{E}[\delta q]\Delta + (\mathbb{E}[\delta q])^2. \end{aligned} \tag{5.2}$$

Claim 5.3.2 *The limiting behaviour of q is given by*

$$dq_t = a(\delta^a, \delta^b)dt + b(\delta^a, \delta^b)dZ_t,$$

where Z_t is a Brownian motion and

$$\begin{aligned} a(\delta^a, \delta^b) = \frac{\Delta}{2}\sum_j \Big[&(1-\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,+}(\xi^b_+) + (1+\pi_j)\beta_j\Lambda^b(\delta^b)p_{j,-}(\xi^b_-), \\ &- (1+\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,+}(\xi^a_+) - (1-\pi_j)\beta_j\Lambda^a(\delta^a)p_{j,-}(\xi^a_-)\Big] \end{aligned}$$

$$b(\delta^a, \delta^b)^2 = \bar{p}_+(\Delta - \mathbb{E}[\delta q])^2 + \bar{p}_-(\Delta + \mathbb{E}[\delta q])^2 + (1 - \bar{p}_+ - \bar{p}_-)(\mathbb{E}[\delta q])^2.$$

In order to test whether this model captures the behaviour of the system well we have computed numerically these terms and compared the value expected under this model to that of the simulation. In Figures 5-13 and 5-14 we show the plots of the actual and expected inventory (with the same parameter choices as before) and the 'detrended' inventory (that is the inventory reduced by its expected value) process compared to the variance expected as computed in (5.2). The actual and expected values are very close and persuade us that the model fits well and that Claim 5.3.2 is a reasonable one.

We also want to understand the dynamics of δW. This is made up of three terms, the first is $q_{t_i-}\delta S_{t_i}$, where q_{t_i-} is the number of units held before any additional orders at time t_i. The other contributions are from the bid-ask spread on any complete orders, computed as

$$\Delta(\delta^a\mathbb{E}[\chi_i; \mathcal{B}] + \delta^b\mathbb{E}[\chi_i; \mathcal{S}]) = \Delta(\delta^a\bar{p}_- + \delta^b\bar{p}_+).$$

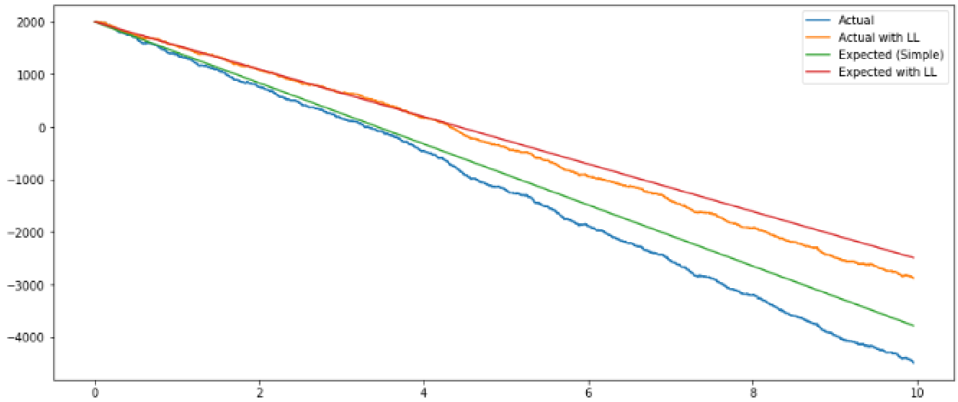


Figure 5-13: A plot of the expected value of the inventory level using this model against the simulated values, with and without last look. We note that all of the lines are decreasing since we have simulated using a fixed control, rather than with an optimal one.

And from the loss due to anticipating trades, given as

$$\frac{\Delta}{2\mu_0}\sum_j \Big[(1-\pi_j)\beta_j\Lambda^b(\delta^b)m_{j,+}(\xi^b_+) + (1+\pi_j)\beta_j\Lambda^b(\delta^b)m_{j,-}(\xi^b_-)$$
$$-(1+\pi_j)\beta_j\Lambda^a(\delta^a)m_{j,+}(\xi^a_+) - (1-\pi_j)\beta_j\Lambda^a(\delta^a)m_{j,-}(\xi^a_-)\Big],$$

where $m_{j,+}(\xi)$ is the expected value of δS, conditional on $\xi > \delta S > 0$ and $t_i \in \mathcal{D}_j$, and similarly for $m_{j,-}$, which is the expected value of δS, conditional on $\xi < \delta S < 0$. Note that $m_{j,+}(\xi) > 0$ and $m_{j,-}(\xi) < 0$. In the symmetric case, where e.g. ξ^b_- and ξ^b_+ are chosen so that $m_{j,+}(\xi^b_+) + m_{j,-}(\xi^b_-) = 0$, then the term above simplifies to

$$-\frac{\Delta}{\mu_0}\sum_j \pi_j\beta_j\left[\Lambda^b(\delta^b)m_{j,+}(\xi^b_+) + \Lambda^a(\delta^a)m_{j,+}(\xi^a_+)\right].$$

We can try to do the same thing for the variance of the quantities. We have

$$\delta W = q_{t_i-}\delta S + \delta q\delta S + (\delta q)_+\delta^b + (\delta q)_-\delta^a,$$

where $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$.

Note that using standard properties of variance, e.g. $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + \mathbb{C}ov(X,Y)$, and $\mathbb{C}ov(X,Y) \leq \sqrt{\mathbb{V}(X)\mathbb{V}(Y)}$, if $\mathbb{V}(X) \gg \mathbb{V}(Y)$, then

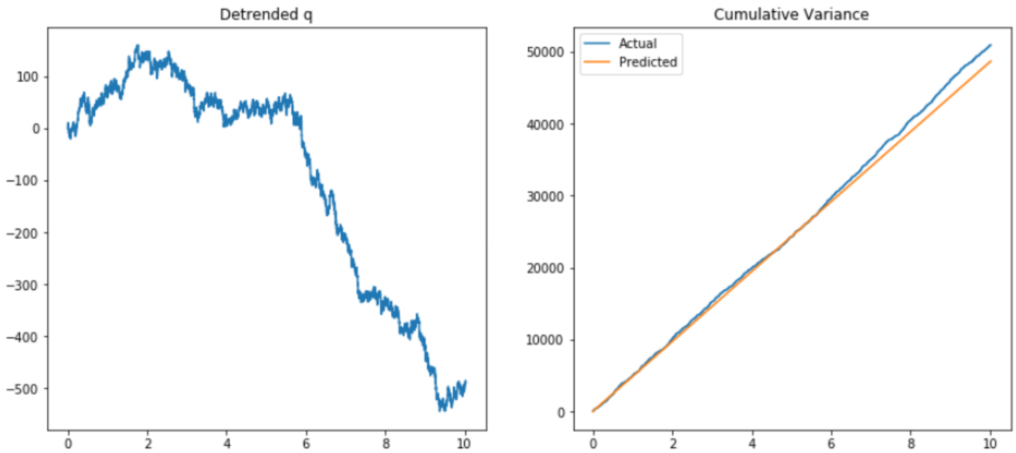


Figure 5-14: The left panel shows a plot of the 'detrended' inventory level. That is, the simulated inventory less its expected value. The right panel shows the cumulative variance from the simulation against that as calculated via (5.2).

$\mathbb{V}(X+Y) \approx \mathbb{V}(X)$. Since we expect $q \gg \delta q$, then $\mathbb{V}(\delta W) \approx q_{t_i}^2 \mathbb{V}(\delta S)$.

Finally, we analyse the correlation between the wealth W and the position q. Since $\delta W \approx q\delta S$, it is sufficient to consider the correlation between S and q.

Writing $\delta S = \sum_{1\le j\le J} \delta S^j \mathbf{1}_{\mathcal{D}_j}$ and $\delta q = \Delta \sum_{1\le j\le J} \mathbf{1}_{\mathcal{D}_j} \left[\mathbf{1}_{\mathcal{U}\cap\mathcal{S}} \mathbf{1}\{\delta S^j \in (0,\xi_+^b\} + \dots\right]$ we get

$$\begin{aligned}\mathbb{C}ov(\delta S\delta q) = \frac{\Delta}{2\mu_0}\sum_j \Big[&(1-\pi_j)\beta_j\Lambda^b(\delta^b)m_{j,+}(\xi_+^b) - (1+\pi_j)\beta_j\Lambda^b(\delta^b)m_{j,-}(\xi_-^b) \\ &- (1+\pi_j)\beta_j\Lambda^a(\delta^a)m_{j,+}(\xi_+^a) + (1-\pi_j)\beta_j\Lambda^a(\delta^a)m_{j,-}(\xi_-^a)\Big]. \qquad (5.3)\end{aligned}$$

Note the similarity to the loss due to anticipating trades above, except for the sign changes.

In Figure 5-15 we compare the correlation between the inventory and wealth process as simulated and as predicted by (5.3) and find an encouraging similarity.

Finally in Figure 5-16 we examine the drift terms in the wealth process. We see that changes in the wealth are a combination of changes to the value of inventory as a result of changes in the asset price as well as profits from market

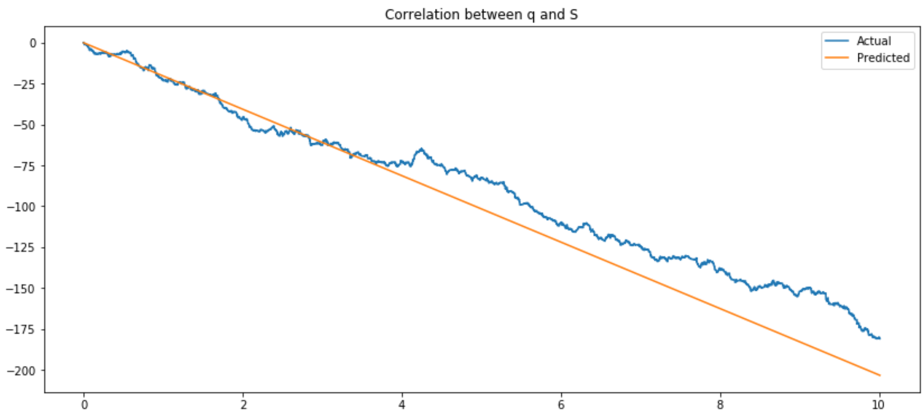


Figure 5-15: The plot shows the correlation between the inventory and wealth process as simulated and as predicted by (5.3).

making activity. To see these profits clearly we subtract $q_t \delta W_t$ from the profit and then can view that we have a good fit between the expected profits from market making and those simulated, and similarly when we look at just the anticipating trades.

5.3.5 A Suitable Continuous Model for Last Look

We conclude that the following model is a reasonable approximation to the micro-founded model.

$$dS_t = \sigma_0 dB_t, \quad dq_t = a(\delta_t^a, \delta_t^b)dt + b(\delta_t^a, \delta_t^b)dZ_t, \quad dW_t = q dS_t + \eta(\delta_t^a, \delta_t^b)dt,$$

where B_t, Z_t are Brownian motions with correlation ρ, and

$$\begin{aligned} a(\delta^a, \delta^b) = \frac{\Delta}{2} \sum_j \Big[&(1-\pi_j)\beta_j \Lambda^b(\delta^b) p_{j,+}(\xi_+^b) + (1+\pi_j)\beta_j \Lambda^b(\delta^b) p_{j,-}(\xi_-^b) \\ &- (1+\pi_j)\beta_j \Lambda^a(\delta^a) p_{j,+}(\xi_+^a) - (1-\pi_j)\beta_j \Lambda^a(\delta^a) p_{j,-}(\xi_-^a) \Big], \end{aligned}$$

$$b(\delta^a, \delta^b)^2 = \bar{p}_+(\Delta - \mathbb{E}[\delta q])^2 + \bar{p}_-(\Delta + \mathbb{E}[\delta q])^2 + (1 - \bar{p}_+ - \bar{p}_-)(\mathbb{E}[\delta q])^2,$$

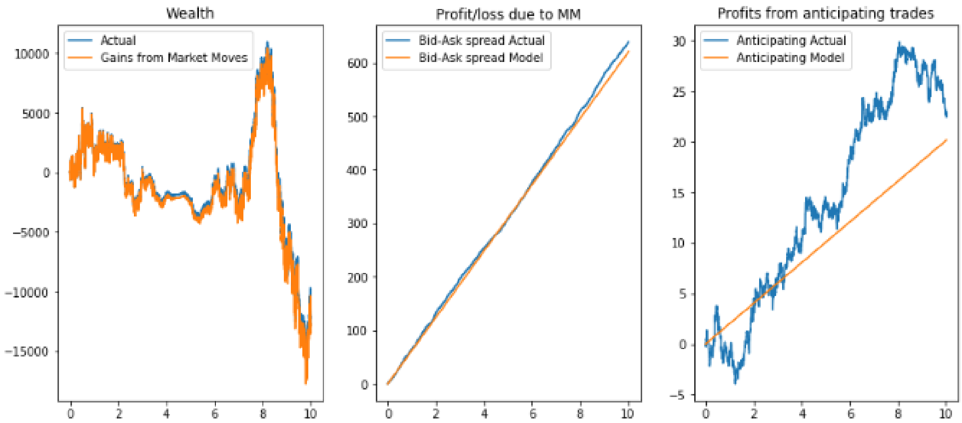


Figure 5-16: In the first panel we see that changes to the wealth arise from changes to the underlying asset price. We subtract $q_t S_t$ from this to see the profits arising from market making activity in the second panel. In the third panel we plot just the profits from anticipating trades as predicted and as simulated.

$$\begin{aligned}\eta(\delta^a, \delta^b) =& \Delta(\delta^a \bar{p}_- + \delta^b \bar{p}_+) \\ & - \frac{\Delta}{2} \sum_j \Big[(1-\pi_j)\beta_j \Lambda^b(\delta^b) m_{j,+}(\xi^b_+) + (1+\pi_j)\beta_j \Lambda^b(\delta^b) m_{j,-}(\xi^b_-) \\ & \quad - (1+\pi_j)\beta_j \Lambda^a(\delta^a) m_{j,+}(\xi^a_+) - (1-\pi_j)\beta_j \Lambda^a(\delta^a) m_{j,-}(\xi^a_-) \Big],\end{aligned}$$

$$\begin{aligned}\rho(\delta^a, \delta^b) = \frac{\Delta}{2\sigma_0 b(\delta^a, \delta^b)} \sum_j & \Big[(1-\pi_j)\beta_j \Lambda^b(\delta^b) m_{j,+}(\xi^b_+) - (1+\pi_j)\beta_j \Lambda^b(\delta^b) m_{j,-}(\xi^b_-) \\ & - (1+\pi_j)\beta_j \Lambda^a(\delta^a) m_{j,+}(\xi^a_+) + (1-\pi_j)\beta_j \Lambda^a(\delta^a) m_{j,-}(\xi^a_-) \Big].\end{aligned}$$

Note that most of these expressions can be rewritten in terms of the controls in the form

$$c_1 \Lambda^a(\delta^a) + c_2 \Lambda^b(\delta^b).$$

We note also that the plots we have presented in this section have been computed under a fixed control that we have chosen fairly arbitrarily. We expect that if we were to apply something closer to an optimal control that the inventory should be mean reverting to 0 and so we would expect to see more fluctuations in some of the plots.

The conclusion that we are encouraged by is that the model appears to be quite suitable for analysis in a similar way to the continuous model of Chapter 4. We leave a fuller exploration of this for future work, but we hope to be able to extend the work of Chapter 4 to this case, by providing a result similar to Proposition 4.4.1 that allows for this correlation between inventory and wealth process to be incorporated. Also we would like to use the work of Section 4.5 and the large deviation framework to be able to optimise over various parameters and understand sensitivity to the parameters chosen for the last look criteria and defining the toxicity of order flow.

5.4 Summary

In this chapter we have successfully adapted some of the major existing models of market making, as well as our own models to the last look case, finding natural adaptations of the optimal strategies and to the long run value of market making. In the final section we have also been able to propose a new model for market making with last look that can capture well situations where the market maker may be trading at an informational disadvantage. The way in which this model has been set up allows it to be analysed via the numerics of Chapter 4 and so provides a setting that could be very useful for optimising over the many parameters involved.

Appendix A

The Truncated Normal Distribution

In Chapter 5 we consider last look mechanisms which cancel trades when the normally distributed increments exceed certain levels. Many of the results that follow then naturally involve properties of the the truncated normal distribution. We collect below for easy reference the most important results we make use of.

Suppose that X is a normally distributed random variable with mean μ and variance σ^2 and $-\infty \leq a < b \leq \infty$. Then conditioning X on $a < X < b$ results in a truncated normal distribution. Below we set out separately results in the case of 'two-sided truncation' where $-\infty < a < b < \infty$ and of 'one-sided truncation' where either $= -\infty$ or $b = \infty$.

Two sided truncation
If $X \sim N(\mu, \sigma^2)$ we have that

$$\mathbb{E}(X|a < X < b) = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}.$$

We also make extensive use of the moment generating function

$$\mathbb{E}(e^{tX}|a < X < b) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left[\frac{\Phi\left(\frac{b-\mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{a-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right]$$

where Φ and ϕ are the standard normal c.d.f. and p.d.f. respectively.

One sided truncation

If $X \sim N(\mu, \sigma^2)$ we have that

$$\mathbb{E}(X|a < X) = \mu + \sigma\frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

and

$$\mathbb{E}(X|X < b) = \mu - \sigma\frac{\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}$$

as well as the moment generating functions

$$\mathbb{E}(e^{tX}|a < X) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left[\frac{1 - \Phi\left(\frac{a-\mu}{\sigma} - \sigma t\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right]$$

and

$$\mathbb{E}(e^{tX}|X < b) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left[\frac{\Phi\left(\frac{b-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}\right].$$

Appendix B

The Quantum Harmonic Oscillator

Some of the problems we will look at later will require us to consider eigenvalue problems for linear operators, and one that will appear frequently is that of the quantum harmonic oscillator $\mathcal{L}^{QHO}g : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, which we will define for $\beta \in \mathbb{R}$ as

$$\mathcal{L}^{QHO}g := g''(x) - \beta x^2 g(x).$$

The full spectral theory of such operators is well studied. The full list of eigenfunctions and corresponding eigenvalues can be specified, for example in terms of the Hermite polynomials (see e.g. [89] or [38]), but we will only make use of the principal eigenfunction and eigenvalue in the work that follows, so we state here for reference the following proposition.

Proposition B.0.1 *The leading (normalised) eigenfunction and eigenvalue pair* (f, λ) *of the operator* $\mathcal{L}^{QHO}$ *defined above are given by*

$$f(x) = \left(\frac{\beta}{\pi}\right)^{\frac{1}{4}} e^{-\frac{\sqrt{\beta}}{2}x^2}, \quad \lambda = -\sqrt{\beta}.$$

We do not prove this standard result but it is easy to verify that

$$f''(x) - \beta x^2 g(x) = \lambda f(x), \quad ||f(x)||_{L^2} = 1.$$

In fact, we will be working with operators of the form

$$\mathcal{L}g := \alpha g''(x) - (\kappa + \beta x^2)g(x)$$

and so for convenience later we also state the following slightly more general result that is an immediate consequence of Proposition B.0.1.

Proposition B.0.2 *The leading (normalised) eigenfunction and eigenvalue pair* (f, λ) *of the operator* $\mathcal{L}g := (\kappa + \beta x^2)g(x) - \alpha g''(x)$ *are given by*

$$f(x) = \left(\frac{\beta}{\alpha\pi}\right)^{\frac{1}{4}} e^{-\frac{1}{2}\sqrt{\frac{\beta}{\alpha}}x^2}, \quad \lambda = \kappa - \sqrt{\alpha\beta}.$$

In particular this pair satisfies

$$(\kappa + \beta x^2)f(x) - \alpha f''(x) = \lambda f(x) \quad ||f(x)||_{L^2} = 1.$$

and a proof of this proposition requires little more than matching terms to Proposition B.0.1.

Bibliography

[1] Marco Avellaneda and Sasha Stoikov. "High-frequency trading in a limit order book". In: *Quantitative Finance* 8.3 (2008), pp. 217–224.

[2] Kerry Back. "Insider trading in continuous time". In: *The Review of Financial Studies* 5.3 (1992), pp. 387–409.

[3] Kerry Back, C Henry Cao, and Gregory A Willard. "Imperfect competition among informed traders". In: *The journal of finance* 55.5 (2000), pp. 2117–2155.

[4] Kerry Back and Hal Pedersen. "Long-lived information and intraday patterns". In: *Journal of financial markets* 1.3-4 (1998), pp. 385–402.

[5] Shmuel Baruch. "Insider trading and risk aversion". In: *Journal of Financial Markets* 5.4 (2002), pp. 451–464.

[6] Richard Bellman. *An introduction to the theory of dynamic programming.* Tech. rep. RAND CORP SANTA MONICA CA, 1953.

[7] Richard Bellman. "On the theory of dynamic programming". In: *Proceedings of the national Academy of Sciences* 38.8 (1952), pp. 716–719.

[8] A Bensoussan, J Frehse, and H Nagai. "Some results on risk-sensitive control with full observation". In: *Applied Mathematics and Optimization* 37.1 (1998), pp. 1–41.

[9] Philippe Bergault et al. "Closed-form approximations in multi-asset market making". In: *arXiv preprint arXiv:1810.04383* (2018).

[10] Philippe Bergault et al. "Closed-form approximations in multi-asset market making". In: *Applied Mathematical Finance* 28.2 (2021), pp. 101–142.

[11] Lorenzo Bertini, Alessandra Faggionato, and Davide Gabrielli. "Large deviations of the empirical flow for continuous time Markov chains". In: *Annales de l'IHP Probabilités et statistiques.* Vol. 51. 3. 2015, pp. 867–900.

[12] Tomasz R Bielecki and Stanley R Pliska. "Risk-sensitive dynamic asset management". In: *Applied Mathematics and Optimization* 39.3 (1999), pp. 337–360.

[13] FF Bonsall. "Linear operators in complete positive cones". In: *Proceedings of the London mathematical Society* 3.1 (1958), pp. 53–75.

[14] David Borthwick. *Spectral theory of infinite-area hyperbolic surfaces.* Springer, 2007.

[15] Bruno Bouchard and Nizar Touzi. "Weak dynamic programming principle for viscosity solutions". In: *SIAM Journal on Control and Optimization* 49.3 (2011), pp. 948–962.

[16] Jean-Philippe Bouchaud et al. *Trades, quotes and prices: financial markets under the microscope.* Cambridge University Press, 2018.

[17] Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations.* Vol. 2. 3. Springer, 2011.

[18] René Caldentey and Ennio Stacchetti. "Insider trading with a random deadline". In: *Econometrica* 78.1 (2010), pp. 245–283.

[19] Alvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. "Enhancing trading strategies with order book signals". In: *Applied Mathematical Finance* 25.1 (2018), pp. 1–35.

[20] Alvaro Cartea, Sebastian Jaimungal, and Jamie Walton. "Foreign exchange markets with last look". In: *Mathematics and Financial Economics* 13.1 (2019), pp. 1–30.

[21] Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. "Algorithmic trading with model uncertainty". In: *SIAM Journal on Financial Mathematics* 8.1 (2017), pp. 635–671.

[22] Álvaro Cartea and Sebastian Jaimungal. "Risk metrics and fine tuning of high-frequency trading strategies". In: *Mathematical Finance* 25.3 (2015), pp. 576–611.

[23] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and high-frequency trading.* Cambridge University Press, 2015.

[24] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. "Buy low, sell high: A high frequency trading perspective". In: *SIAM Journal on Financial Mathematics* 5.1 (2014), pp. 415–444.

[25] Umut Çetin and Albina Danilova. *Dynamic Markov Bridges and Market Microstructure: Theory and Applications*. Vol. 90. Springer, 2018.

[26] Umut Çetin and Albina Danilova. "Markovian Nash equilibrium in financial markets with asymmetric information and related forward–backward systems". In: (2016).

[27] Raphaël Chetrite and Hugo Touchette. "Nonequilibrium Markov processes conditioned on large deviations". In: *Annales Henri Poincaré*. Vol. 16. 9. Springer. 2015, pp. 2005–2057.

[28] Kyung-Ha Cho. "Continuous auctions and insider trading: uniqueness and risk aversion". In: *Finance and Stochastics* 7 (2003), pp. 47–71.

[29] Pierre Collin-Dufresne and Vyacheslav Fos. "Insider trading, stochastic liquidity and equilibrium prices". In: (2014).

[30] Global Foreign Exchange Committee. *FX Global Code 2018*. accessed 18th August 2022. URL: `https://www.globalfxc.org/docs/fx_global.pdf`.

[31] José Manuel Corcuera et al. "Kyle-Back's model with Lévy noise". In: *Preprint series. Pure mathematics http://urn. nb. no/URN: NBN: no-8076* (2010).

[32] Harald Cramér. "Sur un nouveau théoreme-limite de la théorie des probabilités". In: *Actual. Sci. Ind.* 736 (1938), pp. 5–23.

[33] Ian Crowther and Marc Glover. "Rights of last look: a legitimate tool or abuse of the auction process?" English. In: *Butterworths Journal of International Banking and Financial Law* 27.10 (Nov. 2012), pp. 608–611. ISSN: 0269-2694.

[34] Marcos Lopez De Prado. *Advances in financial machine learning*. John Wiley & Sons, 2018.

[35] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Vol. 38. Springer Science & Business Media, 2009.

[36] Paul Dupuis and William M McEneaney. "Risk-sensitive and robust escape criteria". In: *SIAM journal on control and optimization* 35.6 (1997), pp. 2021–2049.

[37] Joanne Elliott. "Eigenfunction expansions associated with singular differential operators". In: *Transactions of the American Mathematical Society* 78.2 (1955), pp. 406–425.

[38] Daniel A Fleisch. *A Student's Guide to the Schrödinger Equation*. Cambridge University Press, 2020.

[39] Wendell H Fleming and William M McEneaney. "Risk-sensitive control on an infinite time horizon". In: *SIAM Journal on Control and Optimization* 33.6 (1995), pp. 1881–1915.

[40] Wendell H Fleming and SJ Sheu. "Risk-sensitive control and an optimal investment model". In: *Mathematical Finance* 10.2 (2000), pp. 197–213.

[41] Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*. Vol. 25. Springer Science & Business Media, 2006.

[42] Pietro Fodra and Mauricio Labadie. "High-frequency market-making for multi-dimensional markov processes". In: *arXiv preprint arXiv:1303.7177* (2013).

[43] Pietro Fodra and Mauricio Labadie. "High-frequency market-making with inventory constraints and directional bets". In: *arXiv preprint arXiv:1206.4810* (2012).

[44] Pietro Fodra and Huyên Pham. "High frequency trading and asymptotics for small risk aversion in a Markov renewal model". In: *SIAM Journal on Financial Mathematics* 6.1 (2015), pp. 656–684.

[45] Avner Friedman. *Partial differential equations of parabolic type*. Courier Dover Publications, 2008.

[46] Xavier Gabaix et al. "Institutional investors and stock market volatility". In: (2010).

[47] Mark B Garman. "Market microstructure". In: *Journal of financial Economics* 3.3 (1976), pp. 257–275.

[48] Lawrence R Glosten and Paul R Milgrom. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". In: *Journal of financial economics* 14.1 (1985), pp. 71–100.

[49] Parameswaran Gopikrishnan et al. "Statistical properties of share volume traded in financial markets". In: *Physical review e* 62.4 (2000), R4493.

[50] Sanford J Grossman and Merton H Miller. "Liquidity and market structure". In: *the Journal of Finance* 43.3 (1988), pp. 617–633.

[51] Olivier Guéant. "Optimal market making". In: *Applied Mathematical Finance* 24.2 (2017), pp. 112–154.

[52] Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. “Dealing with the inventory risk: a solution to the market making problem”. In: *Mathematics and financial economics* 7.4 (2013), pp. 477–507.

[53] Fabien Guilbaud and Huyen Pham. “Optimal high-frequency trading with limit and market orders”. In: *Quantitative Finance* 13.1 (2013), pp. 79–94.

[54] Allen Hart, James Hook, and Jonathan Dawes. “Embedding and approximation theorems for echo state networks”. In: *Neural Networks* 128 (2020), pp. 234–247.

[55] Allen G Hart, James L Hook, and Jonathan HP Dawes. “Echo state networks trained by Tikhonov least squares are L2 (μ) approximators of ergodic dynamical systems”. In: *Physica D: Nonlinear Phenomena* 421 (2021), p. 132882.

[56] Allen G Hart et al. “Using Echo State Networks to Approximate Value Functions for Control”. In: *arXiv preprint arXiv:2102.06258* (2021).

[57] Thomas Ho and Hans R Stoll. “Optimal dealer pricing under transactions and return uncertainty”. In: *Journal of Financial economics* 9.1 (1981), pp. 47–73.

[58] Frank den Hollander. *Large Deviations (Fields Institute Monographs)*. English. Paperback. American Mathematical Society, June 12, 2008, p. 146. ISBN: 978-0821844359.

[59] Roger D Huang and Hans R Stoll. “The components of the bid-ask spread: A general approach”. In: *The Review of Financial Studies* 10.4 (1997), pp. 995–1034.

[60] Hitoshi Ishii. “On the equivalence of two notions of weak solutions, viscosity solutions and distribution solutions”. In: *Funkcial. Ekvac* 38.1 (1995), pp. 101–120.

[61] David Jacobson. “Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games”. In: *IEEE Transactions on Automatic control* 18.2 (1973), pp. 124–131.

[62] Nicole El-Karoui and Said Hamadène. “BSDEs and risk-sensitive control, zero-sum and nonzero-sum game problems of stochastic functional differential equations”. In: *Stochastic Processes and their Applications* 107.1 (2003), pp. 145–169.

[63] Arturo Kohatsu-Higa and Salvador Ortiz-Latorre. "Weak Kyle–back equilibrium models for Max and ArgMax". In: *SIAM Journal on Financial Mathematics* 1.1 (2010), pp. 179–211.

[64] Albert S Kyle. "Continuous auctions and insider trading". In: *Econometrica: Journal of the Econometric Society* (1985), pp. 1315–1335.

[65] Olga A Ladyvzenskaja, Vsevolod Alekseevich Solonnikov, and Nina N Uraltseva. *Linear and quasi-linear equations of parabolic type*. Vol. 23. American Mathematical Soc., 1988.

[66] Baron Law and Frederi Viens. "Market making under a weakly consistent limit order book model". In: *High Frequency* 2.3-4 (2019), pp. 215–238.

[67] Charles-Albert Lehalle and Sophie Laruelle. *Market microstructure in practice*. World Scientific, 2018.

[68] Jacques Louis Lions and Enrico Magenes. *Non-homogeneous boundary value problems and applications: Vol. 1*. Vol. 181. Springer Science & Business Media, 2012.

[69] XTX Markets. *Aggretation and Liquidity in the FX Market*. accessed 21 Feb 2020. 2019. URL: `https://www.xtxmarkets.com/wp-content/uploads/2019/08/Aggregation-and-Liquidity-Management-in-FX.pdf`.

[70] XTX Markets. *XTX Markets Disclosures on eFX Trading Practices*. accessed 16th December 2022. URL: `https://www.xtxmarkets.com/articles/disclosures/`.

[71] Sergei Maslov and Mark Mills. "Price fluctuations from the order book perspective—empirical facts and a simple model". In: *Physica A: Statistical Mechanics and its Applications* 299.1-2 (2001), pp. 234–246.

[72] Hideo Nagai. "Bellman equations of risk-sensitive control". In: *SIAM journal on control and optimization* 34.1 (1996), pp. 74–101.

[73] Roger D Nussbaum. "Eigenvectors of nonlinear positive operators and the linear Krein-Rutman theorem". In: *Fixed point theory*. Springer, 1981, pp. 309–330.

[74] *NYDFS Last Look Settlement (website)*. accessed 20 Feb 2020. URL: `http://barclaysdfslastlooksettlement.com/`.

[75] Kaj Nyström, Sidi Mohamed Ould Aly, and Changyong Zhang. "Market making and portfolio liquidation under uncertainty". In: *International Journal of Theoretical and Applied Finance* 17.05 (2014), p. 1450034.

[76] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[77] Bernt Øksendal and Agnès Sulem. "Stochastic Control of Jump Diffusions". In: *Applied Stochastic Control of Jump Diffusions*. Springer, 2019, pp. 93–155.

[78] Roel Oomen. "Execution in an aggregator". In: *Quantitative Finance* 17.3 (2017), pp. 383–404.

[79] Roel Oomen. "Last look". In: *Quantitative Finance* 17.7 (2017), pp. 1057–1070.

[80] Beresford N Parlett. *The symmetric eigenvalue problem*. SIAM, 1998.

[81] Huyên Pham. "A risk-sensitive control dual approach to a large deviations control problem". In: *Systems & control letters* 49.4 (2003), pp. 295–309.

[82] Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*. Vol. 61. Springer Science & Business Media, 2009.

[83] Ross G Pinsky. *Positive harmonic functions and diffusion*. Vol. 45. Cambridge university press, 1995.

[84] Robert J Plemmons. "M-matrix characterizations. I—nonsingular M-matrices". In: *Linear Algebra and its applications* 18.2 (1977), pp. 175–188.

[85] Marc Potters and Jean-Philippe Bouchaud. "More statistical properties of order books and price impact". In: *Physica A: Statistical Mechanics and its Applications* 324.1-2 (2003), pp. 133–140.

[86] Paul H Rabinowitz and Henri Berestycki. *Théorie du degré topologique et applications à des problèmes aux limites non linéaires*. Université Paris VI, Laboratoire analyse numérique, 1975.

[87] Natasha Rega-Jones. *How the top 50 liquidity providers tackle last look*. accessed 21 Feb 2020. 2019. URL: https://www.risk.net/derivatives/6888556/how-the-top-50-liquidity-providers-tackle-forex-last-look.

[88] Helmut H Schaefer. "Topological Vector Spaces (4th printing)". In: *New York* (1980).

[89] Franz Schwabl. *Quantum mechanics*. Springer Science & Business Media, 2007.

[90] Avanidhar Subrahmanyam. "Risk aversion, market liquidity, and price efficiency". In: *The Review of Financial Studies* 4.3 (1991), pp. 417–441.

[91] John F Toland. "Self-adjoint operators and cones". In: *Journal of the London Mathematical Society* 53.1 (1996), pp. 167–183.

[92] Nizar Touzi and Agnès Tourin. *Optimal stochastic control, stochastic target problems, and backward SDE.* Vol. 29. Springer, 2013.

[93] Philipp Weber and Bernd Rosenow. "Order book approach to price impact". In: *Quantitative Finance* 5.4 (2005), pp. 357–364.

[94] Peter Whittle. "Risk-sensitive linear/quadratic/Gaussian control". In: *Advances in Applied Probability* 13.4 (1981), pp. 764–777.

Chapter 6

Publication (Alternative Format Thesis Section)

This final Chapter 6 is presented 'by publication'. The full text of the paper [56] is included, which is a joint work with Hart, A.G., Cox, A.M.G., Isupova, O. and Dawes, J.H.P, and to which the author of this thesis contributed around 20% of the work. A further statement of authorship is included after these introductory remarks and before the paper.

At the time of publication this paper is under review. The paper involves some novel results about Echo State Networks, which are a type of single-layer recurrent neural network with randomly chosen internal weights and a trainable output layer. The results about the theory of ESNs are primarily attributable to other authors.

The paper also includes some applications of Echo State Networks, and in particular the author of this thesis contributed most significantly to the development of and application to the market making problem presented as well as to the overall idea for the collaboration and the organisation and preparation of the paper.

This PhD thesis as well as that of the co-author Allen Hart, was undertaken with funding the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), grant number EP/L015684/1. The paper in this

chapter was also born out of the collaborative environment of SAMBa. Hart had been working on mathematical results about Echo State Networks and had an idea in mind of a deterministic system to which they could be applied. The author of this thesis had begun to think about market making models and suggested adapting the results to include a stochastic case, and then the collaborative work began along with their supervisors in applying the results and adapting the ESN theory to suit the stochastic case.

The technical results surrounding the ESNs in the paper are primarily those of Hart, building on his previous work in [54] and [55] although there was also joint work in extending these to the stochastic case. The author's main contribution to the paper however is in Section 5, providing a simplified mathematical framework driven by the market making problem and applying and interpreting the outputs of the ESN in this case. At the time of writing that paper, much of the work in this thesis was only partly formed and so the model there was driven by the intuition about the market making problem, but since then we have now also been able to given further justification that this intuition was sound.

Appendix B: Statement of Authorship

<table>
<tr><td colspan="4">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="4">Using Echo State Networks to Approximate Value Functions for Control Problems</td></tr>
<tr><td colspan="4">Publication status (tick one)
Draft manuscript [ ] Submitted [ ] In review [X] Accepted [ ] Published [ ]</td></tr>
<tr><td>Publication details (reference)</td><td colspan="3">Hart, A.G.; Olding, K.R.; Cox, A.M.G.; Isupova, O.; Dawes, J.H.P. Using Echo State Networks to Approximate Value Functions for Control Problems. arXiv 2021, arXiv:2102.06258.

Under review at `Neural Networks’</td></tr>
<tr><td colspan="4">Copyright status (tick the appropriate statement)
The material has been published with a CC-BY license [X] The publisher has granted permission to replicate the material included here [ ]</td></tr>
<tr><td>Candidate’s contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="3">Formulation of ideas (20%):

The candidate played an integral role in the formulation of ideas for the paper. The more theoretical work on Echo State Networks is primarily the work of Hart and his supervisors (Dawes and Isupova) but the candidate contributed very significantly to the other sections of the paper, in particular the idea for and the execution of the application to the stochastic control problem and in structuring and managing the collaboration between authors.

Design of methodology (20%):

The candidate considerably contributed to the stochastic control application section of the paper. He also contributed to other sections of the paper and the overall organisation of the ideas contained in it, but his primary contributions were the stochastic control section.

Experimental work (10%):

The candidate contributed to the numerical work in the paper through discussions about the methods and interpretation of the results, though was not directly involved with producing the code used.

Presentation of data in journal format:
N/A</td></tr>
<tr><td>Statement from Candidate</td><td colspan="3">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed (typed signature)</td><td>Kevin Olding</td><td>Date</td><td>1st December 2022</td></tr>
</table>

Using Echo State Networks to Approximate Value Functions for Control

Allen G. Hart[a], Kevin R. Olding[a], Alexander M.G. Cox[a], Olga Isupova[b], Jonathan H.P. Dawes[a]

[a]*Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK*
[b]*Department of Computer Science, University of Bath, Bath, BA2 7AY, UK*

Abstract

An Echo State Network (ESN) is a type of single-layer recurrent neural network with randomly-chosen internal weights and a trainable output layer. We prove under mild conditions that a sufficiently large Echo State Network can approximate the value function of a broad class of stochastic and deterministic control problems. Such control problems are generally non-Markovian.

We describe how the ESN can form the basis for novel and computationally efficient reinforcement learning algorithms in a non-Markovian framework. We demonstrate this theory with two examples. In the first, we use an ESN to solve a deterministic, partially observed, control problem which is a simple game we call 'Bee World'. In the second example, we consider a stochastic control problem inspired by a market making problem in mathematical finance. In both cases we can compare the dynamics of the algorithms with analytic solutions to show that even after only a single reinforcement policy iteration the algorithms arrive at a good policy.

Keywords: Liquid State Machines, Reservoir Computing, Stochastic Optimal Control, Mathematical Finance, Reinforcement Learning

---

1. Introduction

An Echo State Network (ESN) is a special type of single-layer recurrent neural network introduced at the turn of the millennium by [1] and [2] to study time series. Training is fast because the training step involves only the selection of weights in the output layer rather than updating the internal weights in the recurrent layer. Furthermore, the simple formulation of ESNs renders them amenable to mathematical analysis. Given a time series z_k (where k is the discrete time index) of d-dimensional data points, an ESN is set up as follows. We randomly generate a $n \times n$ reservoir matrix $\boldsymbol{A}$, a $n \times d$ input matrix $\boldsymbol{C}$ and a $n \times 1$ bias vector $\boldsymbol{\zeta}$. Then we iteratively generate a sequence of n-dimensional *reservoir state* vectors x_k according to

$$x_{k+1} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}z_k + \boldsymbol{\zeta})$$

where $\sigma(x)_i = \max(0, x_i)$ is the rectified linear unit (ReLU) activation function applied component-wise to the n-dimensional vector x. Observe that the kth reservoir state x_k depends on all past data-points $\dots, z_{k-2}, z_{k-1}$ and therefore captures non-Markovian temporal correlations in the data. If the 2-norm of the reservoir matrix satisfies $\|\boldsymbol{A}\|_2 < 1$ then as n tends to infinity, the influence on the reservoir state x_{k+n} of the data points $\dots, z_{k-2}, z_{k-1}$ in the distant past becomes arbitrarily small. This is called the *fading memory property* and is closely related to the *echo state property* (ESP) introduced in the context of ESNs by [1]. The ESP is the statement that the sequence of reservoir states $(x_k)_{k\in\mathbb{Z}}$ is, for a given input data sequence $(z_k)_{k\in\mathbb{Z}}$, uniquely determined. We can interpret the reservoir state vectors as the latent vectors which encode the infinite past observations in lower dimensional form.

When an ESN has the ESP, it can be applied to a class of supervised learning problems where we have a time series of d dimensional data points r_k, called *targets*, that depend on all previous input time series data $\dots, z_{k-3}, z_{k-2}, z_{k-1}$ and we seek to learn the relationship between the sequence of past states and the target for each k. We can train an ESN to solve this problem by finding the $m \times d$ matrix W that minimises

$$\sum_{k=0}^{\ell-1} \|W^\top x_k - r_k\|^2 + \lambda\|W\|^2,$$

where ℓ is the number of labelled data points, and $\lambda > 0$ is the Tikhonov regularisation (a.k.a. ridge regression) parameter. Throughout this paper, $\|\cdot\|$ denotes the matrix 2-norm, vector 2-norm or absolute value, depending on whether the input is a matrix, vector, or scalar, respectively.

This minimisation problem can be solved using regularised linear least squares regression, and hence we can both obtain W quickly, and guarantee that W is the global optimum. This compares extremely favourably with training a (deep) neural network with stochastic gradient descent and backpropagation which takes considerably longer, and may not converge to the global optimum [3].

Despite the training procedure being entirely linear, ESNs are universal approximators, and can therefore model arbitrarily complex relationships between the sequence of past data points and the targets. This is made formal in a recent result by [4] that we review here and then build on. We emphasise that not only are ESNs theoretically very promising, they have performed remarkably well in practice on problems ranging from seizure detection, to robot control, handwriting recognition, and financial forecasting, where ESNs have won competitions [5], [6], [7], [8]. Impressively, ESNs outperformed RNNs and LSTMs at a chaotic time series prediction task by a factor of over 2400 [9]. ESNs have also proved themselves competitive in various tasks in reinforcement learning [10] and control [11].

Even in cases where practitioners prefer to use other recurrent neural networks (RNNs), such as Long Short Term Memory networks (LSTMs), the rigorous theory of ESNs should prove useful in architecture design. In [12], it is shown that different deep neural network architectures can be ranked by randomly initialising the internal weights and training only the outer weights by linear regression. Once the best performing architecture (with random internal weights) has been identified, the authors then train the internal weights of the highest ranking architecture. This is much faster than training the internal weights (a nonlinear problem) for every architecture. The ranking of architectures with random internal weights closely approximates the ranking of architectures with optimised internal weights. From our point of view, the authors are essentially approximating fully trained networks with (non-recurrent) ESNs.

In a sequence of papers, [13], [14], and [4] recently analysed ESNs in the context of nonlinear filters and functionals. Roughly speaking, a *filter* U is a map from a bi-infinite sequence $\dots, z_{-2}, z_{-1}, z_0, z_1, z_2, \dots$ of real vectors to another bi-infinite sequence of real vectors $\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots$, and a *functional* H maps a bi-infinite sequence $\dots, z_{-2}, z_{-1}, z_0, z_1, z_2, \dots$ of real vectors to a single real vector or number. We can view an ESN as a filter that maps an input sequence $\dots, z_{-2}, z_1, z_0, z_1, z_2, \dots$ to a reservoir sequence $\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots$, or a funtional that maps $\dots, z_{-2}, z_1, z_0, z_1, z_2, \dots$ to the lone reservoir state x_0. The theory of filters and functionals is therefore a natural theoretical setting for ESNs. Within this theory, this paper presents three novel results.

Our first result assumes that we have a time series of data z_k and a set of targets r_k that depend on all previous data points $\dots, z_{k-2}, z_{k-1}$ via a functional $\mathcal{R}$ which sends infinite sequences of data points to targets. We then have a supervised learning problem of finding the relationship between the data and targets. In the special case that $z_k = r_k$, this problem is *time series forecasting*. Our first novel result states that if we have sufficiently many data points z_k, drawn from a stationary, ergodic, and bounded process $\boldsymbol{Z}$, which need not be Markovian, and we obtain W using regularised linear least squares, then a sufficiently large ESN will approximate, as closely as required, the functional $\mathcal{R}$ sending inputs $\dots, z_{k-2}, z_{k-1}$ to the targets r_k.

This result has applications in the *statistical inference of dynamical systems*, which was recently reviewed by [15]. This area of research is especially focused on statistical inference (i.e learning) of stationary ergodic processes. Furthermore, we can use this result in the context of reinforcement learning (RL) and optimal control. We envisage an agent operating under a given policy in the parlance of reinforcement learning or control in the parlance of control theory that generates a sequence of (reward, action, observation) triples $z_k = (r_k, a_k, \omega_k)$. Then the functional V that maps previous (reward, action, observation) triples

$\ldots, z_{k-2}, z_{k-1}$ to rewards z_k models the reward functional arbitrarily well. The set up does not assume the RL problem is Markovian, and allows for a continuous state space.

Our second novel result generalises the first, and encompasses the case where the functional V is the value functional of a stochastic control process, or Partially Observed Markov Decision Process (POMDP). By training an ESN to approximate the value functional, we establish a stepping stone toward developing an offline reinforcement learning algorithm supported by an ESN that can solve a large class of control problems. Moreover, since ESNs are recurrent, they can be used for non-Markovian problems, where a reinforcement learning agent must exploit its *memory* of past observations, actions and rewards. Our third result is presented in the context of building an online reinforcement algorithm that can, under certain conditions, determine the optimal value function for a given policy.

These results are part of a general push to take machine learning ideas typically applied to (partially observed) Markov processes and generalising them to hold on stationary ergodic processes. We can see for example [16] consider to clustering problems typically defined Markov processes applied to stationary ergodic processes.

We demonstrate some of these theoretical results numerically on two examples. The first is a deterministic game which we call 'Bee World'. The goal of the game for the bee is to navigate a time varying distribution of nectar in order to maximise the total future discounted value of the nectar acquired over all future time. The optimal trajectory can be found explicitly via the calculus of variations but the constraint that the bee has a maximum speed of flight leads to unexpectedly complicated solution paths; it therefore provides a straightforward but not entirely trivial control problem. Since the bee does not have access to the entire state space, and only observes the nectar it collects at each moment in time, the problem is therefore a partially observed Markov Decision Process which requires memory of the past to solve. We demonstrate how a simple and easily-configurable reinforcement learning algorithm supported by an ESN can learn to play Bee World with respectable skill.

The second numerical example is inspired by a market making problem in mathematical finance. The mathematical formulation of this problem reduces to a seeking to control a one dimensional Brownian motion so that it stays near the origin. The cost of straying from the origin is quadratic in the distance from the origin, and the cost of applying a push toward the origin is quadratic in the strength of the push. The market maker must therefore balance the cost of applying the control against the cost of allowing the motion to drift too far from the origin. We briefly discuss the financial motivation for this problem, then solve it analytically in continuous and discrete time. The set up most commonly seen in the literature is continuous time, but only in discrete time is the problem suitable for an ESN. We then compare the optimal discrete time solution to a solution learned by a reinforcement learning agent supported by an ESN.

Finally, we note that our approach to the Market making problem is loosely related to the recent paper by [11] who introduce QuaSiModO: Quantization-Simulation-Modeling-Optimization. These authors analyse the interplay between the following four aspects:

1. Quantising the action space $\mathcal{A}$.
2. Simulating a system under a given control/policy.

3. Modelling the full system given a partial/full observation of the state space.
4. Optimising the control/policy.

The structure of the remainder of the paper closely follows the summary of results presented above. In section 2 we set up the mathematical formalism for ESNs that we wish then to approximate. Section 3 introduces our novel theoretical results, while sections 4 and 5 respectively present applications to the deterministic ('Bee World'), and then the stochastic ('market maker') optimal control problems. We conclude in section 6.

2. Background

In this section, we introduce the theory and notation of nonlinear filters (in relation to ESNs) developed by [13], [14], and [4]. First, we denote by $(\mathbb{R}^d)^{\mathbb{Z}}$ the set of maps with domain $\mathbb{Z}$ and codomain $\mathbb{R}^d$. This is the set of bi-infinite $\mathbb{R}^d$–valued real sequences.

A *filter* is a map $U : (\mathbb{R}^d)^{\mathbb{Z}} \to (\mathbb{R}^n)^{\mathbb{Z}}$. A filter U is called *causal* if inputs from the past and present $\dots, z_{-2}, z_{-1}, z_0$ contribute to $U(z)$ but states in the future $z_1, z_2 \dots$ do not. More formally U is casual if $\forall\ z, y \in (\mathbb{R}^d)^{\mathbb{Z}}$ that satisfy $z_k = y_k\ \forall\ k \leq 0$ it follows that $U(z) = U(y)$. We define the time shift filter $T : (\mathbb{R}^d)^{\mathbb{Z}} \to (\mathbb{R}^n)^{\mathbb{Z}}$ by $T(z)_k = T(z)_{k+1}$ which we interpret as the map that steps forward one unit of time. A filter U is called *time invariant* if U commutes with the time shift operator T. If U is causal and time invariant filter then we call U a causal time invariant (CTI) filter.

A functional is a map $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}^n$. In [14] it is shown that there is a bijection between the space of CTI filters and the space of functionals. To see this, take a functional H and define the kth term of the associated filter U via $U(z)_k = HT^k(z)$. Conversely, given a filter U, the associated functional H is given by $H(z) = U(z)_0$

We can view an ESN as a CTI filter from the space of input sequences $\dots, z_{-1}, z_0, z_1, \dots$ to the space of reservoir sequences $\dots, x_{-1}, x_0, x_1, \dots$. To make this connection between ESNs and filters formal, we will first present a generalisation of an Echo State Network called a reservoir system.

Definition 2.1. (Reservoir system) Let $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^s$. Then we call the following system of equations

$$\begin{aligned} x_{k+1} &= F(x_k, z_k) \\ r_k &= h(x_k) \end{aligned} \tag{1}$$

a reservoir system.

Remark 2.2. We can see that if

$$\begin{aligned} F(x, z) &= \sigma(\boldsymbol{A}x + \boldsymbol{C}z + \boldsymbol{\zeta}) \\ h(x) &= W^\top x \end{aligned}$$

then we retrieve an ESN with $n \times n$ reservoir matrix $\boldsymbol{A}$, $n \times d$ input matrix $\boldsymbol{C}$, bias vector $\boldsymbol{\zeta} \in \mathbb{R}^n$, linear output layer $W \in \mathbb{R}^n$, and activation function $\sigma = \text{ReLU}$, defined in the introduction.

We require that the reservoir system induces a *unique* filter from the input sequence to the reservoir sequence. This property is the Echo State Property that we briefly mentioned in the introduction.

Definition 2.3. (Echo State Property [1]) A reservoir system has the Echo State Property (ESP) if for any $(z_k)_{k\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}}$ there exists a unique $(x_k)_{k\in\mathbb{Z}} \in (\mathbb{R}^n)^{\mathbb{Z}}$ that satisfy the equations of the reservoir system (1).

To any reservoir system with the Echo State property we can associate a unique CTI reservoir filter $U : (\mathbb{R}^d)^{\mathbb{Z}} \to (\mathbb{R}^n)^{\mathbb{Z}}$ defined by $U(z) = x$. To this reservoir filter, we may assign a CTI reservoir functional $H : (\mathbb{R}^m)^{\mathbb{Z}} \to \mathbb{R}^d$ defined by $H(z) = x_0$. In a supervised learning context, we have a time series of data points $\ldots, z_{-2}, z_{-1}, z_0$ and a time series of targets $\ldots, r_{-1}, r_0$ that each depend on *all* previous data points. The output functional $h \circ H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ is the map we use to approximate the relationship between the data and the targets, so $h \circ H(\ldots, z_{-2}, z_{-1}, z_0, z_1, z_2, \ldots) \approx r_k$. Note that $h \circ H$ is *causal*, so does not peer into the future and use data $z_1, z_2, \ldots$ that have not yet been revealed. When the reservoir system is an ESN, the map h is the linear map $W^\top$ obtained by least squares ridge regression, so that $W^\top H(\ldots, z_{-2}, z_{-1}, z_0, z_1, \ldots) \approx r_k$. We assume there exists a *true* map from the data to the targets that we label $\mathcal{R} : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ so that $\mathcal{R}(\ldots, z_{-2}, z_{-1}, z_0, z_1, \ldots) = r_k$. Our goal is to find W such that $W^\top H \approx \mathcal{R}$.

Definition 2.4. (ESN filter and functional) If an ESN has the ESP then we will write $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ to denote the reservoir functional associated to an ESN with parameters $\boldsymbol{A}, \boldsymbol{C}$ and $\boldsymbol{\zeta}$. We will also write $H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ to denote the output functional $W^\top H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ (defined by left multiplication of $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ by the linear readout layer)

Next, we will present a procedure, introduced by [4], for randomly generating the ESN's internal weights $\boldsymbol{A}, \boldsymbol{C}$ and biases $\boldsymbol{\zeta}$, which ensures the ESN has ESP and allows for the universal approximation of target functionals $\mathcal{R}$. The procedure differs from the procedure commonly seen in the literature, where $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ are populated with i.i.d Gaussians, or i.i.d uniform deviates, and then $\boldsymbol{A}$ is rescaled so that its 2-norm (or spectral radius) is less than 1. Furthermore, the procedure introduced by [4] depends on some details of the input process, which must satisfy mild conditions stated below.

Definition 2.5. (Admissible input process) A $(\mathbb{R}^d)^{\mathbb{Z}}$ valued random variable $\boldsymbol{Z}$ is called an admissible process if for any $T \in \mathbb{N}$ there exists $M_T > 0$ such that for all $k \in \mathbb{Z}$

$$\|\boldsymbol{Z}_{k-T}, \boldsymbol{Z}_{k-T+1}, \ldots, \boldsymbol{Z}_k\| \leq M_T \tag{2}$$

Lebesgue-almost surely.

We will now present a procedure by which the matrices $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ are randomly generated.

Procedure: Initialising the random weights of an ESN.

Let $N \in \mathbb{N}$, $R > 0$ be the input parameters for the procedure. Suppose that $\boldsymbol{Z}$ is an admissible input process. Consequently, for any $T_0 \in \mathbb{N}$ there exists M_{T_0} such that (for $k = 0$ in (2))

$$\|\boldsymbol{Z}_{-T_0}, \boldsymbol{Z}_{-T_0+1}, \ldots, \boldsymbol{Z}_0\| \leq M_T$$

Lebesgue-almost surely. Then, for a given T_0, we initialise the ESN reservoir matrix $\boldsymbol{A}$, input matrix $\boldsymbol{C}$, and biases $\boldsymbol{\zeta}$ according to the following procedure.

1. Draw N i.i.d. samples $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$ from the uniform distribution on $B_R \subset \mathbb{R}^{d(T_j+1)}$ where B_R is the ball of radius R and centre 0, and draw N i.i.d. samples $\boldsymbol{\zeta}_1, \ldots \boldsymbol{\zeta}_N$ from the uniform distribution on $[-\max(M_{T_0}R, 1), \max(M_{T_0}R, 1)]$.
2. Let S and c be shift matrices defined

$$S = \begin{bmatrix} 0_{d,dT_0} & 0_{d,d} \\ I_{dT_0} & 0_{dT_0,d} \end{bmatrix} \qquad c = \begin{bmatrix} I_d \\ 0_{dT_0,d} \end{bmatrix}$$

and set

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{A}_1^\top \\ \boldsymbol{A}_2^\top \\ \vdots \\ \boldsymbol{A}_N^\top \end{bmatrix} \qquad \bar{\boldsymbol{A}} = \begin{bmatrix} S & 0_{d(T_0+1),N} \\ \boldsymbol{a}S & 0_{N,N} \end{bmatrix}$$

$$\bar{\boldsymbol{C}} = \begin{bmatrix} c \\ \boldsymbol{a}c \end{bmatrix} \qquad \bar{\boldsymbol{\zeta}} = \begin{bmatrix} 0_{d(T_0+1)} \\ \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_N \end{bmatrix}$$

so that

$$\boldsymbol{A} = \begin{bmatrix} \bar{\boldsymbol{A}} & -\bar{\boldsymbol{A}} \\ -\bar{\boldsymbol{A}} & \bar{\boldsymbol{A}} \end{bmatrix} \qquad \boldsymbol{C} = \begin{bmatrix} \bar{\boldsymbol{C}} \\ -\bar{\boldsymbol{C}} \end{bmatrix} \qquad \boldsymbol{\zeta} = \begin{bmatrix} \bar{\boldsymbol{\zeta}} \\ -\bar{\boldsymbol{\zeta}} \end{bmatrix}.$$

We are now ready to present the key result by [4], (which generalises a result by [17]) and which holds in the following supervised learning context. Given time series data z_k (from an admissible process $\boldsymbol{Z}$) and a time series of targets r_k depending on all previous data $\ldots, z_{k-2}, z_{k-1}$ we wish to approximate the functional that sends $\ldots, z_{k-2}, z_{k-1}$ to r_k. We will denote this functional $\mathcal{R}$. The problem of approximating $\mathcal{R}$ given the data and targets is a supervised learning problem. The result can be summarised as follows. Suppose we have an ESN with weights $\boldsymbol{A}, \boldsymbol{C}$ and biases $\boldsymbol{\zeta}$ randomly generated by procedure 1. Then, the ESN admits a linear readout matrix W for which the ESN equipped with the matrix W (denoted $H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$) approximates the relationship $\mathcal{R}$ between data points $\ldots, z_{k-2}, z_{k-1}$ and targets r_k as closely as is required.

Theorem 2.6 ([4])**.** *Suppose that $\boldsymbol{Z}$ is an admissible input process. Let $\mathcal{R} : (D_n)^{\mathbb{Z}} \to \mathbb{R}$ (where D_n is a compact subset of $\mathbb{R}^n$) be CTI and measurable with respect to some measure μ such that $\mathbb{E}_\mu[|\mathcal{R}(\boldsymbol{Z})|^2] < \infty$.*

Then for any $\epsilon > 0$ and $\delta \in (0,1)$ there exists $N, T_0 \in \mathbb{N}$, $R > 0$ such that, with probability $(1-\delta)$, the ESN with parameters $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ generated by the procedure in definition 1 (with inputs N, T_0, R) has the ESP and admits a readout layer $W \in \mathbb{R}^{2(d(T_0+1)+N)}$ such that

$$\left(\mathbb{E}_\mu\left[\left\|H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \mathcal{R}(\boldsymbol{Z})\right\|^2 \,\middle|\, \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}\right]\right)^{1/2} := \left(\int_{(\mathbb{R}^d)^{\mathbb{Z}}} \left\|H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(z) - \mathcal{R}(z)\right\|^2 d\mu(z)\right)^{1/2} < \epsilon.$$

3. Novel results for ESNs

Theorem 2.6 is an existence result stating that *there exists* a linear readout layer W yielding an arbitrarily good approximation. Our first novel contribution is to strengthen the result under additional assumptions. The new result states that, given a sufficiently large ESN and sufficiently many training data z_k drawn from a stationary, ergodic and bounded process $\boldsymbol{Z}$, if we train an ESN using regularised least squares then the arbitrarily good readout layer W *will* be attained (with probability as close to 1 as desired). This result is analogous to the main result by [18] who prove a similar theorem for ESNs trained on deterministic inputs. Before we introduce the result we will present the definition of a stationary process, an ergodic process, and the ergodic theorem.

Definition 3.1. (Stationary Process [15]) A stochastic process $(\boldsymbol{Z}_k)_{k\in\mathbb{Z}} \equiv \boldsymbol{Z}$ is stationary if for any $\ell \in \mathbb{N}$ and finite subset $I \subset \mathbb{Z}$ the joint distribution $(\boldsymbol{Z}_i)_{i\in I}$ is equal to the joint distribution $(\boldsymbol{Z}_{i+\ell})_{i\in I}$.

Definition 3.2. (Stationary Ergodic Process [15]) A stationary stochastic process $(\boldsymbol{Z}_k)_{k\in\mathbb{Z}} \equiv \boldsymbol{Z}$ is called ergodic if for every $\ell \in \mathbb{N}$ and every pair of Borel sets A, B

$$\begin{aligned}\lim_{\ell\to\infty} \frac{1}{\ell} \sum_{k=0}^{\ell-1} &\mathbb{P}\Big((\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_\ell) \in A, (\boldsymbol{Z}_k, \dots, \boldsymbol{Z}_{k+\ell}) \in B\Big) \\ =&\mathbb{P}\Big((\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_\ell) \in A\Big)\mathbb{P}\Big((\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_\ell) \in B\Big).\end{aligned}$$

Every stationary ergodic processes Z satisfies the celebrated Ergodic Theorem.

Theorem 3.3. *(Ergodic Theorem) If $(\boldsymbol{Z}_k)_{k\in\mathbb{Z}} \equiv \boldsymbol{Z}$ is a stationary ergodic process then for any $i \in \mathbb{Z}$*

$$\mathbb{E}_\mu[\boldsymbol{Z}_i] = \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{k=0}^{\ell-1} \boldsymbol{Z}_{i+k}$$

almost surely.

Our result holds in the following supervised learning context. Given time series data z_k (from an admissible, stationary, ergodic, bounded process $\boldsymbol{Z}$) and a time series of targets r_k depending on all previous data $\dots, z_{k-2}, z_{k-1}$ we wish to approximate the mapping from $\dots, z_{k-2}, z_{k-1}$ to r_k. This mapping is denoted $\mathcal{R}$. Our result states that an ESN with weights $\boldsymbol{A}, \boldsymbol{C}$ and biases $\boldsymbol{\zeta}$ randomly generated by the procedure in definition 1, which is fed the training data z_k, and then trained by regularised least squares, will yield a matrix W. This ESN equipped with the matrix W (denoted $H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$) will approximate the relationship $\mathcal{R}$ between data points $\dots, z_{k-2}, z_{k-1}$ and targets r_k as closely as required.

Theorem 3.4. *Suppose that $\boldsymbol{Z}$ is an admissible input process, that is also stationary and ergodic, with invariant measure μ. Let $\mathcal{R} : (D_n)^{\mathbb{Z}} \to \mathbb{R}$ (where D_n is a compact subset of $\mathbb{R}^n$) be CTI, μ-measurable, and satisfy $\mathbb{E}_\mu[|\mathcal{R}(\boldsymbol{Z})|^2] < \infty$. Let z be an arbitrary realisation of $\boldsymbol{Z}$*

Then for any $\epsilon > 0$ and $\delta \in (0,1)$ there exist $N, T_0 \in \mathbb{N}$, $R > 0$, $\lambda^* > 0$ and $\ell \in \mathbb{N}$ such that the ESN with parameters $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ generated by the procedure in Definition 1 (with inputs N, T_0, R), and $W_\ell^* \in \mathbb{R}^{2(d(T_0+1)+N)}$ which minimises (over $W \in \mathbb{R}^{2(d(T_0+1)+N)}$) the least squares problem

$$\frac{1}{\ell}\sum_{k=0}^{\ell-1} \left\| H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \mathcal{R}T^{-k}(z) \right\|^2 + \lambda \left\| W \right\|^2,$$

(where $\lambda \in (0, \lambda^*)$) satisfies with probability $(1-\delta)$ the inequality

$$\mathbb{E}_\mu \left[\left\| H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \mathcal{R}(\boldsymbol{Z}) \right\|^2 \middle| \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] < \epsilon.$$

Proof. Later in this paper, we state and prove a more general result (Theorem 3.6) which reduces to this result in the special case $\gamma = 0$. □

In summary, we have stated that for any $\epsilon > 0$ and $\delta \in (0,1)$ there exists an ESN of dimension $n = 2(d(T_0+1)+N)$ with output layer W trained by the Tikhonov-regularised least squares procedure against ℓ training points, whose output functional approximates the target arbitrarily closely with arbitrarily high probability. The theorem is (sadly) non constructive in the sense that the number of neurons n, number of training points ℓ and regularisation parameter λ^* are not computed for a given ϵ and δ. Ideally, we would establish uniform bounds on the number of number of neurons n and data points ℓ required for an approximation with tolerance ϵ to hold with probability δ. Though less ideal, one could establish an asymptotic order of convergence using the central limit theorem (CLT). The CLT (roughly) states that the error between the time average and the space average of a stationary ergodic process converges in law to a normal distribution with standard deviation of the order $1/\sqrt{\ell}$ as the number of data points ℓ grows to infinity. The CLT is stated below.

Theorem 3.5. *(Central Limit Theorem [15]) If $(\boldsymbol{Z}_k)_{k\in\mathbb{Z}}$ is a stationary ergodic process then there exists a covariance matrix Σ such that for any $i \in \mathbb{Z}$ and Borel set A*

$$\lim_{\ell\to\infty} \mathbb{P}\bigg(\frac{1}{\sqrt{\ell}} \sum_{k=0}^{\ell-1} (\boldsymbol{Z}_{i+k} - \mathbb{E}_\mu[\boldsymbol{Z}_i]) \bigg) = \mathbb{P}\big(\mathcal{N}(0,\Sigma) \in A\big).$$

In other words, the random variables

$$\frac{1}{\sqrt{\ell}} \sum_{k=0}^{\ell-1} (\boldsymbol{Z}_{i+k} - \mathbb{E}_\mu[\boldsymbol{Z}_i])$$

converge in distribution to the multivariate normal $\mathcal{N}(0, \Sigma)$ *as* $\ell \to \infty$*.*

This suggests that the approximation of the target functional $\mathcal{R}$ also converges with order $1/\sqrt{\ell}$ as the number of data points increases. Furthermore, related results by [4] use the CLT to establish uniform bounds on the number of neurons $n = 2(d(T+1)+N)$ required for a given approximation. This strongly suggests that the approximation in Theorem 3.4 converges with order $1/\sqrt{N}$.

We will now pivot towards our second novel result, which generalises the first. Suppose that we have a contraction mapping Φ on the space of functionals, and we seek a W^* such that the ESN functional $H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ approximates the unique fixed point of Φ. The existence of the unique fixed point is guaranteed by Banach's fixed point theorem. Finding the fixed point of a contraction mapping has applications in reinforcement learning because the optimal value function (and optimal quality function) of a Markov Decision Process (MDP) is a fixed point of a *Bellman operator*. The theory we are presenting here can be viewed as a generalisation of an MDP because the input processes we are considering may have long time correlations (violating the Markov property) which can only be recognised by filters with sufficiently long and robust memories; like Echo State Networks.

We can observe first of all if Φ is the constant map $\Phi(H) = \mathcal{R}$, then Φ is clearly a contraction mapping with fixed point $\mathcal{R}$. In this case, the problem is exactly the same as that solved by Theorem 3.4. We are especially interested in the case of Φ taking the form of the *Bellman Value operator*. To make this formal, we will consider a stationary ergodic process $\boldsymbol{Z}$ with invariant measure μ. Then we define the map $T_{\boldsymbol{Z}}$ as a CTI filter on the bi-infinite sequences $(D_N)^{\mathbb{Z}}$, which returns the random variable:

$$T_{\boldsymbol{Z}}(z)_k = \begin{cases} T_{\boldsymbol{Z}}(z)_{k+1} & \text{if } k < 0 \\ \boldsymbol{Z}_{k+1} \mid \boldsymbol{Z}_j = z_j \ \forall j \leq 0 & \text{if } k \geq 0. \end{cases}$$

Next, we introduce $\mathcal{R} : (D_N)^{\mathbb{Z}} \to \mathbb{R}$ as the CTI reward functional, giving a reward (or expectation over a distribution of rewards) to an agent that has observed a given sequence of (reward, action, observation) triples. We let $\gamma \in [0,1)$ denote the discount factor, and define the operator

$$\Phi(H)(z) := \mathcal{R}(z) + \gamma \mathbb{E}_\mu[H T_{\boldsymbol{Z}}(z)]. \tag{3}$$

In this case, Φ is a contraction mapping with Lipschitz constant γ. With this, we will define the CTI value functional $V : (D_N)^{\mathbb{Z}} \to \mathbb{R}$ (with respect to the process $\boldsymbol{Z}$) as

$$V(z) := \mathbb{E}_\mu\left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R} T^k(\boldsymbol{Z}) \;\middle|\; \boldsymbol{Z}_j = z_j \ \forall j \leq 0 \right].$$

The value functional V takes a sequence of (reward, action, observation) triples and returns the expected discounted sum of future rewards. Furthermore, the value function V is the unique fixed point of the Bellman operator Φ. Re-arranging the definition of $V(z)$ above, we have that:

$$\begin{aligned}
V(z) &= \mathbb{E}_\mu\left[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}T^k(\boldsymbol{Z})\ \middle|\ \boldsymbol{Z}_j = z_j\ \forall j\leq 0\right]\\
&= \mathbb{E}_\mu\left[\sum_{k=1}^{\infty}\gamma^k\mathcal{R}T^k(\boldsymbol{Z})\ \middle|\ \boldsymbol{Z}_j = z_j\ \forall j\leq 0\right] + \mathcal{R}(z)\\
&= \gamma\mathbb{E}_\mu\left[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}T^{k+1}(\boldsymbol{Z})\ \middle|\ \boldsymbol{Z}_j = z_j\ \forall j\leq 0\right] + \mathcal{R}(z)\\
&= \gamma\mathbb{E}_\mu\left[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}T^{k}(\boldsymbol{Z})\ \middle|\ \boldsymbol{Z}_j = z_j\ \forall j< 0\right] + \mathcal{R}(z)
\end{aligned}$$

where we have carried out straightforward relabellings of the indexing of terms in the sum by k. Then by the law of total expectation we may write this last expression as

$$\begin{aligned}
V(z) &= \gamma\mathbb{E}_\mu\left[\mathbb{E}_\mu\left[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}T^k(\boldsymbol{Z})\ \middle|\ \boldsymbol{Z}_j = T_{\boldsymbol{Z}}(z)_j\ \forall j\leq 0\right]\right] + \mathcal{R}(z)\\
&= \gamma\mathbb{E}_\mu[VT_{\boldsymbol{Z}}(z)] + \mathcal{R}(z) = \Phi(V)(z),
\end{aligned}$$

which shows that V is indeed a fixed point of Φ, and so is the unique such, since Φ is a contraction.

Our goal is now to seek a W^* such that the ESN functional $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}$ closely approximates the unique fixed point V of Φ. One approach is to collect a dataset from a single training trajectory, and then perform least squares regression to find W^*. This is an example of *offline* learning (in the reinforcement learning parlance) because the training occurs after the data has been collected. This is in contrast to *online* learning where training takes place dynamically as new data becomes available. We will make this offline approach formal in the following theorem.

Theorem 3.6. *Suppose that $\boldsymbol{Z}$ is an admissible input process, that is also stationary and ergodic with invariant measure μ. Let $\mathcal{R} : (D_N)^{\mathbb{Z}} \to \mathbb{R}$ be μ-measurable and satisfy $\mathbb{E}[|\mathcal{R}(\boldsymbol{Z})|^2] < \infty$ and define Φ using* (3) *on the μ-measurable functionals H that satisfy $\mathbb{E}_\mu[|H(\boldsymbol{Z})|^2] < \infty$. Let $\gamma \in [0, 1)$. Let z be an arbitrary realisation of $\boldsymbol{Z}$*

Then for any $\epsilon > 0$, $\delta \in (0, 1)$ there exists $N, T_0 \in \mathbb{N}, R, \lambda^* > 0$ and $\ell \in \mathbb{N}$ such that the ESN with parameters $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ generated by procedure 1 (with inputs N, T_0, R), and $W^*_\ell \in \mathbb{R}^{2(d(T_0+1)+N)}$ minimising (over $W \in \mathbb{R}^{2(d(T_0+1)+N)}$) the least squares problem

$$\frac{1}{\ell}\sum_{k=0}^{\ell-1}\left\|W^\top(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}(z)\right\|^2 + \lambda\|W\|^2$$

where $\lambda \in (0, \lambda^*)$*, then with probability* $(1-\delta)$

$$\mathbb{E}_\mu \left[\left\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*_\ell}(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*_\ell}(\boldsymbol{Z}) \right\|^2 \middle| \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] < \epsilon.$$

Proof. First let V be the unique fixed point of the contraction mapping Φ whose existence and uniqueness is guaranteed by Banach's fixed point theorem. Denote the Lipschitz constant of Φ with the symbol τ. Then we fix $\epsilon > 0$ and $\delta \in (0,1)$, then by Theorem 2.6 there exists with probability $(1-\delta)$ a linear readout $W \in \mathbb{R}^{2(d(T_0+1)+N)}$ such that

$$\mathbb{E}_\mu \left[\left\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - V(\boldsymbol{Z}) \right\|^2 \middle| \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] < \frac{\epsilon}{5(1+\tau)}. \tag{4}$$

Then it follows that

$$\begin{aligned}
&\mathbb{E}_\mu \left[\left\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W} - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W} \right\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \\
&\quad = \mathbb{E}_\mu \left[\left\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) + V(\boldsymbol{Z}) - V(\boldsymbol{Z}) \right\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \\
&\quad \leq \mathbb{E}_\mu[\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - V(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] + \mathbb{E}_\mu[\| V(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\quad = \mathbb{E}_\mu[\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - V(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] + \mathbb{E}_\mu[\| \Phi V(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\quad \leq \mathbb{E}_\mu[\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - V(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] + \tau \mathbb{E}_\mu[\| V(\boldsymbol{Z}) - H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\quad = (1+\tau)\mathbb{E}_\mu[\| V(\boldsymbol{Z}) - H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\quad < (1+\tau)\frac{\epsilon}{5(1+\tau)} \text{ by (4)} \\
&\quad < \frac{\epsilon}{5}
\end{aligned}$$

which yields the estimate

$$\mathbb{E}_\mu[\| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W} - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] < \frac{\epsilon}{5}. \tag{5}$$

Now, we can choose λ^* such that for any $\lambda \in (0, \lambda^*)$

$$\lambda \|W\|^2 < \frac{\epsilon}{5}. \tag{6}$$

Next we define a sequence of vectors $(W^*_j)_{j\in\mathbb{N}}$ by

$$W^*_j = \underset{U \in \mathbb{R}^{2(d(T_0+1)+N)}}{\arg\min} \left(\frac{1}{j} \sum_{k=0}^{j-1} \| H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{U} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{U} T^{1-k}(z) - \mathcal{R} T^{-k}(z) \|^2 + \lambda \|U\|^2 \right).$$

We may view $\arg\min$ as continuous map on the space of strictly convex C^1 functions that returns their unique minimiser. The regularised linear least squares problem is a strictly convex C^1 problem, so we may define $W_\infty^* \in \mathbb{R}^{2d(T_0+1)+N}$ by

$$\begin{aligned}
W_\infty^* &:= \arg\min_U \left(\mathbb{E}_\mu[\|H_U^{A,C,\zeta}(\boldsymbol{Z}) - \gamma H_U^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z}) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] + \lambda \|U\|^2 \right) \\
&= \arg\min_U \lim_{j\to\infty} \left(\frac{1}{j} \sum_{k=0}^{j-1} \|H_U^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H_U^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k} - \mathcal{R} T^{-k}(z)\|^2 + \lambda\|U\|^2 \right) \\
&= \lim_{j\to\infty} \arg\min_U \left(\frac{1}{j} \sum_{k=0}^{j-1} \|H_U^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H_U^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k} - \mathcal{R} T^{-k}(z)\|^2 + \lambda\|U\|^2 \right) \\
&= \lim_{j\to\infty} W_j^*
\end{aligned}$$

where the second and third equalities hold by the Ergodic Theorem and continuity of arg min respectively. Now, we may choose $\ell \in \mathbb{N}$ sufficiently large that

$$\begin{aligned}
\Big| &\mathbb{E}_\mu[\|W_\ell^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z})) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&- \mathbb{E}_\mu[\|W_\infty^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z})) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \Big| < \frac{\epsilon}{5},
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
\Bigg| &\lim_{j\to\infty} \left(\frac{1}{j} \sum_{k=0}^{j-1} \|W_j^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k}(z)) - \mathcal{R} T^{-k}(z)\|^2 + \lambda \|W_j^*\|^2 \right) \\
&- \frac{1}{\ell} \sum_{k=0}^{\ell-1} \|W_\ell^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k}(z)) - \mathcal{R} T^{-k}(z)\|^2 + \lambda \|W_\ell^*\|^2 \Bigg| < \frac{\epsilon}{5},
\end{aligned} \tag{8}$$

and by the Ergodic Theorem

$$\begin{aligned}
\Bigg| &\frac{1}{\ell} \sum_{k=0}^{\ell-1} \|W^\top(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k}(z)) - \mathcal{R}(z)\|^2 \\
&- \lim_{j\to\infty} \frac{1}{j} \sum_{k=0}^{j-1} \|W^\top(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k}(z)) - \mathcal{R}(z)\|^2 \Bigg| < \frac{\epsilon}{5}.
\end{aligned} \tag{9}$$

Now the proof proceeds directly

$$\begin{aligned}
&\mathbb{E}_\mu[\|H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \Phi H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\qquad = \mathbb{E}_\mu[\|H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z}) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\qquad = \mathbb{E}_\mu[\|W_\ell^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z})) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}].
\end{aligned}$$

Then we apply (7) which yields

$$\begin{aligned}
&\mathbb{E}_\mu[\|H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \Phi H_{W_\ell^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] \\
&\qquad < \mathbb{E}_\mu[\|W_\infty^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T(\boldsymbol{Z})) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}] + \frac{\epsilon}{5}.
\end{aligned}$$

Then we apply the Ergodic Theorem

$$
\begin{aligned}
&\mathbb{E}_{\mu}[\|W_{\infty}^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T(\boldsymbol{Z})) - \mathcal{R}(\boldsymbol{Z})\|^2 | \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}] + \frac{\epsilon}{5} \\
&\quad = \lim_{j\to\infty}\left(\frac{1}{j}\sum_{k=0}^{j-1}\|W_{\infty}^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2\right) + \frac{\epsilon}{5} \\
&\quad \leq \lim_{j\to\infty}\left(\frac{1}{j}\sum_{k=0}^{j-1}\|W_{\infty}^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2\right) + \lambda\|W_{\infty}^{*}\|^2 + \frac{\epsilon}{5} \\
&\quad = \lim_{j\to\infty}\left(\frac{1}{j}\sum_{k=0}^{j-1}\|W_{j}^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2 + \lambda\|W_{j}^{*}\|^2\right) + \frac{\epsilon}{5} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{then apply (8)} \\
&\quad < \frac{1}{\ell}\sum_{k=0}^{\ell-1}\|W_{\ell}^{*\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2 + \lambda\|W_{\ell}^{*}\|^2 + \frac{2\epsilon}{5} \\
&\quad \leq \frac{1}{\ell}\sum_{k=0}^{\ell-1}\|W^{\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2 + \lambda\|W\|^2 + \frac{2\epsilon}{5} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{then apply (9)} \\
&\quad < \lim_{j\to\infty}\left(\frac{1}{j}\sum_{k=0}^{j-1}\|W^{\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2\right) + \lambda\|W\|^2 + \frac{3\epsilon}{5} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{then apply (6)} \\
&\quad < \lim_{j\to\infty}\left(\frac{1}{j}\sum_{k=0}^{j-1}\|W^{\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^{1-k}(z)) - \mathcal{R}T^{-k}(z)\|^2\right) + \frac{4\epsilon}{5} \\
&\quad \text{Then apply the Ergodic Theorem again} \\
&\quad = \mathbb{E}_{\mu}[\|W^{\top}(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T(\boldsymbol{Z}) - \mathcal{R}(\boldsymbol{Z}))\|^2 | \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}] + \frac{4\epsilon}{5} \\
&\quad = \mathbb{E}_{\mu}[\|H_{W}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} - \Phi H_{W}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}\|^2 | \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}] + \frac{4\epsilon}{5} \\
&\quad \text{then apply (5)} \\
&\quad < \epsilon
\end{aligned}
$$

$\square$

3.1. Connection to Partially Observed Markov Decision Processes

Theorem 3.6 applies to a reinforcement learning scenario where the observations are a stationary and ergodic process. This includes the case where observations emerge from a partially observed, stationary and ergodic Markov decision process. These are themselves a special case of a partially observed Markov decision process (POMDP) which are a common scenario studied in the reinforcement learning community. In particular, the results in this paper apply to POMDPs in the special case that the underlying Markov process is stationary and ergodic. However, there exist stationary ergodic processes, which satisfy the conditions of Theorem 3.6, which are not the output of any partially observed decision Markov process.

The approach that we set out in this paper has a lot in common with POMDPs, but there are some subtle differences which we will clarify here. First of all, the value function in this paper is defined in terms of the complete sequence of (reward, action, observation) triples, rather than the current belief state. One advantage of our approach is that a belief state does not need to be computed explicitly, nor do any assumptions need to made about the relationship between the hidden state of the environment and the observations. In the setting of this paper, the reservoir states x_k (which are explicitly computed by evaluating $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^k(z)$ can be interpreted as latent states, very much like the latent states for POMDPs. We also stress that the value function V and reservoir functionals $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ and $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*_\ell}$ are causal and time invariant (CTI) so we are never using future information that is unavailable in the present, despite the input sequences being bi-infinite. Indeed, one of the strengths of our approach is that the learning procedure will be able to learn the impact of any unobserved or hidden states via the latent states x_k and linear regression.

3.2. Training ESNs with online learning

In some reinforcement learning applications, it is useful - or even essential - for the optimisation of W to occur dynamically as new data comes in; such algorithms are called *online* learning algorithms. In this section, we will present and discuss some preliminary novel results surrounding online learning algorithms that use ESNs. We will first introduce a lemma, stating that, under reasonable conditions, the ODE

$$\frac{d}{dt}W = -h(W) := -\mathbb{E}_\mu\Big[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_W(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_W(\boldsymbol{Z})\big)\Big] \tag{10}$$

converges exponentially quickly to a globally asymptotic fixed point W^*, for which the associated ESN functional $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}$ is *close* to the unique fixed point of Φ. By *close* we mean that the orthogonal projection of $\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}$ onto the finite dimensional vector space of functionals $\{H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_W \mid W \in \mathbb{R}^d\}$ is $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}$. Unlike the previous result (Theorem 3.6) we do not need to assume that the contraction mapping satisfies $\Phi(H) = R + \gamma\mathbb{E}[HT_{\boldsymbol{Z}}]$. We could choose for example $\Phi(H) = R + \gamma \sup_\pi \mathbb{E}[HT_{\boldsymbol{Z}(\pi)}]$ where $\boldsymbol{Z}(\pi)$ is a process under a control π. The fixed point of this operator is the *optimal* value function V^*.

Lemma 3.7. *Let $\boldsymbol{Z}$ be an admissible input process. Let $\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}$ be a $n \times n$, $n \times d$, and $n \times 1$ dimensional random reservoir matrix, input matrix and bias vector. Let $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$ and $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_W$ denote the associated ESN functionals. Let Φ be a contraction mapping, with Lipschitz constant $0 \leq \tau < 1$, on the space of CTI filters $H : (D_N)^{\mathbb{Z}} \to \mathbb{R}$ that are μ-measurable and satisfy $\mathbb{E}[H(\boldsymbol{Z})^2] < \infty$. Suppose further that $0 \leq \tau < \kappa^{-1}$ where κ is the condition number of the autocorrelation matrix*

$$\Sigma = \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\top}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right].$$

Then there exists a $\delta > 0$ such that the ODE

$$\frac{d}{dt}W = -h(W) := -\mathbb{E}_\mu\Big[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big(H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) - \Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big) \;\Big|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\Big]$$

satisfies

$$\frac{d}{dt}\|W - W^*\| \leq -\delta\|W - W^*\| \tag{11}$$

where W^* is a globally asymptotic fixed point. W^* enjoys the further property that

$$H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} = \mathcal{P}\Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$$

where $\mathcal{P}$ denotes the $L^2(\mu)$ orthogonal projection operator on the μ-measurable filters H satisfying $\mathbb{E}[H(\boldsymbol{Z})^2] < \infty$ and is defined

$$\mathcal{P}H(z) := H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\top}(z)\Sigma^{-1}\mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})H(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right].$$

Proof. To show that W^* is a globally asymptotic fixed point it suffices to show that there exists a $\delta > 0$ such that

$$(W - W^*)\cdot(h(W^*) - h(W)) \leq -\delta\|(W - W^*)\|^2$$

as this implies

$$\frac{d}{dt}\|W - W^*\| \leq -\delta\|W - W^*\|.$$

To construct this δ, we first note that

$$h(W) = \Sigma W - \mathbb{E}_\mu\Big[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big) \;\Big|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\Big]$$

so, by a direct computation we have

$$
\begin{aligned}
&(W - W^*) \cdot (h(W^*) - h(W)) \\
&\quad = (W - W^*) \cdot \left(\mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] - \mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \right) \\
&\quad - (W - W^*) \cdot \left(\Sigma W - \Sigma W^* \right) \\
&\quad = (W - W^*) \cdot \left(\mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] - \mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \right) \\
&\quad - (W - W^*)^\top \Sigma \big(W - W^* \big) \\
&\quad \leq (W - W^*) \cdot \left(\mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] - \mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \right) \\
&\quad - \sigma \|W - W^*\|^2 \;\; \text{where } \sigma \text{ is the smallest eigenvalue of } \Sigma \\
&\quad = (W - W^*) \cdot \left(\mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big) - H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) \Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \right) \\
&\quad - \sigma \|W - W^*\|^2 \\
&\quad \leq (W - W^*) \cdot \left(\mathbb{E}_\mu \left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) H_W^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big) - H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z}) H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})) \;\middle|\; \boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta} \right] \right) \tau \\
&\quad - \sigma \|W - W^*\|^2 \;\; \text{because } \tau \text{ is the Lipschitz constant for } \Phi \\
&\quad = \tau (W - W^*)^\top \Sigma (W - W^*) - \sigma \|W - W^*\|^2 \\
&\quad \leq \tau \rho \|W - W^*\|^2 - \sigma \|W - W^*\|^2 \;\; \text{where } \rho \text{ is the largest eigenvalue of } \Sigma \\
&\quad = -(\sigma - \tau\rho) \|W - W^*\|^2,
\end{aligned}
$$

so we can set $\delta := \sigma - \tau\rho$ and notice $\delta > 0$ because $0 \leq \tau < \kappa^{-1} = \sigma/\rho$. Next, to show that

$$H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} = \mathcal{P} \Phi H_{W^*}^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}$$

we observe that since W^* is an equilibrium point of the ODE

$$\dot{W} = -h(W)$$

it follows that $h(W^*) = 0$ and therefore

$$\begin{aligned}
0 &= \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z})\big) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
\implies 0 &= \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\top}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]W^* - \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
\implies 0 &= \Sigma W^* - \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
\text{so,}\;\; \Sigma W^* &= \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
\text{so,}\;\; W^* &= \Sigma^{-1}\mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
\text{so,}\;\; H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*} &= H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\top}\Sigma^{-1}\mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}(\boldsymbol{Z}) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right]\\
&= \mathcal{P}\Phi(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*}).
\end{aligned}$$

□

One rather restrictive condition of this lemma is that the Lipschitz constant τ of the contraction Φ must be less than the reciprocal condition number κ^{-1}. κ is a measure of *how orthonormal* the columns of the autocorrelation matrix Σ are. In particular, if the columns are indeed orthonormal, then $\kappa = 1$ and this condition ceases to be restrictive at all. If the columns are close to being linearly dependant, then κ is large so the requirement that κ^{-1} is small becomes troublesome. If indeed there is a linear dependence, the matrix Σ is not even invertible and the theorem breaks down completely. If we interpret $H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})$ as a vector of features, then κ grows with the correlation between features. Higher correlation between the features imposes a greater constraint on the Lipschitz constant τ. If we have no inter-feature correlation then $\kappa = 1$ and we have no restriction at all on τ.

To actually solve ODE (10) we may need to compute

$$h(W) := \mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}(\boldsymbol{Z})\big) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right] \tag{12}$$

which may, or may not, be practical. For example, if the process $\boldsymbol{Z}$ is ergodic, we can approximate (12) by taking a sufficiently long time average of

$$H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}T^k(z)\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}T^k(z) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W}T^k(z)\big).$$

Alternatively, we may approach the problem of solving (10) by first considering the explicit Euler method (with time-steps $\alpha_k > 0$)

$$\begin{aligned}
W_{k+1} &= W_k - \alpha_k h(W_k)\\
&= W_k - \alpha_k\mathbb{E}_\mu\left[H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(\boldsymbol{Z})\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W_k}(\boldsymbol{Z}) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W_k}(\boldsymbol{Z})\big) \;\middle|\; \boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}\right],
\end{aligned}$$

then we might (heuristically) expect the algorithm

$$W_{k+1} = W_k - \alpha_k H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^k(z)\big(H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W_k} T^k(z) - \Phi H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W_k} T^k(z)\big) \tag{13}$$

to converge to W^*, where α_k are positive definite real numbers that satisfy

$$\sum_{k=1}^{\infty} \alpha_k = \infty \qquad \sum_{k=1}^{\infty} \alpha_k^2 < \infty.$$

We believe this heuristic could be made rigorous under mild assumptions, because algorithm (13) closely resembles the major algorithm extensively studied in [19] and [20] for which similar results hold. Theorems 17 and 2.1.1. appearing in [19] and [20] respectively suggest that an algorithm much like (13) converges almost surely to W^* if its associated ODE (reminiscent of (10)) satisfies condition (11), and the input process $\boldsymbol{Z}$ is strongly mixing. The conjecture that algorithm (13) converges to W^* is also reminiscent of Theorem 3.1 by [21], and related results by [22]. These results are closely related to Q-learning and stochastic gradient descent. We note that (sadly) finding the fixed point of the general contraction mapping Φ renders the estimation of W a nonlinear problem.

The theory yields an online reinforcement learning algorithm which we state below. We envision that the agent chooses a fixed policy π and continues executing the policy for $\ell - 1$ time steps. Under this policy, the agent makes observations z_k and receives rewards r_k. We define $z_k(a)$ as the input to the ESN at time k if the agent had instead executed action $a \in \mathcal{A}$ at time k.

Algorithm 2: Online Learning

1: Choose initial output layer W_0 and reservoir state x_0
2: Randomly generate $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ according to procedure 1
3: **for** each k from 0 to $\ell - 1$
4: Compute $W_{k+1} = W_k - \alpha_k x_k \big(W_k^\top x_k - r_k - \max_a\{W_k^\top \sigma(\boldsymbol{A}x_k + \boldsymbol{C}z_k(a) + \boldsymbol{\zeta})\}\big)$
5: Compute $x_{k+1} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}z_k + \boldsymbol{\zeta})$

---

4. Bee World

To demonstrate the theory presented in section 3, we created a game called *Bee World* and show that a simple reinforcement learning algorithm supported by an ESN can learn to play Bee World with respectable skill. The game is designed so that the theory presented previously is easy to visualise, rather than because the game is hard to master.

Bee World is set on the circle of unit circumference, which we denote by S^1, and represent as an interval with edges identified. At every point y on the circle, there is a non-negative quantity of nectar which may be enjoyed by the bee without depletion. 'Without depletion' means that the bee takes a negligible amount of nectar from the point y, so the bee occupying point y does not cause the amount of nectar at y to change. Furthermore, the nectar at every point y varies with time t according to the prescribed function

$$n(y, t) = 1 + \cos(\omega t)\sin(2\pi y) \tag{14}$$

(which we chose somewhat arbitrarily) that is unknown to the bee. Thus, the amount of nectar enjoyed by the bee at time t is a value that lies in the interval $[0, 2]$, which we will denote $\mathcal{N}$. Time advances in discrete integer steps $t = 0, 1, 2, \ldots$, and at any time point t a bee at point y observes the quantity of nectar $r \in \mathcal{N}$ at point y and nothing else. Having made this observation, the bee may choose to move anywhere in the interval $(y - c, y + c)$ for some fixed $0 < c < 1$ and arrive at its chosen destination at time $t + 1$. The interval of possible moves $(-c, c)$ is called the action space and is denoted $\mathcal{A}$. The goal of the bee is to devise a policy whereby, given all its previous observations, the bee makes a decision as to where to move next, such that the discounted sum over all future nectar is as great as possible. The space of all previous (reward, action) pairs $(\mathcal{N} \times \mathcal{A})^{\mathbb{Z}_-}$ is contained by the space of bi-infinite sequences $(\mathbb{R}^2)^{\mathbb{Z}}$. The agent playing Bee World makes no observations beyond the rewards (nectar) and actions, but we could easily envision a more general game where the agent makes observations from a set Ω and therefore makes its decisions based on a left sequence of (reward, action, observation) triples.

The policy adopted by the bee may be realised as a deterministic policy $\pi : (\mathcal{N} \times \mathcal{A})^{\mathbb{Z}} \to \mathcal{A}$ (a CTI functional) for which the bee executes an action $a \in \mathcal{A}$ determined by the history of (reward, action) pairs. Alternatively, the bee may adopt a stochastic policy, for which every state history of (reward, action) pairs admits a distribution over actions $\mathcal{A}$ from which the bee makes a random choice.

Though the evolution of Bee World is Markovian (and deterministic), the bee makes only a partial observation of the state of Bee World (i.e the amount of nectar the bee observes at time t) so the bee must take advantage of its *memory* to reconstruct the true state and find an optimal policy. This need for memory renders the problem suitable for an ESN, while ruling out the conventional theory of Markov Decision Processes. The problem of playing Bee World can therefore be formulated as a Partially Observed Markov Decision Process.

4.1. Approximating the value functional

Under a policy π, the nectar-action pairs experienced by the bee yield a realisation of the $(\mathcal{N}, \mathcal{A})^{\mathbb{Z}}$-valued random variable $\boldsymbol{Z}$. It therefore makes sense to define the value functional $V : (\mathcal{N} \times \mathcal{A})^{\mathbb{Z}} \to \mathbb{R}$ associated to $\boldsymbol{Z}$ by

$$V(z) = \mathbb{E}_{\mu}\left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R} T^k(\boldsymbol{Z}) \;\middle|\; \boldsymbol{Z}_j = z_j \; \forall j \leq 0 \right] \tag{15}$$

where $\mathcal{R} : (\mathcal{N} \times \mathcal{A})^{\mathbb{Z}} \to \mathbb{R}$ is the reward functional defined by $\mathcal{R}(z) = r_0$, where r_0 is the nectar collected at time 0, T is the shift operator, and $\gamma \in [0, 1)$ is the discount factor representing the relative importance of near and long term nectar consumption. We can see after a simple rearrangement of (15) that

$$V(z) = \mathcal{R}(z) + \gamma \mathbb{E}_{\mu}[V T_{\boldsymbol{Z}}(z)]$$

so V is the unique fixed point of the contraction mapping Φ defined by

$$\Phi(H)(z) := \mathcal{R}(z) + \gamma \mathbb{E}_\mu[HT_{\boldsymbol{Z}}(z)]$$

as discussed in Section 3. Thus, by Theorem 3.6, we can approximate the value function V using an ESN trained by regularised least squares as long as the nectar-action pairs $z \in (\mathcal{N} \times \mathcal{A})^{\mathbb{Z}}$ are drawn from a suitable ergodic process $\boldsymbol{Z}$. Therefore, we chose an initial policy π_0 such that $\boldsymbol{Z}$ is ergodic. In particular, we chose a stochastic policy $\pi_0(z) \sim U(-c, c)$ for all histories of (reward, action) pairs $z \in (\mathcal{N} \times \mathcal{A})^{\mathbb{Z}}$ so that the bee takes a uniform sample from the action space $\mathcal{A} = (-c, c)$ at any point $y \in S^1$. For the purpose of playing a game, we set $c = 0.1$ and $\gamma = 0.5$. We allowed the bee to execute this policy for 2000 time steps and recorded the observed nectar at every time. The first 250 time steps are plotted in Figure 1.

Next, we set up an ESN of dimension $n = 300$, with reservoir matrix, input matrix, and bias $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ populated with i.i.d uniform random variables $U(-0.05, 0.05)$. $\boldsymbol{A}$ was then multiplied by a scaling factor such that the 2-norm of $\boldsymbol{A}$ satisfies $\|\boldsymbol{A}\|_2 = 1$. We choose an activation function $\sigma(x) := \max(0, x)$. We should pause here and note that ESN described here differs slightly from the ESN described in procedure 1. We instead generated $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ in a traditional way, which is empirically observed to be highly successful, as demonstrated in the literature, rather than the more cumbersome method described in procedure 1. These numerical results suggest that procedure 1 can be simplified.

We then computed a sequence of reservoir states $x_k \in \mathbb{R}^{300}$ for the ESN using the iteration

$$x_{k+1} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}z_k + \boldsymbol{\zeta})$$

where $x_0 = 0$ and each $z_k \in (\mathcal{N} \times \mathcal{A})$ comprises 2 components: the first is the quantity of nectar observed by the bee at time k, and the second is the action $a \in (-c, c)$ executed at time k under policy π_0. Now we return our attention to Theorem 3.6, and see that the W_ℓ^* minimising (over W)

$$\frac{1}{\ell} \sum_{k=0}^{\ell-1} \|W^\top (H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{-k}(z) - \gamma H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}} T^{1-k}(z)) - \mathcal{R}(z)\|^2 + \lambda \|W\|^2$$

converges to W minimising

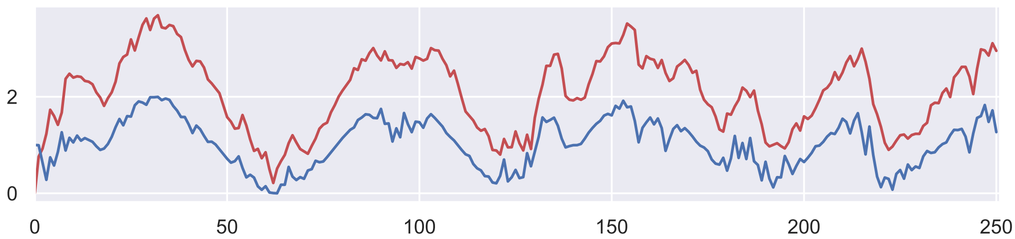
$$\|W^\top (x_k - \gamma x_{k+1}) - r_k\|^2 + \lambda \|W\|^2 \tag{16}$$

so we can immediately reformulate (16) as the least squares problem

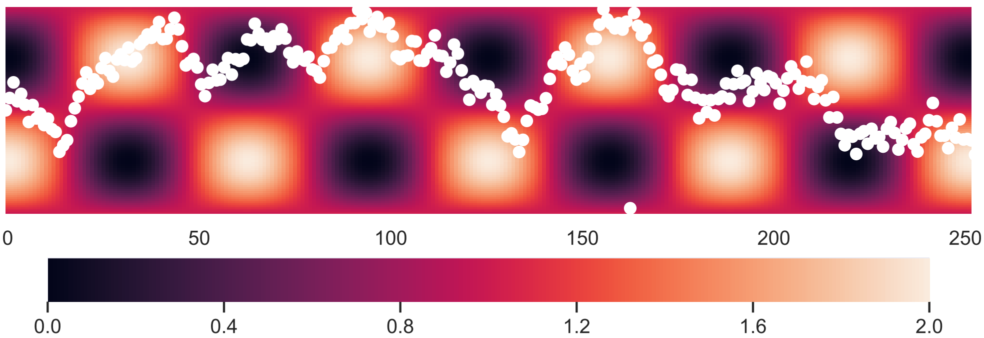
$$W = (\Xi^\top \Xi + \lambda I)^{-1} \Xi^\top U$$

where Ξ is the matrix with kth column

$$\Xi_k := x_k - \gamma x_{k+1}$$



(a) The nectar collected (blue) and the approximate value function under the initial policy π_0 (red) is plotted for the first 250 time steps (x-axis).



(b) The nectar function $n(y,t)$ (14) at every point represented as a heat map in the (t,y) plane, with the position of the bee at time t under the initial policy indicated by the overlaid white circles.

Figure 1: Dynamics of Bee World where the bee executes the initial policy $\pi_0(z) \sim U(-0.1, 0.1)$ for the first 250 time steps.

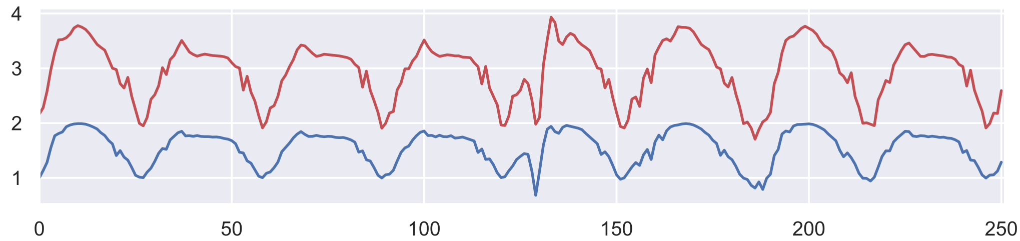
and U has kth entry r_k the kth quantity of nectar, and λ is the regularisation parameter which we set to 10^{-9}. We solved this linear system using the SVD. Now

$$V(z) \approx H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}_{W^*_\ell}(z) \equiv (W^*_\ell)^\top H^{\boldsymbol{A},\boldsymbol{C},\boldsymbol{\zeta}}(z) \equiv W^\top x$$

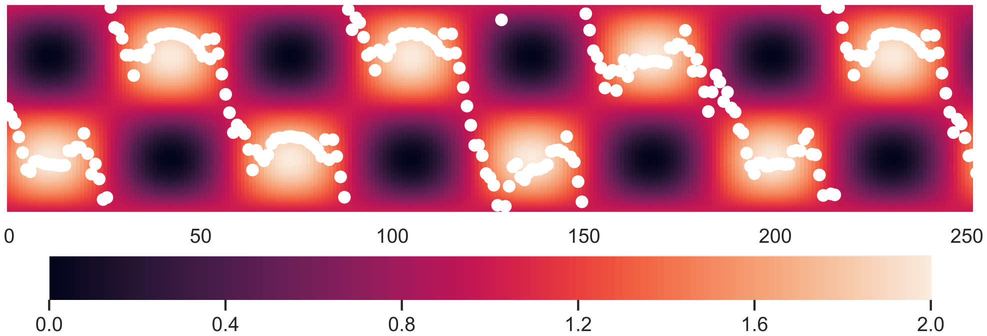
where x is the reservoir state associated to the left infinite input sequence z. Furthermore, the map $(W^\top \cdot)$ therefore approximates the unique fixed point of Φ (by Theorem 3.6) and this fixed point is exactly the value functional we are looking for. Thus, we can easily compute the approximate value of an arbitrary reservoir state x under the initial policy π by computing the inner product $W^\top x$. We illustrate this in Figure 1a by plotting, at each time $k = 1, \ldots, 250$, the value of every observed state to accompany the observed nectar.

4.2. Updating the policy

Having computed an approximate value function under the initial policy $\pi_0(z) \sim U(-0.1, 0.1)$, we were faced with the problem of how to improve upon this policy. Exploring efficient and effective algorithms for iteratively improving a policy is a rich area of reinforcement learning research, but outside the scope of this section. Instead, we implemented a simple and greedy approach. For a given reservoir state x we consider 100 actions $a_1, a_2, \ldots a_{100}$ uniformly sampled over $\mathcal{A} = (-0.1, 0.1)$, then for each action we consider the nectar-action



(a) The nectar collected (blue) and the approximate value function (red) is plotted for the first 250 time steps (y-axis) under the improved policy π_1.



(b) The nectar function $n(y,t)$ (14) at every point represented as a heat map in the (t,y) plane, with the position of the bee at time t under the improved policy is indicated by the overlaid white circles.

Figure 2: Dynamics of Bee World where the bee executes the improved policy π_1 for the first 250 time steps.

pairs $z^{(1)},\dots,z^{(100)} \in \mathcal{N} \times \mathcal{A}$ where the nectar for each pair is the current nectar; and is therefore the same in every pair. Then we compute the next reservoir states for each pair

$$x_{k+1}^{(i)} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}z_k^{(i)} + \boldsymbol{\zeta})$$

and estimate the value of executing the ith action by computing $W^\top x_{k+1}^{(i)}$. Then we choose to execute the action a^* with the greatest estimated value - which determines our new policy π_1 - which yields a significant improvement over the initial policy π_0, as illustrated in Figure 2. Under the initial policy π_0 the bee collected an average of approximately 1.05 nectar per unit time, in comparison to 1.52 nectar under the improved policy π_1. This is much closer to the optimal value of approximately 1.60, which we obtain in the next section. The algorithm which first approximates the value function, and then updates the policy is described in Algorithm 3.

4.3. An Analytic Solution for Bee World

In this section, we will analyse Bee World so that we can compare the ESN solution to results that we can prove. To make our own lives easier, we consider a smooth version of Bee World, rather than the discrete time version solved by the ESN, so that we can formulate Bee World as a control problem that admits a solution via the Euler-Lagrange equation. We

Algorithm 3: One Step Offline Learning Algorithm (Bee World)

Choose initial reservoir state x_0
Randomly generate $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$
for each k from 0 to $\ell - 1$
    Compute $x_{k+1} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}(r_k, a_k) + \boldsymbol{\zeta})$
Find W that minimises $\sum_{k=0}^{\ell-1} \|W^\top(x_k - \gamma x_{k+1}) - r_k\|^2 + \lambda \|W\|^2$
for each k from ℓ to $L - 1$
    Compute $a^* = \max_a \{W^\top \sigma(\boldsymbol{A}x_k + \boldsymbol{C}(r_k, a) + \boldsymbol{\zeta})\}$
    Compute $x_{k+1} = \sigma(\boldsymbol{A}x_k + \boldsymbol{C}(r_k, a^*) + \boldsymbol{\zeta})$

have the control system

$$\begin{aligned} \dot{\tau} &= 1 \\ \dot{y} &= u(y, \tau) \end{aligned}$$

where u is the controller dependant on y and τ. Then we have a cost function

$$\mathcal{C}(x, \tau, u) = f(u) - n(y, \tau)$$

where $f(x)$ is the penalty term for using the control u and $n(y, \tau)$ is the nectar function. In the above formulation of Bee World

$$f(u) = \begin{cases} 0 & \text{if } -c \leq u \leq c \\ \infty & \text{otherwise} \end{cases}$$

where $c = 0.1$. Then the objective is to find

$$u^* = \arg\min_u \left\{ \int_0^\infty \gamma^t \mathcal{C}(y, \tau, u) \; dt \right\}.$$

We can see that f is not a well defined function so we will introduce the family of functions

$$f_\epsilon(u) = -\epsilon \log(\cos(\pi u / (2c)))$$

where $\epsilon > 0$, and notice that f_ϵ approaches f pointwise as $\epsilon \to 0$. Next, we recall that the stationary points (including the minimum) of the integral functional

$$\mathcal{I}[y] = \int_0^\infty \mathcal{F}(t, y, \dot{y}) \; dt$$

all satisfy the Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial \mathcal{F}}{\partial \dot{y}} - \frac{\partial \mathcal{F}}{\partial y} = 0.$$

So, we let

$$\begin{aligned}\mathcal{F}(t,y,\dot{y}) &= \gamma^t \mathcal{C}(t,y,\dot{y})\\ &= \gamma^t(-\epsilon\log(\cos(\pi\dot{y}/(2c))) - \cos(\omega t)\sin(2\pi y) - 1)\end{aligned}$$

then

$$\begin{aligned}0 &= \frac{d}{dt}\frac{\partial\mathcal{F}}{\partial\dot{y}} - \frac{\partial\mathcal{F}}{\partial y}\\ &= \frac{d}{dt}\left(\gamma^t\frac{d}{d\dot{y}}(-\epsilon\log(\cos(\pi\dot{y}/(2c))))\right) + 2\pi\gamma^t\cos(\omega t)\cos(2\pi y)\\ &= \frac{\pi\epsilon}{2c}\frac{d}{dt}\left(\gamma^t\tan(\pi\dot{y}/(2c))\right) + 2\pi\gamma^t\cos(\omega t)\cos(2\pi y)\\ &= \frac{\pi\epsilon}{(2c)}\left(\log(\gamma)\gamma^t\tan(\pi\dot{y}/(2c)) + \gamma^t\frac{\pi\ddot{y}}{2c}\sec^2(\pi\dot{y}/(2c))\right) + 2\pi\gamma^t\cos(\omega t)\cos(2\pi y)\\ &= \frac{\pi\epsilon}{2c}\left(\log(\gamma)\tan(\pi\dot{y}/(2c)) + \frac{\pi\ddot{y}}{2c}\sec^2(\pi\dot{y}/(2c))\right) + 2\pi\cos(\omega t)\cos(2\pi y),\end{aligned}$$

which we can reformulate as a dynamical system

$$\begin{aligned}\dot{v} &= -\frac{2c\cos^2(\pi v/(2c))}{\pi}\left(\frac{4c\cos(\omega\tau)\cos(2\pi y)}{\epsilon} + \log(\gamma)\tan(\pi v/(2c))\right)\\ \dot{y} &= v\\ \dot{\tau} &= 1\end{aligned}\tag{17}$$

whose solutions are stationary points of the integral functional. For small ϵ, we approach the Bee World problem. We took $\epsilon = 10^{-5}$, $\gamma = 1/2$, initial position $y = 0$, and initial velocity $v = 0$ then simulated a trajectory of the ODE using `scipy.integrate.odeint`. We plotted this in Figure 3. The average nectar collected by under this policy was approximately 1.60.

5. Application to Stochastic Control

ESNs have shown remarkable promise in solving problems in mathematical finance - including by [23], [24], and [25] who used an ESN to predict the future values of stock prices. [26] used an ESN to learn the solution to a credit rating problem and [27] used an ESN to forecast exchange rates, comparing the results to forecasts made with an ARMA model. In this section we will introduce a stochastic optimal control problem arising in the market making problem. We will solve this problem analytically, and compare this to the solution obtained by a reinforcement learning agent supported by an ESN.

5.1. A Market Making Problem

We consider a stochastic control problem inspired by the motivations of a market maker acting in a general financial market. In practice the specific role of a market maker depends

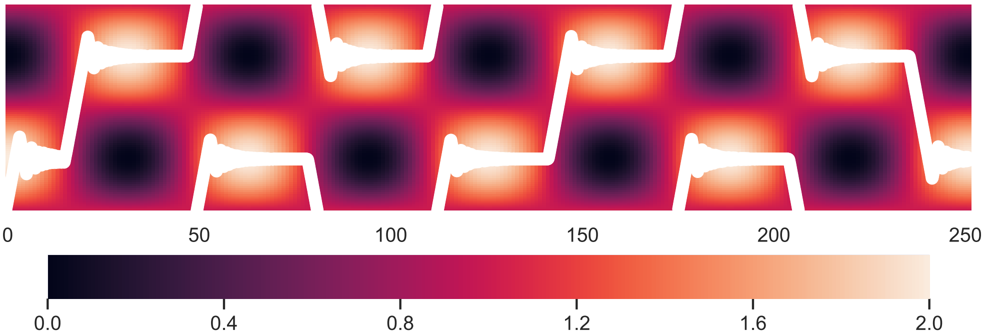


Figure 3: A numerical solution to the ODE (17) with $\epsilon = 10^{-5}$ (white line) superposed on the heat map of the nectar function $n(y, t)$ given in (14). Dark colours indicate regions of low nectar, light regions indicate high values of the nectar function. We observe that the solution trajectory spends much more time near local maxima of the nectar function but has complicated oscillatory fluctuations during transitions between local maxima. The oscillations are likely due to approaching a sort of singularity as $\epsilon \to 0$.

on the particular market, but we consider a market maker who provides liquidity to other market participants by quoting prices at which they are willing to sell (ask) and buy (bid) an asset. By setting the ask price higher than the bid price in general they can profit from the difference when they receive both a buy and sell order at these prices. However, the market maker faces risk, since if they buy a quantity of the asset the market price might move against them before they are able to find a seller.

The market making problem is a complex one, and has been studied extensively since the publication of the paper by [28]. The paper of [29] gives a good overview of much of this work. We consider a stylised version of this problem that focuses on inventory management without considering explicit optimal quoting strategies. We consider that a market maker acting relatively passively around the market price in ordinary conditions would expect to observe a random demand for buy and sell orders. If as a result of random fluctuations they find their inventory has drifted away from zero, they would set prices more competitively on either the ask or bid side to encourage trades to balance their position. Very broadly the conclusions of work on the market making problem are that there is a price to be paid to exert control over the inventory process and bring inventories closer to zero.

Motivated by this insight, we consider the market maker's inventory to be a stochastic process $(\boldsymbol{Y}_t)_{t\geq 0}$ with dynamics

$$d\boldsymbol{Y}_t = \boldsymbol{\pi}_t dt + \sigma d\boldsymbol{W}_t$$

where $(\boldsymbol{W}_t)_{t\geq 0}$ is a standard Brownian motion.

The parameter σ measures the volatility of the incoming order flow, and $(\boldsymbol{\pi}_t)_{t\geq 0}$ is the control process by which the market maker adds drift into their order flow by moving their bid and ask quotes. Naturally, there is a cost involved in applying the control, and a further cost to holding inventory away from zero. We introduce parameters α and β to quantify these effects and model the market maker's profit as a stochastic process solving

$$d\boldsymbol{Z}_t = (r - \alpha\boldsymbol{\pi}_t^2 - \beta\boldsymbol{Y}_t^2)dt$$

where r is the rate of profit the market maker would achieve from the bid-ask spread if they did not have concerns about the asset price movements. We consider the case where the market maker seeks to maximise their long run discounted profit

$$v(y) = \max_{\pi} \mathbb{E}^y\Big[\int_0^\infty e^{-\delta t} d\boldsymbol{Z}_t\Big],$$

where $\mathbb{E}^y$ is the expectation with the process started at $Y_0 = y$. We can show that the market maker's value function and optimal control are

$$v(y) = -\alpha h y^2 + \frac{r - \alpha h \sigma^2}{\delta}, \qquad \pi^*(y) = -hy, \tag{18}$$

where

$$h := \frac{-\alpha\delta + \sqrt{\alpha^2\delta^2 + 4\beta}}{2\alpha}$$

Further, the inventory process $\boldsymbol{Y}_{t\geq 0}$, when controlled by the optimal control $\pi^*(y) = -hy$ is given by the Ornstein-Uhlenbeck process

$$d\boldsymbol{Y}_t = -h\boldsymbol{Y}_t dt + \sigma d\boldsymbol{W}_t$$

whose stationary distribution is a Gaussian $\mathcal{N}\left(0, \frac{\sigma^2}{2h}\right)$.

We observe that this is an infinite horizon, Linear-Quadratic regulator (LQR) type problem, a class of problems which have a long history in the control literature, and more recently have been systematically studied in the reinforcement learning literature. Recent work on online learning for the LQR problem (e.g. [30, 31, 32]) has considered a range of variants of the LQR problem, including cases with uncertainty on the both the dynamics and the reward, and where the state variable may only be partially observed. However most of these approaches work in the setting of model-based learning approaches: that is, they attempt to learn a "model" of the world, and therefore exploit the fact that the LQR structure is known and can be learned from the data; in comparison, [30] still rely on the LQR structure, but do not directly try to learn the "model" of the world. The paper [33] analyses the difference between model-based and model-free approaches to the LQR problem, showing that one should expect an exponential separation between model-based and model-free approaches. In this context, our approach, which does not assume the LQR structure, can also be compared to model-free approaches, such as the classical work of [34], which takes a Q-learning approach.

5.2. Discretised problem

To turn this into a problem into one that can be used to train an Echo State Network we reformulate it in discrete time; we consider a process $\boldsymbol{Y}_0, \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ such that

$$\boldsymbol{Y}_{k+1} - \boldsymbol{Y}_k = \epsilon\boldsymbol{\pi}_k + \sigma\sqrt{\epsilon}\mathcal{N}_k$$

where $(\mathcal{N}_k)_{k\in\mathbb{N}}$ are a sequence of i.i.d. random variables $\mathcal{N}_k \sim \mathcal{N}(0,1)$ for each $k \in \mathbb{N}$, and $\epsilon > 0$ is the time increment. The control is now a sequence $\pi = (\boldsymbol{\pi}_k)_{k\in\mathbb{N}}$. The profit function satisfies $\boldsymbol{Z}_0 = 0$ and

$$d\boldsymbol{Z}_k := \boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k = \epsilon(r - \alpha\boldsymbol{\pi}_k^2 - \beta\boldsymbol{Y}_k^2).$$

and the market maker seeks to maximise the value function

$$v(y) = \max_{\pi} \mathbb{E}^y\Big[\sum_{k=0}^{\infty} e^{-\delta\epsilon k} d\boldsymbol{Z}_k\Big],$$

over choices of the control π where $\mathbb{E}^y$ is the expectation with the process started at $\boldsymbol{Y}_0 = y$.

It can be shown that in the limit as $\epsilon \to 0$, the optimal control and value function for this problem converge precisely to the optimal control and value function in the continuous case.

We state here the results in the case $\epsilon = 1$, the value we will use for the application of the Echo State Network below. Writing $\gamma = e^{-\delta}$, we find in this case that the value function and optimal control are given by

$$v(y) = -\alpha p y^2 + \frac{r - \gamma\alpha p\sigma^2}{1-\gamma}, \qquad \pi^* = -py$$

where

$$p := \frac{(\alpha(\gamma-1)+\gamma\beta) + \sqrt{(\alpha(\gamma-1)+\gamma\beta)^2 + 4\alpha\beta\gamma}}{2\gamma\alpha}$$

.

The process $\boldsymbol{Y}$ controlled by π^* is Markovian, and has transition operator

$$\begin{aligned}(\mathcal{T}s)(y) &= \int_{-\infty}^{\infty} \mathbb{P}(\boldsymbol{Y}_{k+1} = y \mid \boldsymbol{Y}_k = x) s(x)\, dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(y-(1-px))^2}{2\sigma^2}} s(x)\, dx.\end{aligned}$$

It is straightforward to verify that the Gaussian probability density function

$$s^*(y) = \frac{\sqrt{p(2-p)}}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{y^2 p(2-p)}{2\sigma^2}}, \tag{19}$$

is a fixed point of $\mathcal{T}$ and hence that the controlled process has stationary distribution $\mathcal{N}\left(0, \frac{\sigma^2}{p(2-p)}\right)$.

5.3. Solving the Market Making Problem with an ESN

In this section, we seek to solve the the market making problem with a reinforcement learning algorithm supported by an ESN. In this set up, we assume the market maker has no knowledge of the cost function, and no knowledge of the effect of executing an action. The agent must execute a variety of actions in a variety of states to learn about the environment and the effect of its actions. Then, the market maker makes reasonable changes to its policy to arrive at a policy that reduces the long term costs of operation. The policy obtained by the reinforcement learning approach is compared to the optimal policy derived with full knowledge of the system.

5.3.1. Approximating the value functional

For the purpose of running the simulation, we let the cost of operating the control $\alpha = 1$, the cost of straying from the origin $\beta = 1$, the timestep $\epsilon = 1$, and the volatility parameter $\sigma = 1$. We take the baseline profit parameter $r = 0$. The inventory held, and action taken, by the market maker at time k will be denoted y_k and a_k respectively. A sequence of (inventory, action) pairs will be denoted $z \in (\mathbb{R}^2)^{\mathbb{Z}}$ with $z_k = (y_k, a_k)$. The value functional for the market maker problem is defined

$$V(z) = \mathbb{E}_\mu \left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R} T^k(\boldsymbol{Z}) \;\middle|\; \boldsymbol{Z}_j = z_j \;\forall j \leq 0 \right]$$

where $\mathcal{R} : (\mathbb{R}^2)^{\mathbb{Z}} \to \mathbb{R}$ is the reward functional

$$\mathcal{R}(z) = -(\alpha a_{-1}^2 + \beta y_0^2),$$

T is the shift operator, and $\gamma \in [0, 1)$ is the discount factor representing the relative importance of near and long term costs. We can see after a simple rearrangement that

$$V(z) = \mathcal{R}(z) + \gamma \mathbb{E}_\mu [V T_{\boldsymbol{Z}}(z)]$$

so V is the unique fixed point of the contraction mapping Φ defined by

$$\Phi(H)(z) = \mathcal{R}(z) + \gamma \mathbb{E}_\mu [H T_{\boldsymbol{Z}}(z)]$$

as discussed in Section 3. Thus, by Theorem 3.6, we can approximate the value function V using an ESN trained by regularised least squares *if* the (inventory, action) pairs (y_k, a_k) are the realisation of a stationary ergodic process. Consequently, we sought an initial policy π_0 such that the process $\boldsymbol{Z}$ comprising the inventory-action pairs under policy π_0 is stationary and ergodic. In particular, we chose

$$\pi_0(y) \sim \mathcal{N}(0, \sigma_i^2) - \eta y \tag{20}$$

with $\eta = 0.05$ a constant representing the rate of exponential drift toward 0 and $\sigma_i^2 = 1$. We ran this policy for 10000 time steps, and recorded the pairs z_k along with the rewards r_k. Next, we set up an ESN of dimension $n = 300$, with reservoir matrix, input matrix, and bias $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\zeta}$ populated with i.i.d uniform random variables $U(-0.05, 0.05)$. $\boldsymbol{A}$ was then multiplied by a scaling factor such that the 2-norm of $\boldsymbol{A}$ satisfies $\|\boldsymbol{A}\|_2 = 1$. As in the previous example we chose σ to be the ReLU activation function. We then computed reservoir states

$$x_{k+1} = \sigma(\boldsymbol{A} x_k + \boldsymbol{C} z_k + \boldsymbol{\zeta})$$

starting with an initial reservoir state $x_0 = 0$. An arbitrary reservoir state x then encodes the left infinite sequence of (inventory,action) pairs z. We seek an expression for the value of the reservoir state x by solving the least squares problem

$$W = (\Xi^\top \Xi + \lambda I)^{-1} \Xi^\top U$$

(using the singular value decomposition) where Ξ is the matrix with kth column is

$$\Xi_k := x_k - \gamma x_{k+1}$$

and U is the vector of observations where the kth entry is the reward r_k, and λ is the regularisation parameter which we set to `1e-6`. We also chose $\gamma = e^{-1}$. In practice, the discount factor is usually much larger. With this, we obtain an expression for value of the reservoir state x given by $W^\top x$. The results of this policy are shown in Figures 4 and 5. The procedure which estimates the value function and improves upon the policy is described in Algorithm 4.

5.3.2. Updating the policy

We sought to create a new and improved policy based on the observations of under the initial policy using a naïve approach. At each time step, we consider 100 trial actions $a^{(1)}, a^{(2)}, \ldots, a^{(100)}$ drawn from the standard normal distribution $\mathcal{N}(0,1)$ and compute

$$x_{k+1}^{(i)} = \sigma(\boldsymbol{A} x_k + \boldsymbol{C} z_k^{(i)} + \boldsymbol{\zeta})$$

where $z_k^{(i)}$ is the (inventory, action) pair $(y_k, a^{(i)})$, and $a^{(i)}$ is trial action. For each i, we compute $W^\top x_{k+1}^{(i)}$ to obtain the predicted value of executing action $a^{(i)}$. We then choose to execute the action a^* with the greatest predicted value, and update the reservoir state using this (inventory, action) pair (y_k, a^*). This defines our new policy. We ran this new policy for 10,000 time steps and illustrated the results in Figures 6a, and 6b.

Algorithm 4: One Step Offline Learning Algorithm (Market Making)

```
Choose initial reservoir state x_0
Randomly generate A, C, ζ
for each k from 0 to ℓ − 1
    Compute x_{k+1} = σ(A x_k + C(y_k, a_k) + ζ)
Find W that minimises Σ_{k=0}^{ℓ−1} ||W^⊤(x_k − γ x_{k+1}) − r_k||^2 + λ||W||^2
for each k from ℓ to L − 1
    Compute a* = max_a {W^⊤ σ(A x_k + C(y_k, a) + ζ)}
    Compute x_{k+1} = σ(A x_k + C(y_k, a*) + ζ)
```

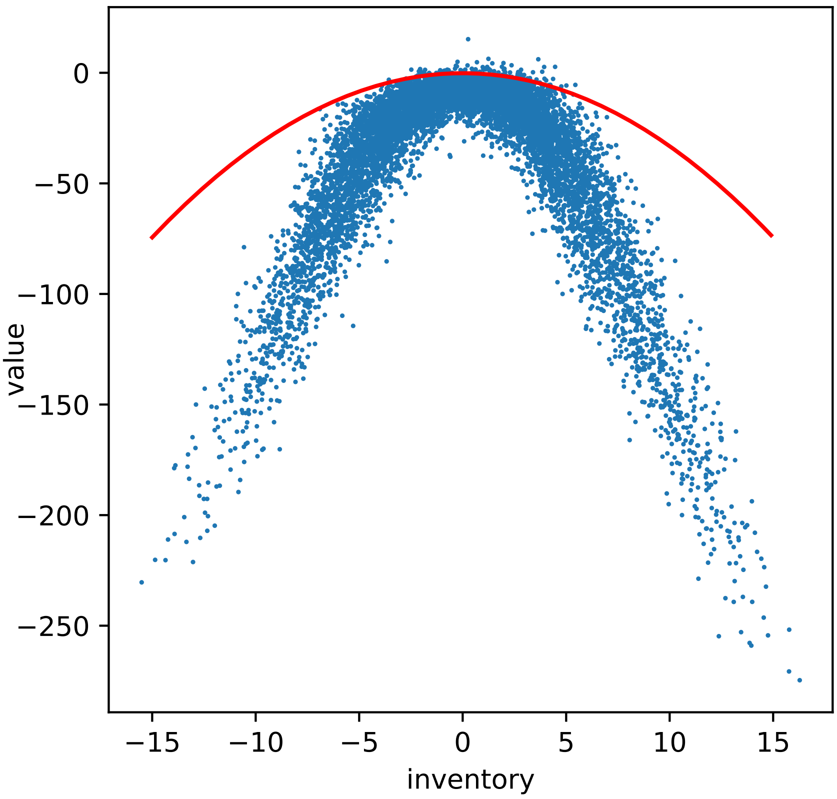
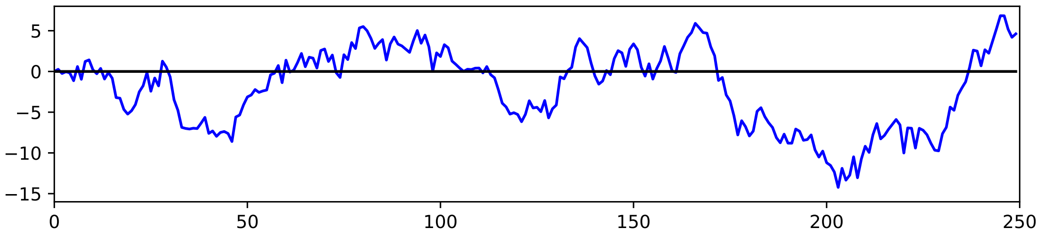
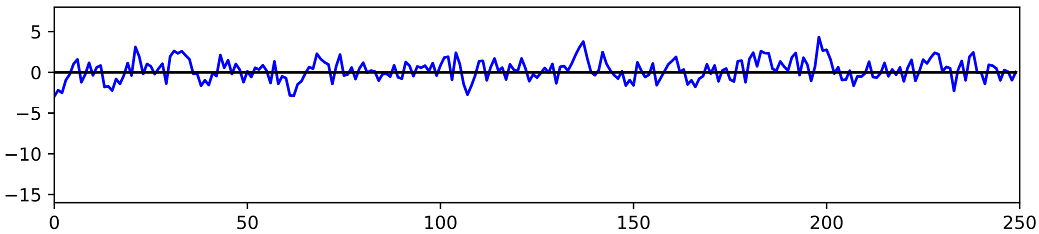


Figure 4: Under the initial policy, the value $V(\boldsymbol{Y})$ (y-axis) learned by the ESN at the inventory $\boldsymbol{Y}$ (x-axis) at each of the 10000 timesteps is shown. The parabolic shape is consistent with the analytically derived optimal value function (19) shown in red. We note that the value function under the initial policy π_0 is not expected to match the value function under the optimal policy π^*.



(a)



(b)

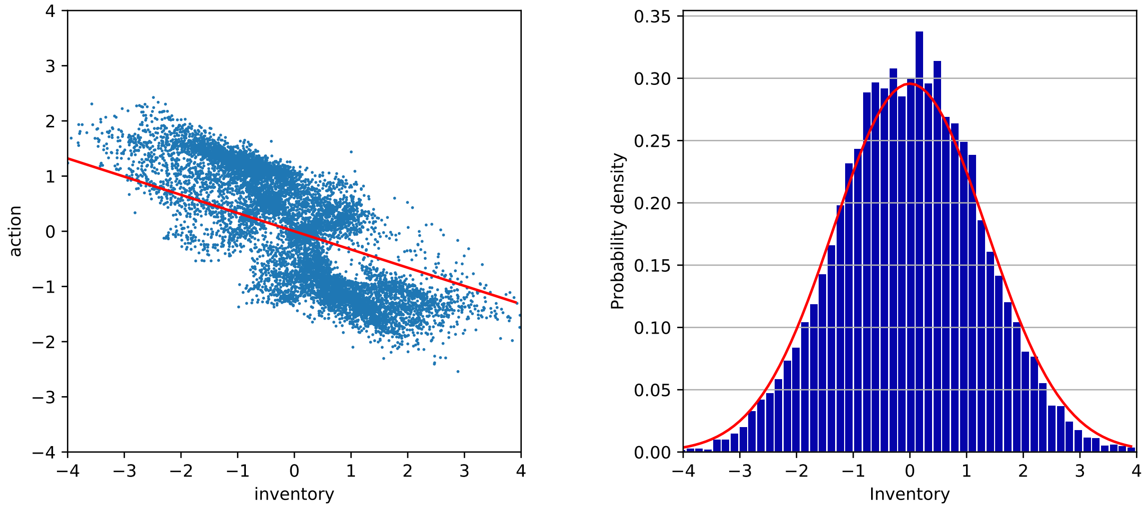
Figure 5: Dynamics of the market maker over time executing (a) the initial policy π_0 and (b) the improved policy π_1. For each plot, the inventory (y-axis) is shown evolving with time (x-axis).

5.4. Comparison between the analytic and learned solutions

The one step reinforcement learning algorithm did not perfectly replicate the analytically derived optimal control, but has moved in a promising direction. We can see in Figure 6a that the inventory process under the improved policy produces (inventory, action) pairs that have some scatter relative to the optimal policy indicated by the red straight line. This suggests that the market maker trained by reinforcement learning is behaving well in some average sense, despite performing many sub-optimal actions. It also appears that the the reinforcement learning algorithm uses the control more aggressively than is optimal. This sub-optimal control results in greater costs than the optimal control. In particular the average cost incurred under the improved policy π_1 is 2.65, while the average cost under the the analytically derived optimal policy is $\sigma/\sqrt{p(2-p)} = 1.35$.

Despite these sub-optimal moves, it seems that the inventory process learned by the market maker has an invariant measure that closely matches the optimal invariant measure. It is reassuring to see that an invariant measure appears, at least numerically, to exist, because the controlled process is assumed to be stationary and ergodic (and therefore admits an invariant measure) in Theorem 3.6.

It is also worth noting that the inventory process, controlled either by the ESN or the optimal control, has support on $\mathbb{R}$, which is not a compact space. Therefore, the conditions of Theorem 3.6 don't technically hold. However, the numerical results here suggest that the ESN has learned the value functional adequately well, suggesting that Theorem 3.6 may hold under relaxed conditions. Of course, realisations of the stochastic processes always explore only bounded subsets of $\mathbb{R}$.



(a)

(b)

Figure 6: (a) Illustrates the (inventory, action) pairs (y_k, a_k) under the improved policy π_1 are represented as points on the scatter plot. The inventory is on the x-axis, and action is on the y-axis. The red line represents the analytically derived optimal control (equation (18)). (b) Illustrates the invariant measure of the inventory process under the improved policy π_1 is approximated with a histogram. The histogram is compared to the analytically derived invariant measure of the optimal control process $\mathcal{N}(0, 1.82)$ (equation (19)).

6. Conclusions and future work

In this paper we have presented three novel mathematical results concerning Echo State Networks trained on data drawn from a stationary ergodic process. The first applies to offline supervised learning. The theorem states that, given a target function, enough training data and a large enough ESN, the least squares training procedure will yield an arbitrarily good approximation to the target function. The second result applies to an agent performing a stochastic policy π. After the agent has collected enough training data, and given a sufficiently large ESN, the least squares training procedure will yield an arbitrarily good approximation to the value function associated to the policy π. The third result is relevant to online reinforcement learning. Though the result is quite preliminary, the lemma is introduced with the intention of developing online algorithms (inspired by Q-learning) to learn the optimal policy for non-Markovian problems.

We demonstrated the second result (which generalises the first) on a deterministic control problem (Bee World) and a stochastic control problem (the market making problem). We chose these 'toy model' problems to understand the performance of the algorithm completely in cases that are solvable analytically, although these optimal solutions themselves are not entirely trivial. The reinforcement learning algorithm we use to improve the policy in both Bee World and the market making problem is extremely simple. It is essentially one iteration of an ϵ-greedy policy [35], with ϵ set to 0. Despite the simplicity of the algorithm, the single iteration considerably improved the policy, resulting in a reasonable approximation to the optimal policy.

It therefore seems a natural direction of future work to develop more sophisticated learning algorithms. Notably the linear upper confidence bound (linUCB) algorithm [35] has a linear structure that fits cleanly into the the linear training framework of the ESN. As this work develops, it will become essential to have a rigorous framework describing the relationship between filters, functionals, random processes and reinforcement learning. The theory presented in this paper tentatively connects these objects using ideas from Markov Decision Processes, but the theory is far from complete.

Acknowledgements

Allen Hart and Kevin Olding are funded through the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), grant number EP/L015684/1.

We thank Jeremy Worsfold for insights about reinforcement learning and the linUCB algorithm, and for refactoring the Bee World code. We also thank Adam White for useful discussions about reinforcement learning settings.

References

[1] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, GMD-Report 148, German National Research Institute for Computer Science (01 2001).

[2] W. Maass, T. Natschläger, H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Computation 14 (11) (2002) 2531–2560. doi:10.1162/089976602760407955.

[3] M. Schlegel, A. Jacobsen, M. Zaheer, A. Patterson, A. White, M. White, General value function networks, arXiv:1807.06763 (2018).

[4] L. Gonon, L. Grigoryeva, J.-P. Ortega, Approximation bounds for random neural networks and reservoir systems, arXiv:2002.05933 (2020).

[5] M. Lukoševičius, H. Jaeger, Reservoir computing approaches to recurrent neural network training, Computer Science Review 3 (3) (2009) 127 – 149. doi:https://doi.org/10.1016/j.cosrev.2009.03.005.

[6] M. Lukoševičius, H. Jaeger, B. Schrauwen, Reservoir computing trends, Künstliche Intelligenz. 26 (4) (2012) 365–371.

[7] A. Rodan, P. Tino, Minimum complexity echo state network, IEEE transactions on neural networks 22 (1) (2011) 131–144.

[8] F. Triefenbach, A. Jalalvand, B. Schrauwen, J.-P. Martens, Phoneme recognition with large hierarchical reservoirs, in: J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, A. Culotta (Eds.), Advances in Neural Information Processing Systems 23, Curran Associates, Inc., 2010, pp. 2307–2315.

[9] H. Jaeger, H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science 304 (5667) (2004) 78–80. arXiv:https://science.sciencemag.org/content/304/5667/78.full.pdf, doi:10.1126/science.1091277.

[10] I. Szita, V. Gyenes, A. Lőrincz, Reinforcement learning with echo state networks, in: S. D. Kollias, A. Stafylopatis, W. Duch, E. Oja (Eds.), Artificial Neural Networks – ICANN 2006, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 830–839.

[11] S. Peitz, K. Bieker, On the universal transformation of data-driven models to control systems, arXiv:2102.04722 (2021).

[12] A. Saxe, P. W. Koh, Z. Chen, M. Bhand, B. Suresh, A. Ng, On random weights and unsupervised feature learning, In Proceedings of the 28th International Conference on Machine Learning (2011).

[13] L. Grigoryeva, J.-P. Ortega, Echo state networks are universal, Neural Networks 108 (2018) 495 – 508. doi:https://doi.org/10.1016/j.neunet.2018.08.025.

[14] L. Grigoryeva, J.-P. Ortega, Differentiable reservoir computing, Journal of Machine Learning Research 20 (179) (2019) 1–62.
URL http://jmlr.org/papers/v20/19-150.html

[15] K. McGoff, S. Mukherjee, N. Pillai, Statistical inference for dynamical systems: A review, Statist. Surv. 9 (2015) 209–252. doi:10.1214/15-SS111.
URL https://doi.org/10.1214/15-SS111

[16] A. Khaleghi, D. Ryabko, Clustering piecewise stationary processes, in: 2020 IEEE International Symposium on Information Theory (ISIT), 2020, pp. 2753–2758. doi:10.1109/ISIT44484.2020.9174045.

[17] A. Hart, J. Hook, J. Dawes, Embedding and approximation theorems for echo state networks, Neural Networks 128 (2020) 234 – 247. doi:https://doi.org/10.1016/j.neunet.2020.05.013.
URL http://www.sciencedirect.com/science/article/pii/S0893608020301830

[18] A. G. Hart, J. L. Hook, J. H. Dawes, Echo state networks trained by tikhonov least squares are $l^2(\mu)$ approximators of Physica D: Nonlinear Phenomena (2021) 132882doi:https://doi.org/10.1016/j.physd.2021.132882.
URL https://www.sciencedirect.com/science/article/pii/S0167278921000403

[19] A. Benveniste, M. Métivier, P. Prioure, Adaptive Algorithms and Stochastic Approximations, Springer-Verlag, 1990.

[20] V. S. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint, Hindustan Book Agency, 2009.

[21] F. S. Melo, M. I. Ribeiro, Q-learning with linear function approximation, in: N. H. Bshouty, C. Gentile (Eds.), Learning Theory, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 308–322.

[22] Z. Chen, S. Zhang, T. T. Doan, S. T. Maguluri, J.-P. Clarke, Performance of q-learning with linear function approximation: Stability and finite-time analysis, arXiv:1905.11425 (2019).

[23] X. Lin, Z. Yang, Y. Song, Short-term stock price prediction based on echo state networks, Expert Systems with Applications 36 (3, Part 2) (2009) 7313 – 7317. doi:https://doi.org/10.1016/j.eswa.2008.09.049.
URL http://www.sciencedirect.com/science/article/pii/S0957417408006519

[24] H. Zhang, J. Liang, Z. Chai, Stock prediction based on phase space reconstruction and echo state networks, Journal of Algorithms & Computational Technology 7 (1) (2013) 87–100. arXiv:https://doi.org/10.1260/1748-3018.7.1.87, doi:10.1260/1748-3018.7.1.87.
URL https://doi.org/10.1260/1748-3018.7.1.87

[25] J. Dan, W. Guo, W. Shi, B. Fang, T. Zhang, Deterministic echo state networks based stock price forecasting, Abstract and Applied Analysis 2014 (2014) 137148. doi:10.1155/2014/137148.
URL https://doi.org/10.1155/2014/137148

[26] J. Bozsik, Z. Ilonczai, Echo state network-based credit rating system, in: 2012 4th IEEE International Symposium on Logistics and Industrial Informatics, 2012, pp. 185–190.

[27] L. Maciel, F. Gomide, D. Santos, R. Ballini, Exchange rate forecasting using echo state networks for trading strategies, in: 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), 2014, pp. 40–47.

[28] M. Avellaneda, S. Stoikov, High-frequency trading in a limit order book, Quantitative Finance 8 (3) (2008) 217–224. arXiv:https://doi.org/10.1080/14697680701381228, doi:10.1080/14697680701381228.
URL https://doi.org/10.1080/14697680701381228

[29] O. Guéant, Optimal market making, Applied Mathematical Finance 24 (2) (2017) 112–154. arXiv:https://doi.org/10.1080/1350486X.2017.1342552, doi:10.1080/1350486X.2017.1342552.
URL https://doi.org/10.1080/1350486X.2017.1342552

[30] M. Fazel, R. Ge, S. Kakade, M. Mesbahi, Global Convergence of Policy Gradient Methods for the Linear

Quadratic Regulator, in: International Conference on Machine Learning, PMLR, 2018, pp. 1467–1476, iSSN: 2640-3498.

[31] Z. Mhammedi, D. J. Foster, M. Simchowitz, D. Misra, W. Sun, A. Krishnamurthy, A. Rakhlin, J. Langford, Learning the linear quadratic regulator from nonlinear observations, arXiv:2010.03799 (2020).

[32] S. Dean, H. Mania, N. Matni, B. Recht, S. Tu, On the Sample Complexity of the Linear Quadratic Regulator, Foundations of Computational Mathematics 20 (4) (2020) 633–679. `doi:10.1007/s10208-019-09426-y`.
URL `https://doi.org/10.1007/s10208-019-09426-y`

[33] S. Tu, B. Recht, The Gap Between Model-Based and Model-Free Methods on the Linear Quadratic Regulator: An Asymptotic in: Conference on Learning Theory, PMLR, 2019, pp. 3036–3083, iSSN: 2640-3498.
URL `http://proceedings.mlr.press/v99/tu19a.html`

[34] S. Bradtke, B. Ydstie, A. Barto, Adaptive linear quadratic control using policy iteration, in: Proceedings of 1994 American Control Conference - ACC '94, Vol. 3, 1994, pp. 3475–3479 vol.3. `doi:10.1109/ACC.1994.735224`.

[35] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT-press, 2015.