이학 석사 학위논문

The effect of basis functions on QLBS

(QLBS 에서 기저 함수의 영향)

2023년 2월

서울대학교 대학원

수리과학부

문상필

The effect of basis functions on QLBS

(QLBS 에서 기저 함수의 영향)

지도교수 이기암

이 논문을 이학 석사 학위논문으로 제출함

2022년 10월

서울대학교 대학원

수리과학부

문상필

문상필의 이학 석사 학위논문을 인준함

2022년 12월

위 원 장	변 순 식	(인)
부 위 원 장	이 기 암	(인)
위 원	박 형 빈	(인)

The effect of basis functions on QLBS

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Master of Science
to the faculty of the Graduate School of
Seoul National University

by

Sangpil Moon

Dissertation Director : Professor Ki-Ahm Lee

Department of Mathematical Sciences
Seoul National University

February 2023

Abstract

The effect of basis functions on QLBS

Sangpil Moon
Department of Mathematical Sciences
The Graduate School
Seoul National University

The question of whether it is suitable to select a set of basis functions in a QLBS model without any restrictions is discussed in this research. In his paper titled "*QLBS: Q-Learner in the Black-Scholes(-Merton) Worlds*", Igor Halperin proposed a discrete-time option hedging and pricing model known as QLBS. In the study, he proved that the QLBS model converges to the Black-Scholes-Merton model as the discrete-time interval converges to 0 under some circumstances, but he left the phenomenon where the discrete-time interval is a specific positive number for future work. In this work, I will demonstrate that, depending on the choice of a set of basis functions, the reward setting in the QLBS model can result in option pricing that is different from the initial outcome expected.

Key words: QLBS, Dynamic programming, Option hedging, Option pricing, Markov Decision Process
Student Number: 2018-26597

Contents

Chapter 1

Introduction

The risk of options is continuously neutralized to obtain the Black-Scholes-Merton (BSM) model. However, in reality, continuously neutralizing is impossible. So there is an inevitable risk of mis-hedging. Igor Halperin proposed a method of neutralizing risk in a discrete-time version of the BSM model and named it the QLBS model. As it seems obvious in the setting, the QLBS model converges to the BSM formulation as the interval of time steps vanishes.

To approximate the action function and Q-function, the QLBS model freely chooses a set of basis functions. In this study, we demonstrate that the reward in the QLBS model may not operate as expected when a set of basis functions is chosen without any restrictions. In the section "*NuQLear experiments*" of the paper "*The QLBS Q-Learner Goes NuQLear:Fitted Q Iteration, Inverse RL, and Option Portfolios*" [3], Igor Halperin experimented and obtained the expected result, but only when the set of basis functions is composed of cubic B-splines; it does not apply to other sets of basis functions. When the interval between discrete times converges to 0, as Igor Halperin has demonstrated, the QLBS model converges to the Black-Scholes-Merton model; nevertheless, the phenomena that occur when the interval is a specific positive value are left for future work.

This paper is organized as follows. First, we establish the common char-

acteristic of the basis function sets that can yield the greatest reward in the QLBS model. We shall demonstrate that, even though such a set of basis functions produces greater reward, the QLBS model does not produce the anticipated outcome. And I will finish the thesis with a conclusion. A model-based Dynamic Programming (DP) method and a data-driven Reinforcement Learning (RL) method are both included in the QLBS model. Since RL converges to the DP result in the same environment, I will only consider DP in this study while considering a set of basis functions.

Chapter 2

QLBS model

The QLBS model described in this chapter is based on Igor Halperin's "*QLBS: Q-Learner in the Black-Scholes(-Merton) Worlds*" [2].

2.1 Discrete portfolio

Let T be the maturity, S_t be the price of the stock at time t, $H_T(S_T)$ be the payoff from the European option terminal at maturity, B_t be a risk-free bank deposit, u_t be the position in the stock at time t, and Π_t be the portfolio. The relationship is as follows at $t \leq T$.

$$\Pi_t = u_t S_t + B_t \tag{2.1.1}$$

Assume that there are no transaction fees and a self-financing constraint, meaning that neither internal nor external money is allowed to flow. The equation below will be reached if you sell all of the remaining stocks for cash with a format of $u_T = 0$ at maturity T.

$$\Pi_T = B_T = H_T(S_T) \tag{2.1.2}$$

Now we may get the relational expression as below, where r is the risk-free interest rate.

$$u_t S_{t+\Delta t} + e^{r\Delta t} B_t = u_{t+\Delta t} S_{t+\Delta t} + B_{t+\Delta t} \tag{2.1.3}$$

$$B_t = e^{-r\Delta t}\left[B_{t+\Delta t} + (u_{t+\Delta t} - u_t) S_{t+\Delta t}\right], \quad t = T - \Delta t, \ldots, \Delta t, 0. \tag{2.1.4}$$

Substitute this expression into (2.1.1) and rearrange it as follows.

$$\begin{aligned}
\Pi_t &= u_t S_t + e^{-r\Delta t}\left[B_{t+\Delta t}\left(u_{t+\Delta t} - u_t\right) S_{t+\Delta t}\right] \\
&= e^{-r\Delta t}\left[B_{t+\Delta t} + u_{t+\Delta t} S_{t+\Delta t} + e^{r\Delta t} u_t S_t - u_t S_{t+\Delta t}\right] \\
&= e^{-r\Delta t}\left[\Pi_{t+\Delta t} - u_t\left(S_{t+\Delta t} - e^{r\Delta t} S_t\right)\right] \\
&= e^{-r\Delta t}\left[\Pi_{t+\Delta t} - u_t \Delta S_t\right], \quad \Delta S_t = S_{t+\Delta t} - e^{r\Delta t} S_t, \quad t = T - \Delta t, \ldots, \Delta t, 0.
\end{aligned}$$

Then the following expression can be obtained.

$$\Pi_t = e^{-r\Delta t}\left[\Pi_{t+\Delta t} - u_t \Delta S_t\right], \quad \Delta S_t = S_{t+\Delta t} - e^{r\Delta t} S_t, \quad t = T - \Delta t, \ldots, \Delta t, 0. \tag{2.1.5}$$

From the viewpoint of minimizing risk, identify $u_t(S_t)$ with the smallest variance while considering S_t as a random variable. When the available cross-sectional information, which refers to the data of all concurrent pathways, is expressed as $\mathcal{F}_t$ at time t, it is obtained as follows.

$$\begin{aligned}
u_t^*(S_t) &= \underset{u}{\operatorname{argmin}}\, Var\left[\Pi_t | \mathcal{F}_t\right] \\
&= \underset{u}{\operatorname{argmin}}\, Var\left[\Pi_{t+\Delta t} - u_t \Delta S_t | \mathcal{F}_t\right], \quad t = T - \Delta t, \ldots, \Delta t, 0.
\end{aligned} \tag{2.1.6}$$

The optimal hedge can be calculated analytically using the derivative of (2.6) with $u_t(S_t)$ as a variable.

$$
\begin{aligned}
&Var\left[\Pi_{t+\Delta t} - u_t(S_t)\Delta S_t|\mathcal{F}_t\right] \\
&\quad = \mathbb{E}\left[\left(\Pi_{t+\Delta t} - u_t(S_t)\Delta S_t\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Pi_{t+\Delta t} - u_t(S_t)\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&\quad = \mathbb{E}\left[\Pi^2_{t+\Delta t}|\mathcal{F}_t\right] - 2u_t(S_t)\mathbb{E}\left[\Pi_{t+\Delta t}\Delta S_t|\mathcal{F}_t\right] + u_t^2(S_t)\mathbb{E}\left[\left(\Delta S_t\right)^2|\mathcal{F}_t\right] \\
&\qquad - \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right]^2 + 2u_t(S_t)\mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right]\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right] - u_t^2(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]^2 \\
&\quad = Var\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right] - 2u_t(S_t)Cov\left[\Pi_{t+\Delta t}, \Delta S_t|\mathcal{F}_t\right] + u_t^2(S_t)Var\left[\Delta S_t\right]
\end{aligned}
$$

This provides

$$
u_t^*(S_t) = \frac{Cov\left[\Pi_{t+\Delta t}, \Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]}, \quad t = T - \Delta t, \ldots, \Delta t, 0. \tag{2.1.7}
$$

2.2 Hedging and pricing at $\Delta t \to 0$

Consider the idea of *fair* option pricing, which is defined as the "time-t expected value of the hedge portfolio Π_t" [2]. Here is the definition.

$$
C_t = \mathbb{E}\left[\Pi_t|\mathcal{F}_t\right] \tag{2.2.1}
$$

Using (2.1.5), it may be written as

$$
\begin{aligned}
C_t &= \mathbb{E}\left[e^{-r\Delta t}\Pi_{t+\Delta t} - e^{-r\Delta t}u_t(S_t)\Delta S_t|\mathcal{F}_t\right] \\
&= e^{-r\Delta t}\mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right] - e^{-r\Delta t}\mathbb{E}\left[u_t(S_t)\Delta S_t|\mathcal{F}_t\right] \\
&= e^{-r\Delta t}\mathbb{E}\left[\mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right]|\mathcal{F}_t\right] - e^{-r\Delta t}u_t(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right] \\
&= e^{-r\Delta t}\mathbb{E}\left[C_{t+\Delta t}|\mathcal{F}_t\right] - e^{-r\Delta t}u_t(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right] \\
&= e^{-r\Delta t}\left(\mathbb{E}\left[C_{t+\Delta t}|\mathcal{F}_t\right] - u_t(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)
\end{aligned} \tag{2.2.2}
$$

(2.1.7) expressed as C_t is:

$$\begin{aligned}
u_t^*(S_t) &= \frac{Cov\left[\Pi_{t+\Delta t}, \Delta S_t | \mathcal{F}_t\right]}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \\
&= \frac{\mathbb{E}\left[\left(\Pi_{t+\Delta t} - \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right]\right)\left(\Delta S_t - \mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} \\
&= \frac{\mathbb{E}\left[\Pi_{t+\Delta t}\Delta S_t|\mathcal{F}_t\right] - \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right]\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[\Pi_{t+\Delta t}\Delta S_t|\mathcal{F}_{t+\Delta t}\right]|\mathcal{F}_t\right] - \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_t\right]\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} \\
&= \frac{\mathbb{E}\left[\Delta S_t\mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right]|\mathcal{F}_t\right] - \mathbb{E}\left[\mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right]|\mathcal{F}_t\right]\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} \\
&= \frac{\mathbb{E}\left[\Delta S_t C_{t+\Delta t}|\mathcal{F}_t\right] - \mathbb{E}\left[C_{t+\Delta t}|\mathcal{F}_t\right]\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} \\
&= \frac{Cov\left[C_{t+\Delta t}, \Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]}
\end{aligned} \tag{2.2.3}$$

It will be demonstrated that C_t closely approximates the solution of the Black-Scholes-Merton Model at the optimal hedge when S_t has a geometric Brownian motion, that is,

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \tag{2.2.4}$$

as $\Delta t \to 0$, where W_t is a standard Brownian motion, and μ and σ are constants.

Using the first-order Taylor expansion, the relationship between $C_{t+\Delta t}$ and C_t is expressed as follows:

$$C_{t+\Delta t} = C_t + \frac{\partial C_t}{\partial S_t}\overline{\Delta} S_t + O\left(\left(\overline{\Delta} S_t\right)^2\right) = C_t + \frac{\partial C_t}{\partial S_t}\overline{\Delta} S_t + O\left(\Delta t\right) \tag{2.2.5}$$

Note that the symbol $\overline{\Delta} S_t$ means $S_{t+\Delta t} - S_t$, which is different from ΔS_t in

(2.1.5). By substituting this into (2.2.3) and $\Delta t \to 0$, we get:

$$
\begin{aligned}
\lim_{\Delta t \to 0} u_t^*(S_t) &= \lim_{\Delta t \to 0} \frac{Cov\left[C_t + \frac{\partial C_t}{\partial S_t}\overline{\Delta} S_t + O(\Delta t), \Delta S_t | \mathcal{F}_t\right]}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \\
&= \lim_{\Delta t \to 0} \frac{1}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \Bigg(\mathbb{E}\left[\left(C_t + \frac{\partial C_t}{\partial S_t}\overline{\Delta} S_t + O(\Delta t)\right) \Delta S_t | \mathcal{F}_t\right] \\
&\qquad - \mathbb{E}\left[C_t + \frac{\partial C_t}{\partial S_t}\overline{\Delta} S_t + O(\Delta t) | \mathcal{F}_t\right] \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] \Bigg) \\
&= \lim_{\Delta t \to 0} \frac{1}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \Bigg(C_t \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] + \frac{\partial C_t}{\partial S_t} \mathbb{E}\left[\overline{\Delta} S_t \Delta S_t | \mathcal{F}_t\right] + \mathbb{E}\left[O(\Delta t) \Delta S_t | \mathcal{F}_t\right] \\
&\quad - C_t \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] - \frac{\partial C_t}{\partial S_t} \mathbb{E}\left[\overline{\Delta} S_t | \mathcal{F}_t\right] \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] - \mathbb{E}\left[O(\Delta t) | \mathcal{F}_t\right] \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] \Bigg) \\
&= \lim_{\Delta t \to 0} \frac{\frac{\partial C_t}{\partial S_t}}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \Bigg(\mathbb{E}\left[\left(\Delta S_t + e^{r\Delta t} S_t - S_t\right) \Delta S_t | \mathcal{F}_t\right] \\
&\qquad - \mathbb{E}\left[\Delta S_t + e^{r\Delta t} S_t - S_t | \mathcal{F}_t\right] \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] \Bigg) \\
&= \lim_{\Delta t \to 0} \frac{\partial C_t}{\partial S_t} \left(\frac{\left(\mathbb{E}\left[\Delta S_t \Delta S_t | \mathcal{F}_t\right] - \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right]\right)}{Var\left[\Delta S_t | \mathcal{F}_t\right]} \right) \\
&= \frac{\partial C_t}{\partial S_t}
\end{aligned}
\tag{2.2.6}
$$

Considering a sufficiently small Δt and an optimal hedge in the second term of (2.2.2),

$$
\begin{aligned}
u_t(S_t) \mathbb{E}\left[\Delta S_t | \mathcal{F}_t\right] &= u_t^*(S_t) \mathbb{E}\left[S_{t+\Delta t} - e^{r\Delta t} S_t | \mathcal{F}_t\right] \\
&= u_t^*(S_t) \mathbb{E}\left[S_{t+\Delta t} - S_t - S_t\left(e^{r\Delta t} - 1\right) | \mathcal{F}_t\right] \\
&\approx u_t^*(S_t) \mathbb{E}\left[dS_t - rS_t dt | \mathcal{F}_t\right] \\
&= u_t^*(S_t) \mathbb{E}\left[\mu S_t dt + \sigma S_t dW_t - rS_t dt | \mathcal{F}_t\right] \\
&= u_t^*(S_t)(\mu - r) S_t dt \approx \frac{\partial C_t}{\partial S_t}(\mu - r) S_t dt
\end{aligned}
\tag{2.2.7}
$$

Using the second-order Taylor expansion, the relationship between $C_{t+\Delta t}$ and

C_t is expressed as follows:

$$\begin{aligned}
C_{t+\Delta t} &= C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}dS_t + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}\left(dS_t\right)^2 + \cdots \\
&= C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}S_t\left(\mu dt + \sigma dW_t\right) \\
&\qquad + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}S_t^2\left(\sigma^2 dW_t^2 + 2\mu\sigma dW_t dt\right) + O\left(dt^2\right) \qquad (2.2.8)
\end{aligned}$$

We obtain the following by substituting this into equation (2.2.2), along with equation (2.2.7) and the optimal hedge.

$$\begin{aligned}
C_t &= e^{-r\Delta t}\left(\mathbb{E}\left[C_{t+\Delta t}|\mathcal{F}_t\right] - u_t^*(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right) \\
&= e^{-r\Delta t}\Bigg(\mathbb{E}\Big[C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}S_t\left(\mu dt + \sigma dW_t\right) \\
&\quad + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}S_t^2\left(\sigma^2 dW_t^2 + 2\mu\sigma dW_t dt\right) + O\left(dt^2\right)|\mathcal{F}_t\Big] - u_t^*(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\Bigg) \\
&= e^{-r\Delta t}\Bigg(C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}S_t\mu dt + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}S_t^2\sigma^2 dt \\
&\qquad + \mathbb{E}\left[O\left(dt^2\right)|\mathcal{F}_t\right] - u_t^*(S_t)\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\Bigg) \\
&\approx e^{-r\Delta t}\Bigg(C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}S_t\mu dt + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}S_t^2\sigma^2 dt \\
&\qquad + \mathbb{E}\left[O\left(dt^2\right)|\mathcal{F}_t\right] - \frac{\partial C_t}{\partial S_t}\left(\mu - r\right)S_t dt\Bigg)
\end{aligned}$$

If the left and right sides are arranged,

$$C_t\left(e^{r\Delta t} - e^{r\cdot 0}\right) \approx \left(\frac{\partial C_t}{\partial t} + rS_t\frac{\partial C_t}{\partial S_t} + \sigma^2 S_t^2\frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}\right)dt + e^{-r\Delta t}\mathbb{E}\left[O\left(dt^2\right)|\mathcal{F}_t\right]$$

If $\Delta t \to 0$ is taken on both sides, the above equation becomes the Black-Scholes equation.

$$\frac{\partial C_t}{\partial t} + rS_t\frac{\partial C_t}{\partial S_t} + \sigma^2 S_t^2\frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2} - rC_t = 0$$

2.3 Transformation to stationary state variables

The QLBS model converts S_t to X_t and uses it.

$$X_t = -\left(\mu - \frac{\sigma^2}{2}\right) t + \log S_t$$

The reason is that, as can be seen below, X_t is expected to lower the interval covered by a set of basis functions rather than S_t, and that S_t and X_t are mutually convertible.

$$\begin{aligned} dX_t &= -\left(\mu - \frac{\sigma^2}{2}\right) dt + d\log S_t \\ &= -\left(\mu - \frac{\sigma^2}{2}\right) dt + \left(\left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t\right) = \sigma dW_t \end{aligned}$$

To explain the contents naturally and simply while concentrating on the topic of this paper, I will utilize S_t as it is rather than X_t.

2.4 Bellman Equations

When looking at $C_t = \mathbb{E}\left[\Pi_t | \mathcal{F}_t\right]$ as the option price from the standpoint of selling the option, there is an unconsidered risk. Since the bank deposit B_t in (2.1.1) is fixed at time $t = 0$ to B_0, it runs the danger of being depleted over time. Here is one of the option price models that accounts for this risk.

$$C_0^{(\text{ask})}(S, u) = \mathbb{E}_0\left[\Pi_0 + \lambda \sum_{t=0}^{T} e^{-rt} Var\left[\Pi_t | \mathcal{F}_t\right] \middle| S_0 = S, u_0 = u\right] \tag{2.4.1}$$

λ is referred to as the risk-aversion parameter in this context, and as its name implies, it controls how sensitively to reject risk. And the index of $\sum_{t=0}^{T}$ is a series of numbers that goes from t to T increasing by Δt. The value function

in the QLBS model is defined as follows when the stock price is S_t at time t.

$$V_t(S_t) = \mathbb{E}\left[-\Pi_t - \lambda \sum_{t'=t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_t\right] \tag{2.4.2}$$

Additionally, the following introduces $\pi\left(t, S_t\right)$.

$$\pi : \{0, \Delta t, \ldots, T - \Delta t\} \times \mathcal{X} \to \mathcal{A}$$

$\mathcal{X}$ is the entire state set, $\mathcal{A}$ is the entire action set, and π is a function of those two variables. In other words, if $a_t \in \mathcal{A}$ and $x_t \in \mathcal{X}$, $a_t = \pi\left(t, x_t\right)$ follows. And this serves as $\mu_t(S_t)$ of the previous exposition.

Let's look at (2.4.2).

$$V_{t+\Delta t}(S_{t+\Delta t}) = \mathbb{E}\left[-\Pi_{t+\Delta t} - \lambda \sum_{t'=t+\Delta t}^{T} e^{-r(t'-(t+\Delta t))} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_{t+\Delta t}\right]$$

$$\begin{aligned} &V_{t+\Delta t}(S_{t+\Delta t}) + \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right] \\ &\qquad = \mathbb{E}\left[-\lambda \sum_{t'=t+\Delta t}^{T} e^{-r(t'-(t+\Delta t))} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_{t+\Delta t}\right] \end{aligned}$$

$$\begin{aligned} &e^{-r\Delta t}\Big(V_{t+\Delta t}(S_{t+\Delta t}) + \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right]\Big) \\ &\qquad = -\lambda \mathbb{E}\left[\sum_{t'=t+\Delta t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_{t+\Delta t}\right] \end{aligned}$$

Therefore, using the above equation and (2.1.5), $V_t^{\pi}(S_t)$ is expressed as follows. Here, $V_t^{\pi}(S_t)$ represents the expected value of the total rewards received

by following policy π at time t and state S_t.

$$
\begin{aligned}
V_t^\pi(S_t) &= \mathbb{E}\left[-\Pi_t - \lambda \sum_{t'=t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[-\Pi_t - \lambda Var\left[\Pi_t|\mathcal{F}_t\right] - \lambda \sum_{t'=t+\Delta t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[-\Pi_t \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[Var\left[\Pi_t|\mathcal{F}_t\right] \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[\sum_{t'=t+\Delta t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[-\Pi_t \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[Var\left[\Pi_t|\mathcal{F}_t\right] \middle| \mathcal{F}_t\right] \\
&\qquad + \mathbb{E}\left[-\lambda \mathbb{E}\left[\sum_{t'=t+\Delta t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}|\mathcal{F}_{t'}\right] \middle| \mathcal{F}_{t+\Delta t}\right] \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[-\Pi_t \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[Var\left[\Pi_t|\mathcal{F}_t\right] \middle| \mathcal{F}_t\right] \\
&\qquad + \mathbb{E}\left[e^{-r\Delta t}\left(V_{t+\Delta t}^\pi(S_{t+\Delta t}) + \mathbb{E}\left[\Pi_{t+\Delta t}|\mathcal{F}_{t+\Delta t}\right]\right) \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[e^{-r\Delta t}\Pi_{t+\Delta t} - \Pi_t \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[Var\left[\Pi_t|\mathcal{F}_t\right] \middle| \mathcal{F}_t\right] + e^{-r\Delta t}\mathbb{E}\left[V_{t+\Delta t}^\pi(S_{t+\Delta t}) \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}\left[e^{-r\Delta t} a_t \Delta S_t \middle| \mathcal{F}_t\right] - \lambda \mathbb{E}\left[Var\left[\Pi_t|\mathcal{F}_t\right] \middle| \mathcal{F}_t\right] + e^{-r\Delta t}\mathbb{E}\left[V_{t+\Delta t}^\pi(S_{t+\Delta t}) \middle| \mathcal{F}_t\right]
\end{aligned}
\tag{2.4.3}
$$

For convenience, let's denote $e^{-r\Delta t}$ as γ.
To obtain the Bellman equation for the QLBS model, let's define $R(S_t, a_t, S_{t+\Delta t})$ as:

$$
R(S_t, a_t, S_{t+\Delta t}) = \gamma a_t \Delta S_t - \lambda Var\left[\Pi_t \middle| \mathcal{F}_t\right] \tag{2.4.4}
$$

And substituting (2.4.4) into (2.4.3) and rearranging (2.4.3), the Bellman equation for the QLBS model can be obtained as follows.

$$
V_t^\pi(S_t) = \mathbb{E}^\pi\left[R(S_t, a_t, S_{t+\Delta t}) + \gamma V_{t+\Delta t}^\pi(S_{t+\Delta t})\right] \tag{2.4.5}
$$

As a result, $R(S_t, a_t, S_{t+\Delta t})$ from equation (2.4.4) is now the reward in the QLBS model. The following factors also affect how the reward $R(S_t, a_t, S_{t+\Delta t})$

is expressed.

$$
\begin{aligned}
Var\left[\Pi_t \middle| \mathcal{F}_t\right] &= \mathbb{E}\left[\Pi_t^2 \middle| \mathcal{F}_t\right] - \left(\mathbb{E}\left[\Pi_t \middle| \mathcal{F}_t\right]\right)^2 \\
&= \gamma^2 \left(\mathbb{E}\left[\left(\Pi_{t+\Delta t} - a_t \Delta S_t\right)^2 \middle| \mathcal{F}_t\right] - \left(\mathbb{E}\left[\Pi_{t+\Delta t} \middle| \mathcal{F}_t\right] - \mathbb{E}\left[a_t \Delta S_t \middle| \mathcal{F}_t\right]\right)^2\right) \\
&= \gamma^2 \Big(\mathbb{E}\left[\Pi_{t+\Delta t}^2 \middle| \mathcal{F}_t\right] - \left(\mathbb{E}\left[\Pi_{t+\Delta t} \middle| \mathcal{F}_t\right]\right)^2 - 2\mathbb{E}\left[\Pi_{t+\Delta t} a_t \Delta S_t \middle| \mathcal{F}_t\right] \\
&+ 2\mathbb{E}\left[\Pi_{t+\Delta t} \middle| \mathcal{F}_t\right] \mathbb{E}\left[a_t \Delta S_t \middle| \mathcal{F}_t\right] + \mathbb{E}\left[\left(a_t \Delta S_t\right)^2 \middle| \mathcal{F}_t\right] - \left(\mathbb{E}\left[a_t \Delta S_t \middle| \mathcal{F}_t\right]\right)^2 \Big) \\
&= \gamma^2 \Big(Var\left[\Pi_{t+\Delta t} \middle| \mathcal{F}_t\right] - 2\, Cov\left[\Pi_{t+\Delta t}, a_t \Delta S_t \middle| \mathcal{F}_t\right] + Var\left[a_t \Delta S_t \middle| \mathcal{F}_t\right]\Big)
\end{aligned}
\tag{2.4.6}
$$

Substituting (2.4.6) into (2.4.4),

$$
\begin{aligned}
R(S_t, a_t, S_{t+\Delta t}) &= \gamma a_t \Delta S_t - \lambda Var\left[\Pi_t \middle| \mathcal{F}_t\right] \qquad (2.4.7) \\
&= \gamma a_t \Delta S_t \\
&- \lambda \gamma^2 \Big(Var\left[\Pi_{t+\Delta t} \middle| \mathcal{F}_t\right] - 2\, Cov\left[\Pi_{t+\Delta t}, a_t \Delta S_t \middle| \mathcal{F}_t\right] + Var\left[a_t \Delta S_t \middle| \mathcal{F}_t\right]\Big)
\end{aligned}
$$

Now, $V_t^\pi(S_t)$ can be calculated backward in time based on (2.4.5).
$V_T^\pi(S_T) = -\Pi_T(S_T) - \lambda Var\left[\Pi_T\right]$ is a terminal condition at time $t = T$ that starts backward recursion.
The definition of the action-value function, or Q-function, is similar to (2.4.2).

$$
Q_t^\pi(s, a) = \mathbb{E}^\pi \left[-\Pi_t(S_t) - \lambda \sum_{t'=t}^{T} e^{-r(t'-t)} Var\left[\Pi_{t'}(S_{t'}) | \mathcal{F}_{t'}\right] \middle| S_t = s, a_t = a \right]
\tag{2.4.8}
$$

The optimal policy π_t^* is definded as follows.

$$
\pi_t^*(S_t) = \arg\max_\pi V_t^\pi(S_t) = \arg\max_{a_t \in \mathcal{A}} Q_t^*(S_t, a_t)
$$

Therefore, the Bellman optimality equation for the action-value function is:

$$
Q_t^*(s, a) = \mathbb{E}\left[R_t\left(S_t, a_t, S_{t+\Delta t}\right) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}\right) \middle| S_t = s, a_t = a \right]
\tag{2.4.9}
$$

, where $t = 0, \Delta t, \ldots, T - \Delta t$.

$$Q_T^* \left(S_T, a_T = 0\right) = -\Pi_T \left(S_T\right) - \lambda Var \left[\Pi_T \left(S_T\right)\right] \tag{2.4.10}$$

is a terminal condition at time $t = T$ that starts backward recursion. Using (2.1.2), $\Pi_T \left(S_T\right)$ can be obtained.

2.5 Optimal Policy

Substituting equation (2.4.7) into (2.4.9) gives:

$$\begin{aligned}
Q_t^*(s,a) &= \gamma \mathbb{E}\left[Q_{t+\Delta t}^* \left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) + a_t \Delta S_t \middle| S_t = s, a_t = a\right] \\
&- \lambda \gamma^2 \mathbb{E}\Bigg[Var\left[\Pi_{t+\Delta t} \middle| S_t = s, a_t = a\right] - 2a_t \, Cov\left[\Pi_{t+\Delta t}, \Delta S_t \middle| S_t = s, a_t = a\right] \\
&\quad + a_t^2 Var\left[\Delta S_t \middle| S_t = s, a_t = a\right] \Bigg| S_t = s, a_t = a\Bigg] \\
&\quad = \gamma \mathbb{E}\left[Q_{t+\Delta t}^* \left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) + a_t \Delta S_t \middle| S_t = s, a_t = a\right] \\
&- \lambda \gamma^2 \Bigg[Var\left[\Pi_{t+\Delta t} \middle| S_t = s, a_t = a\right] - 2a \, Cov\left[\Pi_{t+\Delta t}, \Delta S_t \middle| S_t = s, a_t = a\right] \\
&\quad + a^2 Var\left[\Delta S_t \middle| S_t = s, a_t = a\right]\Bigg]
\end{aligned} \tag{2.5.1}$$

It should be noted here that $\mathbb{E}\left[Q_{t+\Delta t}^* \left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) | S_t = s, a_t = a\right]$ depends on a_t only by the conditional probability $p\left(S_{t+\Delta t} | S_t, a_t\right)$. However, it can be said that $\mathbb{E}\left[Q_{t+\Delta t}^* \left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) | S_t = s, a_t = a\right]$ does not depend on it because the option buyer or seller does not have any impact on the market according to the assumption of the Black-Scholes model. Therefore, $Q_t^* \left(S_t, a_t\right)$

is quadratic with respect to a_t.

$$Q_t^*(S_t, a_t) = \gamma \mathbb{E}\left[Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}^*\right)\middle|\mathcal{F}_t\right] + \gamma a_t \mathbb{E}\left[\Delta S_t\middle|\mathcal{F}_t\right] \tag{2.5.2}$$
$$- \lambda\gamma^2\Big(Var\left[\Pi_{t+\Delta t}\middle|\mathcal{F}_t\right] - 2a_t\ Cov\left[\Pi_{t+\Delta t}, \Delta S_t\middle|\mathcal{F}_t\right] + a_t^2 Var\left[\Delta S_t\middle|\mathcal{F}_t\right]\Big)$$

(2.5.2) is calculated as follows when performing backward recursion by applying the Monte Carlo method to the stock price paths later.

$$Q_t^*(S_t, a_t) = \gamma \mathbb{E}\left[Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) + a_t \Delta S_t\middle|\mathcal{F}_t\right] \tag{2.5.3}$$
$$- \lambda\gamma^2 \mathbb{E}\left[\hat{\Pi}_{t+\Delta t}^2 - 2a_t\hat{\Pi}_{t+\Delta t}\Delta\hat{S}_t + a_t^2\left(\Delta\hat{S}_t\right)^2\middle|\mathcal{F}_t\right], t = 0, \Delta t, \ldots, T - \Delta t$$

In (2.5.3), it is defined as $\hat{\Pi}_{t+\Delta t} := \Pi_{t+\Delta t} - \overline{\Pi}_{t+\Delta t}$, where $\overline{\Pi}_{t+\Delta t}$ is the sample mean of all values of $\Pi_{t+\Delta t}$. $\Delta\hat{S}_t$ is similarly defined.

For reference,

$$Q_t^*(S_t, a_t) = \gamma \mathbb{E}\left[Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}^*\right) + a_t \Delta S_t\middle|\mathcal{F}_t\right]$$

is the result of changing λ to 0 in (2.5.1). In (2.4.10), if $\lambda = 0$, then

$$Q_T^*\left(S_T, a_T = 0\right) = -\Pi_T\left(S_T\right)$$

So, $Q_t^*\left(S_t, a_t\right) = -\Pi_t\left(S_t, a_t\right)$ is obtained by the above equations and (2.1.5). It can be rephrased as follows utilizing the concept of a fair option price (2.2.1).

$$C_t = \gamma \mathbb{E}\left[C_{t+\Delta t} - a_t \Delta S_t\middle|\mathcal{F}_t\right] \tag{2.5.4}$$

This phrase has the same meaning as (2.2.2). Additionally, when $\Delta t \to 0$, as discussed in section 2.2, we have the Black-Scholes equation if we replace a_t with the optimal hedge (2.2.3). In an experiment designed to demonstrate the effect of a set of basis functions, the discussion just made will be employed.

Returning to the main topic, using that $Q_t^*\left(S_t, a_t\right)$ is quadratic with re-

spect to a_t, the optimal action is obtained as follows.

$$a_t^*(S_t) = \frac{Cov\left[\Pi_{t+\Delta t}, \Delta S_t|\mathcal{F}_t\right] + \frac{1}{2\gamma\lambda}\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]}, \quad t = T - \Delta t, \ldots, \Delta t, 0. \tag{2.5.5}$$

Let's now determine the limit of equation (2.5.5) when $\Delta t \to 0$.
Through (2.2.6), it is possible to determine that

$$\lim_{\Delta t \to 0} \frac{Cov\left[\Pi_{t+\Delta t}, \Delta S_t|\mathcal{F}_t\right]}{Var\left[\Delta S_t|\mathcal{F}_t\right]} = \frac{\partial C_t}{\partial S_t} \tag{2.5.6}$$

From the fact that

$$\begin{aligned}
Var\left[\Delta S_t|\mathcal{F}_t\right] &= \mathbb{E}\left[\left(\Delta S_t - \mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[\left(\Delta S_t\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \mathbb{E}\left[\left(S_{t+\Delta t} - e^{r\Delta t}S_t\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \mathbb{E}\left[\left(S_{t+\Delta t} - S_t + S_t - e^{r\Delta t}S_t\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \mathbb{E}\left[\left(dS_t - S_t\left(e^{r\Delta t} - 1\right)\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \mathbb{E}\left[\left(\mu S_t dt + \sigma S_t dW_t - S_t\left(e^{r\Delta t} - 1\right)\right)^2|\mathcal{F}_t\right] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \mathbb{E}\Big[\left(\mu S_t dt\right)^2 + \left(\sigma S_t dW_t\right)^2 + \left(S_t\left(e^{r\Delta t} - 1\right)\right)^2 + 2\mu\sigma S_t^2 dt\ dW_t \\
&\qquad - 2\sigma S_t^2\left(e^{r\Delta t} - 1\right)dW_t - 2\mu S_t^2\left(e^{r\Delta t} - 1\right)dt\Big|\mathcal{F}_t\Big] - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \left(\mu S_t dt\right)^2 + \left(\sigma S_t\right)^2 dt \\
&\qquad + \left(S_t\left(e^{r\Delta t} - 1\right)\right)^2 - 2\mu S_t^2\left(e^{r\Delta t} - 1\right)dt - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&\approx \left(\mu S_t dt\right)^2 + \left(\sigma S_t\right)^2 dt + \left(S_t r\ dt\right)^2 - 2\mu S_t^2 r\left(dt\right)^2 - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&= \left(\left(\mu - r\right)S_t dt\right)^2 + \left(\sigma S_t\right)^2 dt - \left(\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right]\right)^2 \\
&\approx \left(\left(\mu - r\right)S_t dt\right)^2 + \left(\sigma S_t\right)^2 dt - \left(\left(\mu - r\right)S_t dt\right)^2 \\
&= \left(\sigma S_t\right)^2 dt
\end{aligned} \tag{2.5.7}$$

and

$$\begin{aligned}
\mathbb{E}\left[\Delta S_t|\mathcal{F}_t\right] &= \mathbb{E}\left[S_{t+\Delta t} - e^{r\Delta t}S_t|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[S_{t+\Delta t} - S_t + S_t - e^{r\Delta t}S_t|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[dS_t - S_t\left(e^{r\Delta t} - 1\right)|\mathcal{F}_t\right] \\
&\approx \mathbb{E}\left[\mu S_t dt + \sigma S_t dW_t - rS_t dt|\mathcal{F}_t\right] \\
&= (\mu - r)\, S_t dt
\end{aligned} \tag{2.5.8}$$

, the following conclusion can be drawn by combining equations (2.5.6), (2.5.8), and (2.5.7).

$$\lim_{\Delta t \to 0} a_t^*(S_t) = \frac{\partial C_t}{\partial S_t} + \frac{\mu - r}{2\lambda\sigma^2}\frac{1}{S_t}, \quad t = T - \Delta t, \ldots, \Delta t, 0. \tag{2.5.9}$$

From this, it can be seen that if $\mu = r$ or $\lambda \to \infty$ for a risk-aversion parameter λ, the same result as (2.2.6) is obtained when $\Delta t \to 0$. And according to (2.4.1) and (2.4.8), the following relationship can be found.

$$C_0^{(\text{ask})}(S_0, a_0^*) = -Q_0^*(S_0, a_0^*)$$

From now on, I will outline how to use the backward recursion in a Monte Carlo setting to determine option pricing for stock price paths. At this point, the QLBS model assumes that a set of basis functions $\{\Phi_n(x)\}$ has been chosen arbitrarily, and the discussion begins. Assume that $a_t^*(S_t)$ and $Q_t^*(S_t, a_t^*)$ are expanded using $\{\Phi_n(x)\}$ and represented as follows.

$$a_t^*(S_t) = \sum_n^N \phi_{nt}\Phi_n(S_t), \qquad Q_t^*(S_t, a_t^*) = \sum_n^N \omega_{nt}\Phi_n(S_t) \tag{2.5.10}$$

Let's start by determining the coefficients $\{\phi_{nt}\}$ for the optimal action. Assume that there are as many stock price paths as K_{MC}. The optimal action expansion can be achieved by substituting equation (2.5.10) into equation (2.5.3) and minimizing $G_t(\phi)$ (2.5.11) derived by using the Monte Carlo es-

timate.

$$G_t(\phi) = \sum_{k=1}^{K_{MC}} \left(-\sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^k\right) \Delta S_t^k + \gamma\lambda \left(\hat{\Pi}_{t+1}^k - \sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^k\right) \Delta \hat{S}_t^k \right)^2 \right) \tag{2.5.11}$$

The gradient of $G_t(\phi)$ for ϕ is

$$\begin{aligned}
&\frac{\partial G_t}{\partial \phi_{it}} \\
&= \sum_{k=1}^{K_{MC}} \left(-\Phi_i \left(S_t^k\right) \Delta S_t^k - 2\gamma\lambda \left(\hat{\Pi}_{t+\Delta t}^k - \sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^k\right) \Delta \hat{S}_t^k \right) \Phi_i \left(S_t^k\right) \Delta \hat{S}_t^k \right) \\
&= \sum_{k=1}^{K_{MC}} \Bigg(-\left(2\gamma\lambda \hat{\Pi}_{t+\Delta t}^k \Phi_i(S_t^k) \Delta \hat{S}_t^k + \Phi_i(S_t^k) \Delta S_t^k \right) \\
&\qquad\qquad + 2\gamma\lambda \sum_{n}^{N} \phi_{nt} \Phi_n(S_t^k) \Phi_i(S_t^k) \left(\Delta \hat{S}_t^k \right)^2 \Bigg)
\end{aligned} \tag{2.5.12}$$

When $\frac{\partial G_t}{\partial \phi_{it}} = 0$, it is expressed as

$$\begin{aligned}
&\sum_{k=1}^{K_{MC}} \left(\sum_{n}^{N} \phi_{nt} \Phi_n(S_t^k) \Phi_i(S_t^k) \left(\Delta \hat{S}_t^k \right)^2 \right) \\
&\qquad = \sum_{k=1}^{K_{MC}} \left(\hat{\Pi}_{t+\Delta t}^k \Phi_i(S_t^k) \Delta \hat{S}_t^k + \frac{1}{2\gamma\lambda} \Phi_i(S_t^k) \Delta S_t^k \right)
\end{aligned} \tag{2.5.13}$$

Now we define $N \times N$ matrix A and $N \times 1$ matrix B as follows.

$$\left(A^{(t)}\right)_{ij} := \sum_{k=1}^{K_{MC}} \Phi_i(S_t^k) \Phi_j(S_t^k) \left(\Delta \hat{S}_t^k \right)^2 \tag{2.5.14}$$

$$\left(B^{(t)}\right)_{i1} := \sum_{k=1}^{K_{MC}} \left(\hat{\Pi}_{t+\Delta t}^k \Phi_i(S_t^k) \Delta \hat{S}_t^k + \frac{1}{2\gamma\lambda} \Phi_i(S_t^k) \Delta S_t^k \right)$$

From (2.5.14) and (2.5.13),

$$A^{(t)} \phi_t^* = B^{(t)} \tag{2.5.15}$$

Therefore, the coefficients of expansion of the optimal action $a_t^*(S_t)$ are

$$\phi_t^* = \left(A^{(t)}\right)^{-1} B^{(t)} \tag{2.5.16}$$

Going forward, let's look for coefficients $\{\omega_{nt}\}$ for Q-function $Q_t^*(S_t, a_t^*)$. From (2.4.9),

$$R_t\left(S_t, a_t^*, S_{t+\Delta t}\right) + \gamma \max_{a_{t+\Delta t}\in\mathcal{A}} Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}\right) = Q_t^*(S_t, a_t^*) + \epsilon_t \ , \tag{2.5.17}$$

where ϵ_t is a random noise and the mean of ϵ_t is zero. Therefore, by inserting (2.5.10) into (2.5.17) and identifying the coefficients that minimize the square sum of ϵ_t, one can derive the coefficients $\{\omega_{nt}\}$ of expansion of the optimal Q-function $Q_t^*(S_t, a_t^*)$. In terms of formula, it can be viewed as resolving the least squares optimization problem.

$$F_t(\omega) = \tag{2.5.18}$$
$$\sum_{k=1}^{K_{MC}} \left(R_t(S_t^k, a_t^*, S_{t+\Delta t}^k) + \gamma \max_{a_{t+\Delta t}\in\mathcal{A}} Q_{t+\Delta t}^*\left(S_{t+\Delta t}^k, a_{t+\Delta t}\right) - \sum_n^N \omega_{nt}\Phi_n\left(S_t^k\right) \right)^2$$

After obtaining the gradient of $F_t(\omega)$ for ω_{nt} in the same way as for $G_t(\phi)$, if $N \times N$ matrix C and $N \times 1$ matrix D are defined as follows, the relational expression (2.5.20) can be obtained.

$$\left(C^{(t)}\right)_{ij} := \sum_{k=1}^{K_{MC}} \Phi_i(S_t^k)\Phi_j(S_t^k) \tag{2.5.19}$$
$$\left(D^{(t)}\right)_{i1} := \sum_{k=1}^{K_{MC}} \Phi_i(S_t^k) \left(R_t(S_t, a_t^*, S_{t+\Delta t}) + \gamma \max_{a_{t+\Delta t}\in\mathcal{A}} Q_{t+\Delta t}^*\left(S_{t+\Delta t}, a_{t+\Delta t}\right) \right)$$

$$C^{(t)}\omega_t^* = D^{(t)} \tag{2.5.20}$$

Therefore, the coefficients of expansion of the optimal Q-function $Q_t^*(S_t, a_t^*)$ are

$$\omega_t^* = \left(C^{(t)}\right)^{-1} D^{(t)} \tag{2.5.21}$$

Chapter 3

The optimal action and Q-function in QLBS

In this chapter, we will introduce how to find the optimal action and the optimal Q-function among all sets of basis functions. The experiment in chapter 4 will be theoretically supported by the findings in this chapter.

3.1 The optimal action

The inverse matrix of $A^{(t)}$ is utilized in (2.5.16), however $A^{(t)}$ is not necessarily invertible.

Lemma 3.1.1. *Let* $\left(\Delta\hat{S}_t^k\right)^2 > 0$ *for all* $k = 1, 2, \cdots, K_{MC}$.
Then, $A^{(t)}$ *is non-singular if and only if*

$$\begin{bmatrix} \Phi_1\left(S_t^1\right) \\ \Phi_1\left(S_t^2\right) \\ \vdots \\ \Phi_1\left(S_t^{K_{MC}}\right) \end{bmatrix}, \begin{bmatrix} \Phi_2\left(S_t^1\right) \\ \Phi_2\left(S_t^2\right) \\ \vdots \\ \Phi_2\left(S_t^{K_{MC}}\right) \end{bmatrix}, \ldots, and \begin{bmatrix} \Phi_N\left(S_t^1\right) \\ \Phi_N\left(S_t^2\right) \\ \vdots \\ \Phi_N\left(S_t^{K_{MC}}\right) \end{bmatrix} \tag{3.1.1}$$

are linearly independent.

Proof. For all $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\begin{aligned}\mathbb{X}^T A^{(t)} \mathbb{X} &= \sum_i^N x_i \sum_j^N x_j \sum_k^{K_{MC}} \Phi_i\left(S_t^k\right) \Phi_j\left(S_t^k\right) \left(\Delta \hat{S}_t^k\right)^2 \\ &= \sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 \left(\Delta \hat{S}_t^k\right)^2 \geq 0\end{aligned} \tag{3.1.2}$$

If (3.1.1) are linearly independent, then,
for all $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 \left(\Delta \hat{S}_t^k\right)^2 > 0,$$

then $A^{(t)}$ is positive definite. So it has only positive eigenvalues, and is non-singular.
If (3.1.1) are *not* linearly independent, then,
for some $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 \left(\Delta \hat{S}_t^k\right)^2 = 0$$

Since $A^{(t)}$ is symmetric, it is orthogonally diagonalizable, say $A^{(t)} = Q^T D Q$. Because $A^{(t)}$ is positive semi-definite, all of its eigenvalues are non-negative. And with $0 \neq \mathbb{X} \in \mathbb{R}^N$, $Q\mathbb{X} \neq 0$. So $0 = \mathbb{X}^T A^{(t)} \mathbb{X} = \mathbb{X}^T Q^T D Q \mathbb{X}$. Hence at least one of its eigenvalue is zero. It means that $A^{(t)}$ is singular. □

The lemma 3.1.1 has an $\left(\Delta \hat{S}_t^k\right)^2 > 0$ condition. This is natural when S_t is a geometric Brownian motion. Because, if $\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$, then

$$\int_t^{t+\Delta t} \frac{dS_t}{S_t} = \mu \Delta t + \sigma \left(W_{t+\Delta t} - W_t\right) \tag{3.1.3}$$

By Ito's formula,

$$d\left(\ln S_t\right) = \frac{1}{S_t} dS_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\left(dS_t\right)^2 = \frac{dS_t}{S_t} - \frac{1}{2S_t^2}\sigma^2 S_t^2 dt$$

So,

$$\frac{dS_t}{S_t} = d\left(\ln S_t\right) + \frac{1}{2}\sigma^2 dt \tag{3.1.4}$$

The following expression can be obtained by combining (3.1.4) and (3.1.3).

$$\ln S_{t+\Delta t} - \ln S_t + \frac{1}{2}\sigma^2 \Delta t = \mu \Delta t + \sigma\left(W_{t+\Delta t} - W_t\right)$$

$$\ln S_{t+\Delta t} = \ln S_t + \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\left(W_{t+\Delta t} - W_t\right)$$

Therefore,

$$\ln S_{t+\Delta t} \sim \mathcal{N}(\ln S_t + (\mu - \frac{1}{2}\sigma^2)\Delta t, \sigma^2 \Delta t) \tag{3.1.5}$$

From (3.1.5), $\left(\Delta \hat{S}_t^k\right)^2 > 0$ with probability 1.

In other cases, it is necessary to check that $\left(\Delta \hat{S}_t^k\right)^2 > 0$ has been met. For the sake of convenience, it is assumed in this chapter that $\{S_t^k\}$ meet the requirement.

Theorem 3.1.2. *Let time t be fixed. Assume that K_{MC} stock price paths and a set of basis functions $\{\Phi_n\}_{n=1}^N$ are given and that $S_t^i \neq S_t^j$ if $i \neq j$. Then there is a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ that is able to get a reward $-G_t(\phi)$ (see 2.5.11)) greater than or equal to any reward of the set of basis functions $\{\Phi_n\}_{n=1}^N$.*

Proof. For $n = 1, 2, \ldots, N$, define $\widehat{\Phi}_n$ as $\left[\Phi_n(S_t^1), \Phi_n(S_t^2), \cdots, \Phi_n(S_t^{K_{MC}})\right]$. By linear algebra theory, there is a linearly independent subset $\{\widehat{\Phi}'\}$ of $\{\widehat{\Phi}_n\}$ such that $\mathrm{Span}(\{\widehat{\Phi}'\}) = \mathrm{Span}(\{\widehat{\Phi}_n\})$. And then, $\{\widehat{\Phi}'\}$ can be expanded into a basis $\{\widehat{\Phi}_n^*\}$ of $\mathbb{R}^{K_{MC}}$.

Since $S_t^i \neq S_t^j$ when $i \neq j$, we can choose a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ such that has the value $\widehat{\Phi}_n^*$ for $\{S_t^k\}_{k=1}^{K_{MC}}$, that is, $\Psi_n(S_t^k) = k$ th element of $\widehat{\Phi}_n^*$ for all $k = 1, 2, \cdots, K_{MC}$ and for all $n = 1, 2, \cdots, K_{MC}$.

Now, we will show that $\{\Psi_n\}_{n=1}^{K_{MC}}$ is the set of basis functions we are looking for. As looking at (2.5.10) and (2.5.11), we can find out that the Span($\{\widehat{\Phi}_n\}$) determine the range of the reward $-G_t(\phi)$ for $\{\Phi_n\}_{n=1}^{N}$. In other words, the wider Span is, the more range of reward can be obtained. From the fact that Span($\{\widehat{\Phi}_n\}$) = Span($\{\widehat{\Phi}'\}$) $\subseteq$ Span($\{\widehat{\Phi}_n^*\}$), we can get a reward greater than or equal to any reward of the set of basis functions $\{\Phi_n\}_{n=1}^{N}$ with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$. □

In the proof above, there is a part that says 'we can choose a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ such that has the value $\widehat{\Phi}_n^*$ for $\{S_t^k\}_{k=1}^{K_{MC}}$, that is, $\Psi_n(S_t^k) = k$ th element of $\widehat{\Phi}_n^*$ for all $k = 1, 2, \cdots, K_{MC}$ and for all $n = 1, 2, \cdots, K_{MC}$'. Is it really so? Let's consider L^2 space. $\{\Psi_n\}_{n=1}^{K_{MC}}$ may be generated via *spline interpolation* (see Ref. [7]), and since the vectors $\{\left[\Psi_n(S_t^1), \Psi_n(S_t^2), \cdots, \Psi_n(S_t^{K_{MC}})\right]\}_{n=1}^{K_{MC}}$ are linearly independent, $\sum_{n=1}^{K_{MC}} c_n \Psi_n$ cannot be a zero function, where $\{c_n\}$ are constants. Since $\sum_{n=1}^{K_{MC}} c_n \Psi_n$ is continuous and does have a non-zero value, $||\sum_{n=1}^{K_{MC}} c_n \Psi_n||_2 \neq 0$. Therefore, $\{\Psi_n\}_{n=1}^{K_{MC}}$ is linearly independent in L^2.

The premise that $S_t^i \neq S_t^j$ if $i \neq j$ is necessary for the previous theorem 3.1.2. The requirement is met when S_t is a geometric Brownian motion, as can be seen by deriving the equation (3.1.6) below from (3.1.5).

$$\ln S_t \sim \mathcal{N}(\ln S_0 + (\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t) \tag{3.1.6}$$

In other instances, we will first discuss the optimal action among all sets of basis functions for the case that $S_t^i \neq S_t^j$ if $i \neq j$, and then discuss it for the case that $S_t^i = S_t^j$ for some $i \neq j$.

Theorem 3.1.3. *Let time t be fixed and let $\{\Psi_n\}_{n=1}^{K_{MC}}$ be the set of basis functions that we get from theorem 3.1.2. Then ,with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, there is the only one point ξ such that $-G_t(\xi)$ (see (2.5.11)) is the unique global maximum reward of $-G_t(\phi)$.*

Proof. Since the function $G_t(\phi)$ (2.5.11) is differentiable, the gradient of $G_t(\xi)$ must be zero if ξ is to be a local extremum point. By lemma 3.1.1 and the

definition of $\{\Psi_n\}_{n=1}^{K_{MC}}$, $A^{(t)}$ is non-singular. So from (2.5.12) and (2.5.14), we know that ,with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, there is the only one point ξ such that the gradient of $G_t(\xi)$ is zero.
By the second-order Taylor expansion, for any $h \in \mathbb{R}^{K_{MC}}$, there is a real number α between 0 and 1 such that

$$G_t(\phi + h) = G_t(\phi) + \sum_{i=1}^{K_{MC}} \frac{\partial G_t}{\partial \phi_i}(\phi)\, h_i + \frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 G_t}{\partial \phi_j \partial \phi_i}(\phi + \alpha h)\, h_i h_j$$

where h_i means the i th element of h. In particular, when ϕ is ξ, it becomes

$$G_t(\xi + h) = G_t(\xi) + +\frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 G_t}{\partial \phi_j \partial \phi_i}(\xi + \alpha h)\, h_i h_j$$

From equation (2.5.12),

$$\frac{\partial^2 G_t}{\partial \phi_j \partial \phi_i}(\xi + \alpha h) = 2\gamma\lambda \sum_{k=1}^{K_{MC}} \Phi_j(S_t^k)\Phi_i(S_t^k) \left(\Delta \hat{S}_t^k\right)^2 = 2\gamma\lambda \left(A^{(t)}\right)_{ji}$$

Then,

$$\frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 G_t}{\partial \phi_j \partial \phi_i}(\xi + \alpha h)\, h_i h_j = \frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} 2\gamma\lambda \left(A^{(t)}\right)_{ji} h_i h_j = \gamma\lambda\, h^T A^{(t)} h$$

By the fact that $A^{(t)}$ is non-singular, (3.1.2), and lemma 3.1.1,

$$h^T A^{(t)} h > 0 \text{ for all } 0 \neq h \in \mathbb{R}^{K_{MC}}$$

hence $G_t(\xi + h) > G_t(\xi)$, that is to say $G_t(\xi)$ is a global minimum. As mentioned above, it is the unique one. Thus, the reward $-G_t(\xi)$ is the unique global maximum of $-G_t(\phi)$ with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$. □

Theorem 3.1.4. *Let time t be fixed and let $\{\Psi_n\}_{n=1}^{K_{MC}}$ be the set of basis functions that we get from theorem 3.1.2. Assume that an action function $a_t^*(\cdot)$, with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, gets the maximum reward*

$-G_t(\xi)$. *Then, for* $k = 1, 2, \ldots, K_{MC}$, *the action function has values below;*

$$a_t^*(S_t^k) = \frac{\widehat{\Pi}_{t+\Delta t}^k}{\Delta \widehat{S}_t^k} + \frac{1}{2\gamma\lambda} \frac{\Delta S_t^k}{(\Delta \widehat{S}_t^k)^2}$$

Proof. Let $\left[\Psi\right]$ be

$$\begin{bmatrix} \Psi_1(S_t^1)\Delta\widehat{S}_t^1 & \Psi_1(S_t^2)\Delta\widehat{S}_t^2 & \ldots & \Psi_1(S_t^{K_{MC}})\Delta\widehat{S}_t^{K_{MC}} \\ \Psi_2(S_t^1)\Delta\widehat{S}_t^1 & \Psi_2(S_t^2)\Delta\widehat{S}_t^2 & \ldots & \Psi_2(S_t^{K_{MC}})\Delta\widehat{S}_t^{K_{MC}} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{K_{MC}}(S_t^1)\Delta\widehat{S}_t^1 & \Psi_{K_{MC}}(S_t^2)\Delta\widehat{S}_t^2 & \ldots & \Psi_{K_{MC}}(S_t^{K_{MC}})\Delta\widehat{S}_t^{K_{MC}} \end{bmatrix}$$

From (2.5.14), $\left[A_{ij}^{(t)}\right] = [\Psi]\,[\Psi]^T$, where T denote the transpose symbol. By the definition of $\left[\Psi\right]$ from theorem 3.1.2, it is invertible. So is $\left[A_{ij}^{(t)}\right]$. Hence the equation (2.5.16) has a solution, that is to say $\xi = [\xi_1, \xi_2, \ldots, \xi_{K_{MC}}]^T = \left(A^{(t)}\right)^{-1} B^{(t)}$ by theorem 3.1.3. Now, we will calculate the action values that have the maximum reward.

$$
\begin{bmatrix} a_t^*(S_t^1) \\ a_t^*(S_t^2) \\ \vdots \\ a_t^*(S_t^{K_{MC}}) \end{bmatrix} = \begin{bmatrix} \Psi_1(S_t^1) & \Psi_2(S_t^1) & \dots & \Psi_{K_{MC}}(S_t^1) \\ \Psi_1(S_t^2) & \Psi_2(S_t^2) & \dots & \Psi_{K_{MC}}(S_t^2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(S_t^{K_{MC}}) & \Psi_2(S_t^{K_{MC}}) & \dots & \Psi_{K_{MC}}(S_t^{K_{MC}}) \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{K_{MC}} \end{bmatrix}
$$

$$
= \begin{bmatrix} \Psi_1(S_t^1) & \Psi_2(S_t^1) & \dots & \Psi_{K_{MC}}(S_t^1) \\ \Psi_1(S_t^2) & \Psi_2(S_t^2) & \dots & \Psi_{K_{MC}}(S_t^2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(S_t^{K_{MC}}) & \Psi_2(S_t^{K_{MC}}) & \dots & \Psi_{K_{MC}}(S_t^{K_{MC}}) \end{bmatrix} \left[A^{(t)}\right]^{-1} \left[B^{(t)}\right]
$$

$$
= diag\left[\left(\Delta\widehat{S}_t^1\right)^{-1}, \left(\Delta\widehat{S}_t^2\right)^{-1}, \dots, \left(\Delta\widehat{S}_t^{K_{MC}}\right)^{-1}\right] \left[\Psi\right]^T \left[A^{(t)}\right]^{-1} \left[B^{(t)}\right]
$$

$$
= diag\left[\left(\Delta\widehat{S}_t^1\right)^{-1}, \left(\Delta\widehat{S}_t^2\right)^{-1}, \dots, \left(\Delta\widehat{S}_t^{K_{MC}}\right)^{-1}\right] \left[\Psi\right]^T \left[\Psi\right]^{-T} \left[\Psi\right]^{-1} \left[B^{(t)}\right]
$$

$$
= diag\left[\left(\Delta\widehat{S}_t^1\right)^{-1}, \dots, \left(\Delta\widehat{S}_t^{K_{MC}}\right)^{-1}\right] \left[\Psi\right]^{-1} \left[\Psi\right] \begin{bmatrix} \widehat{\Pi}_{t+\Delta t}^1 + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^1}{\Delta\widehat{S}_t^1} \\ \widehat{\Pi}_{t+\Delta t}^2 + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^2}{\Delta\widehat{S}_t^2} \\ \vdots \\ \widehat{\Pi}_{t+\Delta t}^{K_{MC}} + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^{K_{MC}}}{\Delta\widehat{S}_t^{K_{MC}}} \end{bmatrix}
$$

$$
= \begin{bmatrix} \frac{\widehat{\Pi}_{t+\Delta t}^1}{\Delta\widehat{S}_t^1} + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^1}{(\Delta\widehat{S}_t^1)^2} \\ \frac{\widehat{\Pi}_{t+\Delta t}^2}{\Delta\widehat{S}_t^2} + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^2}{(\Delta\widehat{S}_t^2)^2} \\ \vdots \\ \frac{\widehat{\Pi}_{t+\Delta t}^{K_{MC}}}{\Delta\widehat{S}_t^{K_{MC}}} + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^{K_{MC}}}{(\Delta\widehat{S}_t^{K_{MC}})^2} \end{bmatrix}
$$

□

Theorem 3.1.5. *In QLBS model, there is no set of basis functions that is able to have greater reward than that of an action which has the same values on* $\{S_t^k\}_{k=1}^{K_{MC}}$ *as* $a_t^*(S_t^k)$ *of theorem 3.1.4.*

Proof. Let $\{\Phi_n\}$ be an arbitrary set of basis functions. Then, by theorem 3.1.2, there is a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ that is able to get reward

greater than or equal to any of the set of basis functions $\{\Phi_n\}$. So any reward of $\{\Phi_n\}$ can not be greater than the maximum reward of $\{\Psi_n\}_{n=1}^{K_{MC}}$. By theorem 3.1.4, when $-G_t(\phi)$ has the maximum reward with $\{\Psi_n\}_{n=1}^{K_{MC}}$, the action function has the values $a_t^*(S_t^k)$ of theorem 3.1.4, for $k = 1, 2, \ldots, K_{MC}$. $\square$

Now let's discuss the case that $S_t^i = S_t^j$ for some $i \neq j$. To apply the case to the procedure for the case that $S_t^i \neq S_t^j$ if $i \neq j$, it is necessary to slightly modify $G_t(\phi)$. Let's reindex as follows when there are K'_{MC} distinct values among $\{S_t^k\}_{k=1}^{K_{MC}}$. The $\{S_t^{lm}\}$ that reindex $\{S_t^k\}_{k=1}^{K_{MC}}$ contains two indices: l, a group of $\{S_t^k\}_{k=1}^{K_{MC}}$ that consists of K'_{MC} distinct values, and m, a group of $\{S_t^k\}_{k=1}^{K_{MC}}$ that overlap each other for each l. Let M_l denote the total number of m for each l. Then (2.5.11) can be expressed as:

$$
\begin{aligned}
G_t(\phi) &= \sum_{k=1}^{K_{MC}} \left(-\sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^k\right) \Delta S_t^k + \gamma\lambda \left(\hat{\Pi}_{t+\Delta t}^k - \sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^k\right) \Delta \hat{S}_t^k \right)^2 \right) \\
&= \sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \Bigg(-\sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^{lm}\right) \Delta S_t^{lm} \\
&\qquad\qquad + \gamma\lambda \left(\hat{\Pi}_{t+\Delta t}^{lm} - \sum_{n}^{N} \phi_{nt} \Phi_n \left(S_t^{lm}\right) \Delta \hat{S}_t^{lm} \right)^2 \Bigg)
\end{aligned}
$$

This also allows equation (2.5.13) to be expressed as:

$$
\begin{aligned}
&\sum_{k=1}^{K_{MC}} \left(\sum_{n}^{N} \phi_{nt} \Phi_n(S_t^k) \Phi_i(S_t^k) \left(\Delta \hat{S}_t^k\right)^2 \right) \\
&\qquad\qquad = \sum_{k=1}^{K_{MC}} \left(\hat{\Pi}_{t+\Delta t}^k \Phi_i(S_t^k) \Delta \hat{S}_t^k + \frac{1}{2\gamma\lambda} \Phi_i(S_t^k) \Delta S_t^k \right) \\
&\sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \left(\sum_{n}^{N} \phi_{nt} \Phi_n(S_t^{lm}) \Phi_i(S_t^{lm}) \left(\Delta \hat{S}_t^{lm}\right)^2 \right) \\
&\qquad\qquad = \sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \left(\hat{\Pi}_{t+\Delta t}^{lm} \Phi_i(S_t^{lm}) \Delta \hat{S}_t^{lm} + \frac{1}{2\gamma\lambda} \Phi_i(S_t^{lm}) \Delta S_t^{lm} \right)
\end{aligned}
$$

And equation (2.5.14) is

$$
\begin{aligned}
\left(A^{(t)}\right)_{ij} &:= \sum_{k=1}^{K_{MC}} \Phi_i(S_t^k)\Phi_j(S_t^k)\left(\Delta\hat{S}_t^k\right)^2 \\
&= \sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \Phi_i(S_t^{lm})\Phi_j(S_t^{lm})\left(\Delta\hat{S}_t^{lm}\right)^2 \\
&= \sum_{l=1}^{K'_{MC}} \Phi_i(S_t^{l1})\Phi_j(S_t^{l1}) \sum_{m=1}^{M_l}\left(\Delta\hat{S}_t^{lm}\right)^2 \\
\left(B^{(t)}\right)_{i1} &:= \sum_{k=1}^{K_{MC}} \left(\hat{\Pi}_{t+\Delta t}^k \Phi_i(S_t^k)\Delta\hat{S}_t^k + \frac{1}{2\gamma\lambda}\Phi_i(S_t^k)\Delta S_t^k\right) \\
&= \sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \left(\hat{\Pi}_{t+\Delta t}^{lm} \Phi_i(S_t^{lm})\Delta\hat{S}_t^{lm} + \frac{1}{2\gamma\lambda}\Phi_i(S_t^{lm})\Delta S_t^{lm}\right) \\
&= \sum_{l=1}^{K'_{MC}} \Phi_i(S_t^{l1}) \sum_{m=1}^{M_l} \left(\hat{\Pi}_{t+\Delta t}^{lm} \Delta\hat{S}_t^{lm} + \frac{1}{2\gamma\lambda}\Delta S_t^{lm}\right)
\end{aligned}
$$

Now, since $S_t^{lm} \neq S_t^{l'm'}$ if $l \neq l'$, the method for the case that $S_t^i \neq S_t^j$ if $i \neq j$ can be applied. Therefore, we can see that

$$
a_t^*\left(S_t^{l1}\right) = \frac{\sum_{m=1}^{M_l}\left(\hat{\Pi}_{t+\Delta t}^{lm}\Delta\hat{S}_t^{lm} + \frac{1}{2\gamma\lambda}\Delta S_t^{lm}\right)}{\sum_{m=1}^{M_l}\left(\Delta\hat{S}_t^{lm}\right)^2}, \qquad l = 1, 2, \ldots, K'_{MC}
$$

is the optimal action we are looking for.

3.2 The optimal Q-function

The process of finding the optimal Q-function for all sets of basis functions is essentially the same as the process for finding the optimal action for all sets of basis functions.

Lemma 3.2.1. *$C^{(t)}$ is non-singular if and only if*

$$\begin{bmatrix} \Phi_1\left(S_t^1\right) \\ \Phi_1\left(S_t^2\right) \\ \vdots \\ \Phi_1\left(S_t^{K_{MC}}\right) \end{bmatrix}, \begin{bmatrix} \Phi_2\left(S_t^1\right) \\ \Phi_2\left(S_t^2\right) \\ \vdots \\ \Phi_2\left(S_t^{K_{MC}}\right) \end{bmatrix}, \ldots, and \begin{bmatrix} \Phi_N\left(S_t^1\right) \\ \Phi_N\left(S_t^2\right) \\ \vdots \\ \Phi_N\left(S_t^{K_{MC}}\right) \end{bmatrix} \tag{3.2.1}$$

are linearly independent.

Proof. For all $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\begin{aligned} \mathbb{X}^T C^{(t)} \mathbb{X} &= \sum_i^N x_i \sum_j^N x_j \sum_k^{K_{MC}} \Phi_i\left(S_t^k\right) \Phi_j\left(S_t^k\right) \\ &= \sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 \geq 0 \end{aligned} \tag{3.2.2}$$

If (3.2.1) are linearly independent, then,
for all $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 > 0,$$

then $C^{(t)}$ is positive definite. So it has only positive eigenvalues, and is non-singular.
If (3.1.1) are *not* linearly independent, then,
for some $0 \neq (x_1, x_2, \ldots, x_N) = \mathbb{X} \in \mathbb{R}^N$,

$$\sum_k^{K_{MC}} \left(x_1 \Phi_1(S_t^k) + x_2 \Phi_2(S_t^k) + \ldots + x_N \Phi_N(S_t^k)\right)^2 \left(\Delta \hat{S}_t^k\right)^2 = 0$$

Since $C^{(t)}$ is symmetric, it is orthogonally diagonalizable, say $C^{(t)} = Q^T D Q$. Because $C^{(t)}$ is positive semi-definite, all of its eigenvalues are non-negative. And with $0 \neq \mathbb{X} \in \mathbb{R}^N$, $Q\mathbb{X} \neq 0$. So $0 = \mathbb{X}^T C^{(t)} \mathbb{X} = \mathbb{X}^T Q^T D Q \mathbb{X}$. Hence at least one of its eigenvalue is zero. It means that $C^{(t)}$ is singular.

□

Theorem 3.2.2. *Let time t be fixed. Assume that K_{MC} stock price paths and a set of basis functions $\{\Phi_n\}_{n=1}^{N}$ are given and that $S_t^i \neq S_t^j$ if $i \neq j$. Then there is a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ that is able to get a squared error $F_t(\omega)$ (see (2.5.18)) less than or equal to any squared error of the set of basis functions $\{\Phi_n\}_{n=1}^{N}$.*

Proof. For $n = 1, 2, \ldots, N$, define $\widehat{\Phi}_n$ as $\left[\Phi_n(S_t^1), \Phi_n(S_t^2), \cdots, \Phi_n(S_t^{K_{MC}})\right]$. By linear algebra theory, there is a linearly independent subset $\{\widehat{\Phi}'\}$ of $\{\widehat{\Phi}_n\}$ such that $\mathrm{Span}(\{\widehat{\Phi}'\}) = \mathrm{Span}(\{\widehat{\Phi}_n\})$. And then, $\{\widehat{\Phi}'\}$ can be expanded into a basis $\{\widehat{\Phi}_n^*\}$ of $\mathbb{R}^{K_{MC}}$.
Since $S_t^i \neq S_t^j$ when $i \neq j$, we can choose a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ such that has the value $\widehat{\Phi}_n^*$ for $\{S_t^k\}_{k=1}^{K_{MC}}$, that is, $\Psi_n(S_t^k) = k$ th element of $\widehat{\Phi}_n^*$ for all $k = 1, 2, \cdots, K_{MC}$ and for all $n = 1, 2, \cdots, K_{MC}$.
Now, we will show that $\{\Psi_n\}_{n=1}^{K_{MC}}$ is the set of basis functions we are looking for. As looking at (2.5.10) and (2.5.18), we can find out that the $\mathrm{Span}(\{\widehat{\Phi}_n\})$ determines the range of the squared error $F_t(\omega)$ for $\{\Phi_n\}_{n=1}^{N}$. In other words, the wider Span is, the more range of reward can be obtained. From the fact that $\mathrm{Span}(\{\widehat{\Phi}_n\}) = \mathrm{Span}(\{\widehat{\Phi}'\}) \subseteq \mathrm{Span}(\{\widehat{\Phi}_n^*\})$, we can get a squared error less than or equal to any squared error of the set of basis functions $\{\Phi_n\}_{n=1}^{N}$ with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$. □

Theorem 3.2.3. *Let time t be fixed and let $\{\Psi_n\}_{n=1}^{K_{MC}}$ be the set of basis functions that we get from theorem 3.2.2. Then ,with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, there is the only one point ξ such that $F_t(\xi)$ (see (2.5.18)) is the unique global minimum squared error of $F_t(\omega)$.*

Proof. Since the function $F_t(\omega)$ (2.5.18) is differentiable, the gradient of $F_t(\xi)$ must be zero if ξ is to be a local extremum point. By lemma 3.2.1 and the definition of $\{\Psi_n\}_{n=1}^{K_{MC}}$, $C^{(t)}$ is non-singular. So from (2.5.19), we know that ,with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, there is the only one point ξ such that the gradient of $F_t(\xi)$ is zero.

By the second-order Taylor expansion, for any $h \in \mathbb{R}^{K_{MC}}$, there is a real number α between 0 and 1 such that

$$F_t(\omega + h) = F_t(\omega) + \sum_{i=1}^{K_{MC}} \frac{\partial F_t}{\partial \omega_i}(\omega)\, h_i + \frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 F_t}{\partial \omega_j \partial \omega_i}(\omega + \alpha h)\, h_i h_j$$

where h_i means the i th element of h. In particular, when ω is ξ, it becomes

$$F_t(\xi + h) = F_t(\xi) + \frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 F_t}{\partial \omega_j \partial \omega_i}(\xi + \alpha h)\, h_i h_j$$

From equation (2.5.18),

$$\frac{\partial^2 F_t}{\partial \omega_j \partial \omega_i}(\xi + \alpha h) = 2 \sum_{k=1}^{K_{MC}} \Phi_j(S_t^k)\Phi_i(S_t^k) = 2\left(C^{(t)}\right)_{ji}$$

Then,

$$\frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} \frac{\partial^2 F_t}{\partial \omega_j \partial \omega_i}(\xi + \alpha h)\, h_i h_j = \frac{1}{2!} \sum_{j=1}^{K_{MC}} \sum_{i=1}^{K_{MC}} 2\left(C^{(t)}\right)_{ji} h_i h_j = h^T C^{(t)} h$$

By the fact that $C^{(t)}$ is non-singular, (3.2.2), and lemma 3.2.1,

$$h^T C^{(t)} h > 0 \text{ for all } 0 \neq h \in \mathbb{R}^{K_{MC}}$$

hence $F_t(\xi + h) > F_t(\xi)$, that is to say $F_t(\xi)$ is a global minimum. As mentioned above, it is the unique one. Thus, the squared error $F_t(\xi)$ is the unique global minimum of $F_t(\omega)$ with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$. □

Theorem 3.2.4. *Let time t be fixed and let $\{\Psi_n\}_{n=1}^{K_{MC}}$ be the set of basis functions that we get from theorem 3.2.2. Assume that an Q-function $Q_t^*(\cdot, a_t^*)$, with the set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$, gets the minimum squared error $F_t(\xi)$. Then, for $k = 1, 2, \ldots, K_{MC}$, the Q-function has values below;*

$$Q_t^*\left(S_t^k, a_t^*\right) = R_t(S_t^k, a_t^*, S_{t+\Delta t}^k) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^*\left(S_{t+\Delta t}^k, a_{t+\Delta t}\right)$$

Proof. Let $\left[\Psi\right]$ be

$$\begin{bmatrix} \Psi_1(S_t^1) & \Psi_1(S_t^2) & \dots & \Psi_1(S_t^{K_{MC}}) \\ \Psi_2(S_t^1) & \Psi_2(S_t^2) & \dots & \Psi_2(S_t^{K_{MC}}) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{K_{MC}}(S_t^1) & \Psi_{K_{MC}}(S_t^2) & \dots & \Psi_{K_{MC}}(S_t^{K_{MC}}) \end{bmatrix}$$

From (2.5.19), $\left[C_{ij}^{(t)}\right] = [\Psi]\,[\Psi]^T$, where T denote the transpose symbol. By the definition of $\left[\Psi\right]$ from theorem 3.2.2, it is invertible. So is $\left[C_{ij}^{(t)}\right]$. Hence the equation (2.5.21) has a solution, that is to say $\xi = [\xi_1, \xi_2, \dots, \xi_{K_{MC}}]^T = \left(C^{(t)}\right)^{-1} D^{(t)}$ by theorem 3.2.3. Now, we will calculate the Q-function values that have the minimum squared error.

$$
\begin{aligned}
&\begin{bmatrix} Q_t^*(S_t^1, a_t^*) \\ Q_t^*(S_t^2, a_t^*) \\ \vdots \\ Q_t^*(S_t^{K_{MC}}, a_t^*) \end{bmatrix} \\
&= \begin{bmatrix} \Psi_1(S_t^1) & \Psi_2(S_t^1) & \dots & \Psi_{K_{MC}}(S_t^1) \\ \Psi_1(S_t^2) & \Psi_2(S_t^2) & \dots & \Psi_{K_{MC}}(S_t^2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(S_t^{K_{MC}}) & \Psi_2(S_t^{K_{MC}}) & \dots & \Psi_{K_{MC}}(S_t^{K_{MC}}) \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{K_{MC}} \end{bmatrix} \\
&= \begin{bmatrix} \Psi_1(S_t^1) & \Psi_2(S_t^1) & \dots & \Psi_{K_{MC}}(S_t^1) \\ \Psi_1(S_t^2) & \Psi_2(S_t^2) & \dots & \Psi_{K_{MC}}(S_t^2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(S_t^{K_{MC}}) & \Psi_2(S_t^{K_{MC}}) & \dots & \Psi_{K_{MC}}(S_t^{K_{MC}}) \end{bmatrix} \left[C^{(t)}\right]^{-1} \left[D^{(t)}\right] \\
&= \left[\Psi\right]^T \left[C^{(t)}\right]^{-1} \left[D^{(t)}\right] \\
&= \left[\Psi\right]^T \left[\Psi\right]^{-T} \left[\Psi\right]^{-1} \left[D^{(t)}\right] \\
&= \left[\Psi\right]^{-1} \left[\Psi\right] \begin{bmatrix} R_t(S_t^1, a_t^*, S_{t+\Delta t}^1) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^1, a_{t+\Delta t}\right) \\ R_t(S_t^2, a_t^*, S_{t+\Delta t}^2) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^2, a_{t+\Delta t}\right) \\ \vdots \\ R_t(S_t^{K_{MC}}, a_t^*, S_{t+\Delta t}^{K_{MC}}) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^{K_{MC}}, a_{t+\Delta t}\right) \end{bmatrix} \\
&= \begin{bmatrix} R_t(S_t^1, a_t^*, S_{t+\Delta t}^1) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^1, a_{t+\Delta t}\right) \\ R_t(S_t^2, a_t^*, S_{t+\Delta t}^2) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^2, a_{t+\Delta t}\right) \\ \vdots \\ R_t(S_t^{K_{MC}}, a_t^*, S_{t+\Delta t}^{K_{MC}}) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^{K_{MC}}, a_{t+\Delta t}\right) \end{bmatrix}
\end{aligned}
$$

□

Theorem 3.2.5. *In QLBS model, there is no set of basis functions that is able to have less squared error than that of a Q-function which has the same values on* $\{S_t^k\}_{k=1}^{K_{MC}}$ *as* $Q_t^*\left(S_t, a_t^*\right)$ *of theorem 3.2.4.*

Proof. Let $\{\Phi_n\}$ be an arbitrary set of basis functions. Then, by theorem 3.2.2, there is a set of basis functions $\{\Psi_n\}_{n=1}^{K_{MC}}$ that is able to get a squared error less than or equal to any of the set of basis functions $\{\Phi_n\}$. So any squared error of $\{\Phi_n\}$ can not be less than the minimum squared error of $\{\Psi_n\}_{n=1}^{K_{MC}}$. By theorem 3.2.4, when $F_t(\omega)$ has the minimum squared error with $\{\Psi_n\}_{n=1}^{K_{MC}}$, the Q-function has the values $Q_t^*\left(S_t^k, a_t^*\right)$ of theorem 3.2.4, for $k = 1, 2, \ldots, K_{MC}$. □

Now let's discuss the case that $S_t^i = S_t^j$ for some $i \neq j$. To apply the case to the procedure for the case that $S_t^i \neq S_t^j$ if $i \neq j$, it is necessary to slightly modify $F_t(\omega)$. Let's reindex as follows when there are K'_{MC} distinct values among $\{S_t^k\}_{k=1}^{K_{MC}}$. The $\{S_t^{lm}\}$ that reindex $\{S_t^k\}_{k=1}^{K_{MC}}$ contains two indices: l, a group of $\{S_t^k\}_{k=1}^{K_{MC}}$ that consists of K'_{MC} distinct values, and m, a group of $\{S_t^k\}_{k=1}^{K_{MC}}$ that overlap each other for each l. Let M_l denote the total number of m for each l. Then (2.5.18) can be expressed as:

$$
\begin{aligned}
&F_t(\omega) = \\
&\sum_{k=1}^{K_{MC}} \left(R_t(S_t^k, a_t^*, S_{t+\Delta t}^k) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^k, a_{t+\Delta t}\right) - \sum_n^N \omega_{nt} \Phi_n \left(S_t^k\right) \right)^2 \\
&= \sum_{l=1}^{K'_{MC}} \sum_{m=1}^{M_l} \Bigg(R_t(S_t^{lm}, a_t^*, S_{t+\Delta t}^{lm}) \\
&\qquad\qquad + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^{lm}, a_{t+\Delta t}\right) - \sum_n^N \omega_{nt} \Phi_n \left(S_t^{lm}\right) \Bigg)^2
\end{aligned}
$$

Now, since $S_t^{lm} \neq S_t^{l'm'}$ if $l \neq l'$, the method for the case that $S_t^i \neq S_t^j$ if $i \neq j$ can be applied. Therefore, we can see that, for $l = 1, 2, \ldots, K'_{MC}$,

$$
Q_t^*(S_t^{l1}, a_t^*) = \frac{1}{M_l} \sum_{m=1}^{M_l} \left(R_t(S_t^{l1}, a_t^*, S_{t+\Delta t}^{lm}) + \gamma \max_{a_{t+\Delta t} \in \mathcal{A}} Q_{t+\Delta t}^* \left(S_{t+\Delta t}^{lm}, a_{t+\Delta t}\right) \right)
$$

is the optimal Q-function we are looking for.

Chapter 4

Experiment : The optimal is not optimal

Igor Halperin in his paper [2] "*We will leave a detailed investigation of empirical behavior of option prices and hedges in this pre-asymptotic regime to a future work, while concentrating in this paper on a mathematical framework.*" said. The pre-asymptotic regime in this context refers to $\Delta t > 0$ and $\lambda > 0$. Let's experiment with the effect of basis functions in this paper and then examine the outcomes of the experiment.

4.1 Experimental Design

The experiment's goal is to determine whether option pricing will improve if we select a set of basis functions that produce greater rewards in $G_t(\phi)$ (2.5.11). Chapter 3 and (2.5.4) provides the experiment's theoretical background. The experiment is structured as follows.

- Consider a geometric Brownian motion $dS_t = \mu S_t dt + \sigma S_t dW_t$ for the stock price S_t.
  The details of the figures are as follows.
  $S_0 = 100,\ \mu = 0.03,\ \sigma = 0.05,\ T = 1$, and $\Delta t = 1/24$.

- Prepare 100 instances of 500 paths data of such a stock price.

- With that data, the results are obtained for each of the two sets of basis functions in the QLBS model.
  One is a B-spline, and the other is a set of basis functions that can achieve the highest reward in $G_t(\phi)$, $a_t^*(S_t^k) = \frac{\widehat{\Pi}_{t+\Delta t}^k}{\Delta \widehat{S}_t^k} + \frac{1}{2\gamma\lambda}\frac{\Delta S_t^k}{(\Delta \widehat{S}_t^k)^2}$, as detailed in Chapter 3.
  As observed in (2.5.4) and the remarks below, the QLBS model can be used to estimate the option pricing of the BSM model if λ is large enough in (2.5.5).

- Therefore, the error rate is defined as follows.

$$\text{Error rate } := \frac{\mathbb{E}_0\left[\Pi_0\right] - \text{BSM value}}{\text{BSM value}} \times 100$$

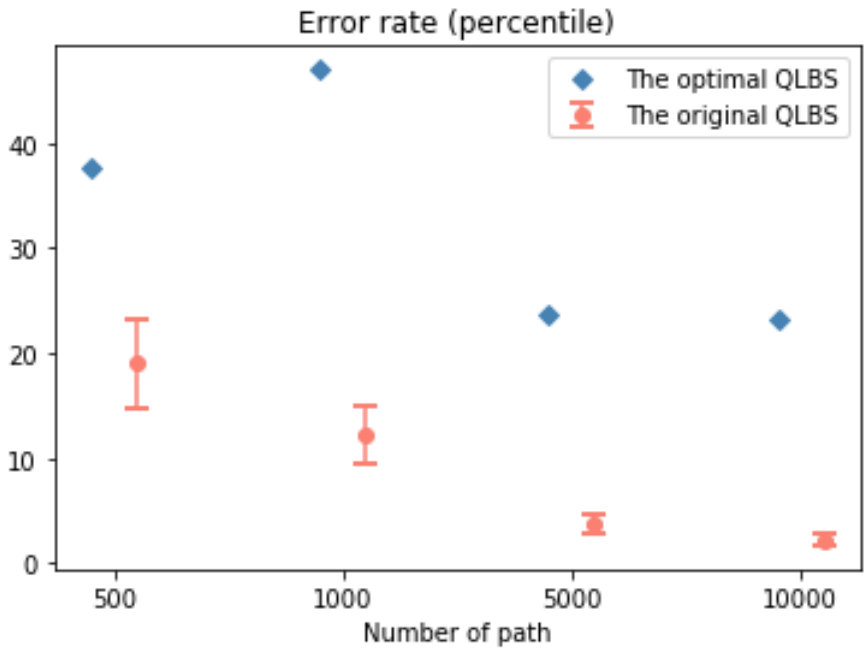
- The mean and standard deviation of the error rate are calculated after 100 iterations, and the results are contrasted.

- In addition to the stock price paths data of 500 bundles, the previous experiment is also conducted using the same approach for stock price paths of 1000, 5000, and 10,000 bundles.

- The experiment's outcomes are then collected and compared.

4.2 Experimental results and analysis

The experimental results were presented in a graph and a table. In the graph, the y-axis is the error rate and the x-axis is the number of stock price paths. The orange circles represent the values of the original QLBS model using B-spline basis functions. The vertical lines above and below the circle show the standard deviation added to and subtracted from the mean. The blue diamond shape indicates the value of the optimal QLBS model using the set of basis functions obtained in Chapter 3. The first column in the table



	mean_original	mean_optimal	std_original	std_optimal
500	19	37.65	4.27	65.47
1000	12.29	46.98	2.71	123.17
5000	3.79	23.76	0.87	29.77
10000	2.22	23.32	0.52	34.73

represents the number of stock price paths. In the remaining columns, the index is in the first row.

Let's analyze the experimental results. First, the original QLBS model demonstrates that as the number of stock price paths rises, the average and standard deviation of the error rate converges to zero. It indicates that option pricing closely tracks the option value predicted by the BSM model. However, in the case of the optimal QLBS model, the average error rate does not decrease, and even the standard deviation does not converge to zero. Option pricing was not carried out with the option value of the BSM model, despite using the same data, and since the standard deviation did not converge, this implies that option pricing was not carried out with any other points. The standard deviation for the optimal QLBS model was too high to display in

the graph.

We currently come to two conclusions.

1. The optimal QLBS model has higher rewards than the original QLBS model. However, the preceding experimental results show that which set of basis functions is used affects option pricing independently of reward.

2. Depending on the choice of a set of basis functions, even with the same quantity of data, the rate at which option pricing will converge varies.

Therefore, choosing a set of basis functions at random, as in the current QLBS model [2], is no longer desirable. We require a theory that will enable us to select a set of basis functions for the QLBS model that is more effective.

Bibliography

[1] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of political economy*, 81(3):637–654, 1973.

[2] Igor Halperin. QLBS: Q-Learner in the Black-Scholes(-Merton) Worlds. *The Journal of derivatives*, 28(1):99–122, 2020.

[3] Igor Halperin. The QLBS Q-Learner Goes NuQLear: Fitted Q Iteration, Inverse RL, and Option Portfolios. *Quantitative finance*, 19(9):1543–1553, 2019.

[4] Andreas J. Grau. Applications of Least-Squares Regressions to Pricing and Hedging of Financial Derivatives. *PhD. thesis, Technische Universität München*, 2007.

[5] Marc Potters, Jean-Philippe Bouchaud, and, Dragan Sestovic. Hedged Monte-Carlo: low variance derivative pricing with objective probabilities. *Physica A.*, 289(3):517–525, 2001.

[6] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction, Second edition. *The MIT Press*, 2018.

[7] Rainer Kress. Numerical analysis. *New York : Springer*, 1998.

국문초록

이 논문은 QLBS model 에서 a set of basis functions 가 제약 없이 선택되는 것이 적절한가에 대해 논의한다. Igor Halperin 은 그의 논문 "*QLBS: Q-Learner in the Black-Scholes(-Merton) Worlds*" 에서 QLBS 라는 discrete-time option hedging and pricing model 을 소개했다. 그는 그 논문에서 특정한 조건 하에 discrete-time 간의 간격이 0 으로 수렴할수록 QLBS model 이 Black-Scholes-Merton model 에 수렴함을 증명하였지만 그 간격이 구체적인 양수일 때의 현상은 future work 로 남겨두었다. 이 논문에서는 QLBS model 에서 설정한 reward 가 a set of basis functions 의 선택에 따라 애초에 기대했던 결과와 다른 option pricing 을 유도할 수 있다는 것을 보일 것이다.

주요어휘: QLBS, 다이나믹 프로그래밍, 옵션 헷징, 옵션 가격결정, 마르코프 결정 과정
학번: 2018-26597

감사의 글

지도해주신 이기암 교수님께 감사드립니다. 면담 시 해주신 말씀을 홀로 되새겨본 적이 많았습니다. 저의 많은 부족한 부분 중 어느 곳을 채워주는 지혜들이었습니다. 감사합니다. 논문 심사를 위해 시간을 내어주신 변순식, 박형빈 교수님께도 감사를 드립니다.

수리과학부 행정실 선생님들께 감사의 말씀을 드립니다. 행정 절차에서 헤맬 때가 많았습니다. 그럴 때마다 귀찮은 내색 한번 없이 매번 친절히 도움을 주셔서 감사했습니다.

연구실 동료, 대학원 동기들 모두 정말 고맙고 각각의 모습이 하나하나 빛나서 같이 지내고 바라보는 내내 행복했습니다.

그리고 나보다 더 소중한 사람. 내 딸.
네가 태어나고 아빠의 세상은 다시 채색되었어.

마지막으로 항상 같이 있지만 고맙다는 말이 가장 인색하게 되는 사람에게 이 기회를 통해 감사함을 전합니다.

나랑 사느라 수고가 많아.
그 생각을 따라 거닐다 보면
네가 안쓰럽고.
내심 고맙고.
자연스레 사랑스럽다.

우리가 만나서 서로의 삶이 겹쳐지고.
나의 삶이 비좁아진다는 것이 너의 온기가 느껴진다는 것과 같아지고.
서로의 삶이 부딪힌다는 것이 서로 토닥여준다는 것과 같아졌어.

벌써 내 삶의 반 이상이 너다.
그것이 굳이 생각해봐야 감사한 줄 아는 일상이 됐네.
고마워.