Total Risk Minimization With Spline Kernel Function

by

Beini Fan

A research paper
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Supervisor: Dr. Yuying Li

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

To solve a quadratic total risk minimization problem, we minimizes the difference between the option payoff and the final value of a self-financing hedging portfolio. To reduce the computational complexity, a method is proposed by Coleman et al. (2008) that a specific function is used to describe the relationship between the holding in the hedging portfolio and the underlying asset price. The spline kernel function with regularization is introduced to model the relationship. We use cross validation to investigate the out-of-sample performance. The results obtained from Monte Carlo simulation indicate that using spline kernel function with regularization could allow more complexity and may lead to a relatively smaller total risk. In addition, the total risk is affected more by the regularization penalty parameter while shows robustness with respect to the number of reference points.

Acknowledgements

I would like to sincerely acknowledge my supervisor, Dr. Yuying Li, whose support is so important that I could not successfully finish this research project without her help. I am deeply grateful for her unfailing guidance and generosity.

Great thanks also go to Dr. Peter Forsyth for reading my paper carefully and providing valuable suggestions for me to make improvements.

I would as well like to thank Dr. Jeff Orchard, Amanda Guderian, Cherisse Mike and other students in Computational Mathematics for their help and concerns.

Dedication

This project report is dedicated to my family.

Table of Contents

List of Tables viii

List of Figures x

1 Introduction 1

2 Risk Minimization Criteria 4

3 Total Risk Minimization Using Spline Kernel With Regularization 11

3.1 Spline Kernel Function 11

3.2 Regularization 14

4 Simulation Results 18

4.1 Simulation Settings 18

4.2 Comparison With Delta Hedging 20

4.2.1 Convergence 20

4.2.2 Normalized P&L 21

4.2.3 Other Risk Measures 23

4.3 Spline Kernel Without Regularization 25

4.3.1 Benchmark 25

4.3.2 Strategy 1 26

4.3.3 Strategy 2 31

4.4 Spline Kernel With Regularization 37
4.4.1 Strategy 3 37
4.4.2 Strategy 4 44
4.4.3 New Placement of Reference Points 47

5 Conclusion **51**
5.1 Technical Issues 52
5.2 Future Work 52

References **53**

List of Tables

4.1 Basic settings for the Monte Carlo simulation 19
4.2 Convergence to the Black Scholes price 20
4.3 Statistics for the P&L of Delta hedging (hedging 24 times) 23
4.4 Statistics for the P&L of Strategy 1 (hedging 24 times) 23
4.5 Statistics for the P&L of Delta hedging (hedging 600 times) 24
4.6 Statistics for the P&L of Strategy 1 (hedging 600 times) 24
4.7 Average total risk for the Benchmark and Strategy 1 with $n = 60$ 27
4.8 Average total cost for the Benchmark and Strategy 1 with $n = 60$ 27
4.9 Average total risk for testing set using Strategy 1 29
4.10 Average total cost for testing set using Strategy 1 29
4.11 VaR and CVaR of total risk for testing set using Strategy 1 30
4.12 VaR and CVaR of total cost for testing set using Strategy 1 31
4.13 Average total risk for the Benchmark and Strategy 2 with $n = 60$ 32
4.14 Average total cost for the Benchmark and Strategy 2 with $n = 60$ 32
4.15 Average total risk for testing set using Strategy 2 34
4.16 Average total cost for testing set using Strategy 2 34
4.17 VaR and CVaR of total risk for testing set using Strategy 2 36
4.18 VaR and CVaR of total cost for testing set using Strategy 2 36
4.19 Average total risk using Strategy 3 with 2D cross validation 41
4.20 Average total cost using Strategy 3 with 2D cross validation 42

4.21 Comparison of average total risk using 2D and 1D cross validation 43
4.22 Average total risk for Strategy 4 with testing data set 46
4.23 Average total cost for Strategy 4 with testing data set 46
4.24 Average total risk for Strategy 3 with new placement of n 49
4.25 Average total cost for Strategy 3 with new placement of n 49

List of Figures

4.1 Frequency versus normalized P&L with $K = 100$ 22
4.2 Average total risk for strategy 1 with cross validation at $K = 90$ 28
4.3 Density plot of total risk and total cost for Strategy 1 30
4.4 Average total risk for strategy 2 with cross validation at $K = 90$ 33
4.5 Density plot of total risk and total cost for Strategy 2 35
4.6 Average total risk for testing set with pairs of λ and n using Strategy 3 . . 38
4.7 Change of total risk with respect to λ for fixed n 40
4.8 Change of total risk with respect to n for fixed λ 41
4.9 Average total risk for testing set using Strategy 4 45
4.10 Average total risk for testing set with pairs of λ and n using Strategy 3-b . 48
4.11 Average total risk for testing set using Strategy 3-b 48

Chapter 1

Introduction

Hedging plays a very important role in risk management. It helps to reduce the risk caused by asset price movements in the financial market, making a portfolio less sensitive to the market changes. For an option, one can hedge it by creating a portfolio containing the underlying asset and a risk-free asset. In the Black-Scholes framework, a method called delta hedging was proposed as a hedging strategy. Delta is the first-order partial derivative of the option value against the underlying asset price, which measures the sensitivity of the option value to the changing of the underlying asset price. Delta hedging uses a self-financing portfolio to exactly replicate the option and hedge the market risk arising from asset price movements. The self-financing portfolio means that the portfolio only needs the initial investment and it does not generate any capital inflow or outflow from the portfolio after the strategy begins.

In the Black-Scholes model, the financial market is complete under a unique risk-neutral probability measure, and the contingent claims are attainable. In addition, the market is arbitrage-free. However, the assumption of the Black-Scholes model is very restrictive. There are a lot of unpredictable factors and restrictions in the real financial market, such as jump risks, transaction limitations, taxes, bid-ask spreads, and so on, which make the completeness impossible to achieve. Since the real market is incomplete, we cannot find a strategy that can hedge the risk of an option completely. What's more, in practice, hedging cannot be done continuously in time to offset perfectly the fluctuations of the underlying asset price. We can only trade the market asset at discrete time points and hedging is also under transaction costs. Therefore, it is difficult to replicate and hedge an option completely in reality.

Many researchers have studied the topic of hedging an option in an incomplete market.

El Karoui and Quenez (1995) tried to price an option by a super-replication method. The idea of this method is to obtain the minimum initial cost of an option with a self-financing hedging strategy such that the final value of the hedging portfolio is always larger than the option payoff. Although the fair value of an option cannot be obtained by the no-arbitrage principle in the incomplete market, the super-replication method succeeds to give a range for the actual value of an option. However, this method has some limitations and we do not pursue it in this project. For instance, the super-replicating strategy for a call option suggested by this method for the Hull-White stochastic model (1987) is to hold the underlying asset (Frey, 1997).

Minimizing a certain measure for the risk of an option is another widely used approach to obtain the value of an option and to get the optimal discretely hedging strategy in the incomplete market. Instead of minimizing the initial cost, this approach aims to obtain the minimum risk which is defined in an appropriate criterion. There are two criteria studied by the researchers, which are the local risk-minimization first studied by Schäl (1994) and the total risk-minimization criterion proposed by Föllmer and Schweizer (1989). The local risk strategy assumes that the final value of the hedging portfolio is equal to the option payoff. It then finds the optimal strategy by minimizing the incremental cost to adjust the portfolio at each hedging time. The local risk-minimization criterion is not self-financing and the optimization problem is solved backwards from the expiry to the beginning. Another approach is to find an optimal hedging strategy that best approximates the option payoff by the final value of a hedging portfolio. For the total risk minimization problem, the optimal strategy is self-financing. It assumes that, except the initial investment, the hedging portfolio does not need any extra funding to adjust the hedging position, which is very different from the local risk minimization problem. In this project, we investigate the total risk minimization problem.

There are two different objective functions studied for the risk minimization problem, which are quadratic functions and piecewise linear functions. Föllmer and Schweizer (1989), Schäl (1994), Schweizer (1995, 2001) studied the quadratic risk-minimization criteria. Coleman et al. (2003) investigate the piecewise linear criterion for local risk minimization and total risk minimization (2008). The quadratic function and piecewise linear function may lead to very different optimal strategies since they penalize the residuals in different ways. The quadratic function penalizes more when the hedging strategy give a bad performance while the piecewise linear function treat all the situations equally. Since the piecewise linear function is computationally more expensive to optimize, Coleman et al. (2008) propose a method to obtain the optimal strategy by using Monte Carlo simulation.

Solving the total risk minimization problem by Monte Carlo simulation introduce a large number of unknowns. In order to reduce the complexity of solving this optimization problem, Coleman et al. (2008) make an assumption that the holding of the risky asset (stock) in the hedging portfolio is determined by a function of the current and past stock prices. Since the value of option is affected mostly by the current and past trading information of the underlying asset, using such a function is reasonable from the financial point of view. By this assumption, Coleman et al. (2008) investigate piecewise linear total risk optimization and assume the relationship between holdings and the asset prices can be described by the cubic spline functions. They generate stock price paths using the Monte Carlo simulation and obtain the best strategy by solving for the parameters in the cubic spline functions.

In this project, we solve the quadratic total risk minimization problem as an illustration. The whole framework can also be applied to the total risk minimization problem with the piecewise linear objective function. Based on the assumption proposed by Coleman et al. (2008), we solve the optimization problem using Monte Carlo simulation. In addition, when obtaining the optimal strategy, we generate both training set and testing set and determine the best parameters by cross validation. However, instead of using the cubic spline function, we treat the holding as a spline kernel function of the underlying asset.

When using the cubic spline function, a finite number of fixed knots is introduced and the cubic spline function only guarantees the continuity and the derivative continuity at those specific knots. Thus, it may cause the discontinuity problem when a data point is not chosen to be a knot. Unlike cubic spline function, the spline kernel function maps the input values into a space with infinite number of knots. Furthermore, it is unnecessary to choose a certain set of knots since the knots are implicit in the spline kernel function, which avoids the explicit placement of spline knots that need to be determined when using cubic spline function.

Since the holding depends nonlinearly on the stock price, Coleman et al. (2012) propose a method that the nonlinear relationship can be approximated by a linear combination of spline kernels. By this method, we assume that the holding of the underlying asset is a linear combination of spline kernel functions evaluated at some reference points. The reference points are chosen from the training set of the underlying asset prices.

Moreover, since the spline kernel function is easier to regularize, we introduce the regularization term to the objective function. Regularization puts constraints on the flexibility of a model, thereby increasing the stability of the model. Using the spline kernel function with regularization, we examine the total risk to see if there is any improvement of the hedging performance.

Chapter 2

Risk Minimization Criteria

In this chapter, we review the literature to show and compare the risk minimization criteria in a mathematical way. For clear notation, we introduce a financial market which is modeled by a filtered probability space $(\Omega, \mathcal{F}, \mathrm{P})$. The time horizon is $[0, T]$ and the discrete hedging times are denoted by $0 = t_0 < t_1 < t_2 < ... < t_M = T$, where $T > 0$ and M is the total number of the discrete hedging times. The filtration $\{\mathcal{F}_j\}_{,j=0,1,...,M}$ corresponds to the time t_j. The stock price process $\{S_j\}_{,j=0,1,...,M}$ is measurable with respect to $\{\mathcal{F}_j\}$, which means that $\mathcal{F}_j$ represents all history information about the stock up to time t_j. It is obvious that $\mathcal{F}_0 = \{\oslash, \Omega\}$ is trivial.

We consider a hedging portfolio V_j containing a risky asset (stock) S_j and a risk-free asset (bond) B_j. At each time point t_j, the discounted stock price is defined by $X_j = \frac{S_j}{B_j}, \forall j = 0,...,M$. At each time t_j, value of the hedging portfolio is

$$V_j = \xi_j X_j + \eta_j \tag{2.1}$$

where stochastic processes ξ_j and η_j are holdings of the stock and the bond respectively.

For all $j = 0, 1, ..., M-1$, we define the change of the discounted stock price between every two hedging times as follows,

$$\Delta X_j = X_{j+1} - X_j, j = 0, 1, ..., M-1.$$

Then the value of the hedging portfolio is changed by $\xi_j \Delta X_j$ due to the change in the discounted stock price before making any portfolio adjustment. We define accumulated gain G_j as the change of the portfolio value caused by the change of the discounted stock price. Therefore, the accumulated gain G_j can be represented by

$$G_j = \xi_j(X_{j+1} - X_j) = \xi_j \Delta X_j, j = 0, 1, ..., M-1 \tag{2.2}$$

and $G_0 = 0$.

The cumulative cost C_j is defined as the cash flow of the portfolio at time t_j after considering the change of the portfolio value from the stock price fluctuation. The cumulative cost has the relationship below with the accumulated gain.

$$C_j = V_j - G_j, j = 0, 1, ..., M-1. \tag{2.3}$$

Assume that we need to hedge a claim, whose payoff at time T is denoted by H and H is $\mathcal{F}_M$-measurable. A market is complete if any claim H is attainable, which means that there exists a self-financing strategy with $V_M = H$ (a.s.). Self-financing indicates that the cumulative cost process $\{C_j\}_{,j=0,1,...,M}$ is constant over time, that is, $C_0 = C_1 = ... = C_{M-1}$. However, in the case of discrete hedging, the financial market is not complete and the hedging strategy is obtained by using some optimality criterion.

One criterion to obtain the optimal strategy is to minimize the local risk of an option. Local risk is defined as the incremental cost of adjusting the hedging portfolio at each time. In this approach, the final value of trading strategy is assumed to be equal to the option payoff, that is $H = V_M$. Then the optimal solution is obtained by minimizing the expected incremental cost at every hedging spot, which leads to the following optimization problem.

$$E((C_{j+1} - C_j)^2 \mid \mathcal{F}_j), 0 \leq j \leq M-1$$

where $E(\cdot \mid \mathcal{F}_j)$ denotes the expected value with respect to the probability measure P and the filtration $\{\mathcal{F}_j\}$.

The holding at each intermediate hedging time is given by solving the optimization problem at that time. The optimal holding at t_{j+1} is served as the condition when solving the optimization problem at t_j. With the end condition that $H = V_M$, the local risk minimization problem can be solved from the maturity to the initial time. In addition, by considering the definition (2.2), and the relationship between the cumulative cost and the accumulated gain shown in the equation (2.3), the objective function for the local risk minimization criterion is as follows,

$$E((C_{j+1} - C_j)^2 \mid \mathcal{F}_j) = E((V_{j+1} - V_j - \xi_j(X_{j+1} - X_j))^2 \mid \mathcal{F}_j).$$

After applying the equation (2.1) to replace V_j with the holdings ξ_j, η_j in the trading portfolio, we can obtain the objective function for the local risk minimization criterion.

$$E((V_{j+1} - V_j - \xi_j(X_{j+1} - X_j))^2 \mid \mathcal{F}_j) = E((X_{j+1}(\xi_{j+1} - \xi_j) + (\eta_{j+1} - \eta_j))^2 \mid \mathcal{F}_j).$$

Schäl (1994) proves the existence of an explicit local risk minimizing strategy when the option payoff H is squared integrable and the discounted stock price X has a bounded mean-variance trade-off, that is:

$$\frac{(E(\Delta X_j \mid \mathcal{F}_j))^2}{Var(\Delta X_j \mid \mathcal{F}_j)} \text{ is P} - \text{a.s. uniformly bounded,} \tag{2.4}$$

where $Var(\cdot \mid \mathcal{F}_j)$ denotes the variance of a random variable with respect to the probability measure P and the filtration $\{\mathcal{F}_j\}$. The analytic solution proposed by Schäl (1994) for the quadratic local risk minimization problem is:

$$\begin{cases} \xi_M^{(l)} = 0, \eta_M^{(l)} = H \\ \xi_j^{(l)} = \frac{Cov(\xi_{j+1}X_{j+1} + \eta_{j+1}, X_{j+1} \mid \mathcal{F}_j)}{Var(X_{j+1} \mid \mathcal{F}_j)}, 0 \le j \le M-1 \\ \eta_j^{(l)} = E((\xi_{j+1} - \xi_j)X_{j+1} + \eta_{j+1} \mid \mathcal{F}_j), 0 \le j \le M-1. \end{cases}$$

where $Cov(\cdot, \cdot \mid \mathcal{F}_j)$ denotes the covariance of two random variables with respect to the probability measure P and the filtration $\{\mathcal{F}_j\}$.

The quadratic total risk minimization criterion aims to minimize the L_2-norm of the difference between option payoff and the hedging portfolio value. In other words, we need to minimize the expectation of squared difference between H and V_M. This strategy is self-financing, which means $C_0 = C_1 = ... = C_{M-1}$. Therefore, the quadratic total risk minimization criterion is to minimize the formula (2.5):

$$E((H - V_M)^2) = E\left((H - V_0 - \sum_{j=0}^{M-1} \xi_j \Delta X_j)^2\right). \tag{2.5}$$

The optimal solution is given by Schweizer (1995) with the condition that the discounted stock price X has a bounded mean-variance trade-off shown in the formula (2.4). The analytic solution is:

$$\begin{cases} V_0^{(t)} = \frac{E(H \prod_{j=0}^{M-1}(1 - \beta_j \Delta X_j))}{E(\prod_{j=0}^{M-1}(1 - \beta_j \Delta X_j))}, \\ \xi_M^{(t)} = 0, \\ \xi_j^{(t)} = \xi_j^{(l)} + \beta_j(V_j^{(l)} - V_0^{(t)} - G_j(\xi^{(t)})) + \gamma_j, 0 \le j \le M-1, \end{cases}$$

where the process β_j and γ_j are given by the formula:

$$\beta_j = \frac{E(\Delta X_j \prod_{k=j+1}^{M-1}(1-\beta_k \Delta X_k)|\mathcal{F}_j)}{E(\Delta X_j^2 \prod_{k=j+1}^{M-1}(1-\beta_k \Delta X_k)^2|\mathcal{F}_j)},$$
$$\gamma_j = \frac{E((V_T^{(l)} - G_T(\xi^{(l)}) - V_j^{(l)} + G_j(\xi^{(l)}))\Delta X_j \prod_{k=j+1}^{M-1}(1-\beta_k \Delta X_k)|\mathcal{F}_j)}{E(\Delta X_j^2 \prod_{k=j+1}^{M-1}(1-\beta_k \Delta X_k)^2|\mathcal{F}_j)}.$$

The existence and uniqueness of this solution have been proven by Schweizer (1995). But the condition for this analytic solution is a bit restrictive so that it cannot be applied to many models, e.g., the stochastic volatility models.

Besides the limited application of the explicit solutions to the quadratic local risk minimization and the total risk minimization criteria, Coleman et al. (2003) propose to use piecewise linear risk minimization criteria. Minimizing quadratic criteria puts more emphasis on the larger errors while minimizing piecewise linear criteria considers the weighted average error. Coleman et al. (2003, 2008) investigate the piecewise linear criterion for local risk minimization and total risk minimization. They also prove that the discrete hedging put-call parity works for both piecewise linear local and total risk minimization criteria. However, there is no analytic solutions for the piecewise linear criteria and it is very expensive to compute. Therefore, Coleman et al. (2003, 2008) propose a method to compute the optimal solution by using Monte Carlo simulation. They generate simulations of a binomial tree model to approximate the stock price path that follows the stochastic differential equation:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ_t. \tag{2.6}$$

Coleman et al. (2003) give the piecewise linear local risk minimization criterion, which is similar to the quadratic ones, given the final condition $\xi_M = 0$ and $\eta_M = H$, and get the optimal holdings by minimizing:

$$E(| X_{j+1}(\xi_{j+1} - \xi_j) + (\eta_{j+1} - \eta_j) || \mathcal{F}_j).$$

The optimal strategy can be computed by solving the L_1-optimization problem backwards in time. From the numerical results, Coleman et al. (2003) conclude that piecewise linear local risk minimization may lead to a large probability of smaller cost and risk but also could result in a small probability of larger cost and risk than the explicit solution of the quadratic local risk minimization criterion.

Coleman et al. (2008) investigate the piecewise linear total risk minimization criterion. This problem is also solved by Monte Carlo simulation. Suppose L simulations are generated, and the option payoff at the maturity for each simulation is denoted by $H^{(k)}$, where $k = 1, 2, .., L$.

For each hedging time $0 \leq j \leq M-1$, the discounted stock price and the change of the discounted stock price are denoted by $X_j^{(k)}$ and $\Delta X_j^{(k)}$ respectively for each simulation. The holding position of the underlying asset is denoted as $\xi_j^{(k)}$ in each simulation path. The objective function for the total risk minimization problem becomes:

$$\min_{V_0, \xi_0, \xi_j^{(k)}} \sum_{k=1}^{L} \left| H^{(k)} - V_0 - \xi_0 \Delta X_0^{(k)} - \sum_{j=1}^{M-1} \xi_j^{(k)} \Delta X_j^{(k)} \right|$$

The dimension of this problem is $L \times M$, which makes it computationally more expensive and much harder to solve than the local risk minimization criterion. To solve this optimization problem, Coleman et al. (2008) put forward a method that the holdings in the hedging strategy is supposed to follow a function of the current and past discounted stock prices. They examine two assumptions of the relationship between the holding and the discounted stock price.

One assumption is that the holding at each hedging time is a function of the discounted stock price at that hedging time.

$$\xi_j = D_j(X) \tag{2.7}$$

This method reduces the number of parameters in the optimization problem to $N \times M$ where N is the number of the parameters in the function $D_j(\cdot)$, which is chosen to represent the relationship between the holding and the discounted stock price at each hedging time. The function $D_0 \equiv \xi_0$ and is a constant. With this assumption, the piecewise linear total risk minimization problem becomes:

$$\min_{V_0, D_j} \sum_{k=1}^{L} \left| H^{(k)} - V_0 - \sum_{j=0}^{M-1} D_j(X_j^{(k)}) \Delta X_j^{(k)} \right|$$

Coleman et al. (2008) apply it to the stochastic differential equation (2.6) by assuming the function $D_j(\cdot)$ as cubic spline functions. Thus, the parameters in the optimization problem become the unknowns in each cubic spline function. However, this constraint may introduce extra risk if the number and the placement of the spline knots are not specified appropriately.

Another assumption is that the holding relates not only to the current discounted stock price at that hedging date, but also to the past discounted stock prices. Thus, Coleman et al. (2008) assume the holding is a linear combination of functions with respect to the current and past discounted stock prices. They introduce more degrees of freedom by allowing the effect of the current discounted stock price X_j at t_j to be different from the effect of the past discounted stock prices $X_0, ..., X_{j-1}$. In this assumption, the holdings in the hedging portfolio becomes:

$$\xi_j = D_j(X_j) + \frac{1}{X_j}\sum_{i=0}^{j-1}\tilde{D}_i(X_i)\Delta X_i, \forall j = 1, ..., M-1 \tag{2.8}$$

where $\tilde{D}_i(\cdot)$ is a function with the different parameters from $D_j(\cdot)$.

After replacing ξ_j in the piecewise total risk minimization problem, the optimization objective function becomes:

$$\min_{V_0, D_j, \tilde{D}_j} \sum_{k=1}^{L} \left| H^{(k)} - V_0 - \sum_{j=0}^{M-1}\left(D_j(X_j^{(k)}) + \sum_{l=0}^{j-1}\tilde{D}_l(X_l^{(k)})\frac{\Delta X_l^{(k)}}{X_j^{(k)}}\right)\Delta X_j^{(k)}\right|$$

This method reduces the complexity of the optimization problem and makes it easier to compute optimal strategy. The assumptions 2.7 and 2.8 are applied to both Black-Scholes model and a stochastic volatility framework. With the numerical results, Coleman et al. (2008) illustrate the pattern of the holdings with different discounted stock prices after a rebalancing time and find that the optimal strategy obtained under such assumptions is similar to the analytic solution, which shows the reliability of this method.

Coleman et al. (2008) use the cubic spline function with a few knots to model the function $D_j(\cdot)$ in their research. The cubic spline function only allows a finite number of knots and the continuity is achieved by constraining the first and second derivatives on either side of each break point to be the same. This create discontinuity at the data points that are not chosen to be the knots. In addition, the result may be sensitive to the choice of the knots, which may not be appropriate when approximating the mathematical relationship between the holding and the discounted stock price.

Instead of the cubic spline function, we introduce the spline generating kernels with an infinite number of knots to model the relationship. The kernel function $\kappa(x, y)$ is the inner product of two mappings $\{\phi_n(x)\}_{n=0}^{+\infty}$ and $\{\phi_n(y)\}_{n=0}^{+\infty}$ as shown below, which allows infinite number of knots.

$$\kappa(x, y) = \langle\{\phi_n(x)\}, \{\phi_n(y)\}\rangle$$

In our case, the mapping is a function of the discounted stock price X. However, instead of giving definition to the mapping $\{\phi_n(\cdot)\}_{n=0}^{+\infty}$, the kernel function is usually specified directly in practice, which leads to computational convenience. In this project, the kernel generating spline with an infinite number knots has the function representation proposed by Vapnik (1998). The detail is shown in Chapter 3.

To model the complex nonlinear relationship between the holding and the underlying asset price, Coleman et al. (2012) propose a method to approximate the nonlinear relationship by a linear combination of spline kernels. By this method, the hedging position function $D_j(X)$ is represented as a linear combination of kernel function values with respect to a chosen set of reference points X^*, i.e.,

$$D_j(X) = \sum_{i=1}^{n} \alpha_{i,j} \kappa(X, X^*_{i,j})$$

where n is the number of the reference points at each hedging time t_j and $\alpha_{i,j}$ are the coefficients.

In this project, we focus on the quadratic total risk minimization criterion using spline kernel function. The relationship between holdings and discounted stock prices at each hedging time is modelled by a linear combination of spline kernels with some reference points. We change the number of the reference points to obtain the best hedging strategy and examine the difference in the hedging performance when the number of reference points changes. Then the regularization term is introduced to the objective function of the optimization problem to see if the hedging performance would be affected more by the regularization penalty parameter compared with the number of the reference points. Cross validation is used when determining the best values of the parameters to avoid overfitting when introducing too many degrees of freedom. The best parameters are chosen when minimizing the total risk for the testing set.

Chapter 3

Total Risk Minimization Using Spline Kernel With Regularization

3.1 Spline Kernel Function

In this project, we use the quadratic total risk minimization criterion as an illustration, which is shown as the formula (2.5). The framework can be applied to the piecewise linear total risk minimization problem. We note that the quadratic total risk minimization criterion aims to minimize the squared difference between the option payoff and the final value of the hedging portfolio, which is an L_2-norm minimization problem. By using the assumption proposed by Coleman et al. (2008) that the holding at each hedging time is a function of the current and past discounted stock prices, we can get the new formulas of the quadratic total risk minimization criterion instead of the formula (2.5). To get the formula (3.1) below, we first assume the holding at each hedging time to be a function of the discounted stock price at that time as shown in the formula (2.7).

Suppose that the number of simulations is L and the number of total hedging times is M. The quadratic total risk minimization becomes the following.

$$\min_{V_0, D_0, \ldots, D_{M-1}} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - D_0 \Delta X_0^{(k)} - \sum_{j=1}^{M-1} D_j(X_j^{(k)}) \Delta X_j^{(k)} \right)^2 \tag{3.1}$$

To model the function $D_j(\cdot)$, we assume a set of the reference points $X_{i,j}^*$, $i = 1, \ldots, n$, where n is the number of the reference points at each hedging time. As we mention in

Chapter 2, the function $D_j(\cdot)$ is a linear combination of $\kappa(\cdot, X^*_{i,j})$, $i = 1, ..., n$, as follows,

$$D_j(X) = \sum_{i=1}^{n} \alpha_{i,j}\kappa(X, X^*_{i,j}) \tag{3.2}$$

where $\kappa(\cdot, X^*_{i,j})$ is a spline kernel function proposed by Vapnik (1998), that is

$$\kappa(X, X^*_{i,j}) = 1 + XX^*_{i,j} + \frac{1}{2}|X - X^*_{i,j}|(X\Lambda X^*_{i,j} + X^*_b)^2 + \frac{(X\Lambda X^*_{i,j} + X^*_b)^3}{3} \tag{3.3}$$

and

$$X\Lambda X^*_{i,j} = \min(X, X^*_{i,j}), \quad X^*_b = \min_{i,j} X^*_{i,j}$$

In this project, the holding D_j depends on the discounted stock price X_j and some selected reference discounted stock prices $X^*_{i,j}$. At each hedging date, the reference price values are pre-determined and X^*_b is the global minimum value of all the reference discounted price values in time. The number of the reference points, n, in the summation sets the degrees of freedom of the function, which needs to be determined.

In addition, for the spline kernel function, the placement of the reference points need to be determined. We assume the reference points to be evenly distributed within the range of the discounted stock price values at each hedging time. The range of the values is given by the minimum and maximum values at that hedging time. Therefore, the placement of the reference points is shown in (3.4).

$$X^*_{i,j} = X^*_{1,j} + \frac{i-1}{n-1}(X^*_{n,j} - X^*_{1,j}), 1 \le i \le n \tag{3.4}$$

where

$$X^*_{n,j} = max(X_j), X^*_{1,j} = min(X_j)$$

After introducing the spline kernel function, the original quadratic total risk minimization problem becomes:

$$\min_{V_0, D_0, \alpha_{i,j}} \frac{1}{L}\sum_{k=1}^{L}\left(H^{(k)} - V_0 - D_0\Delta X_0^{(k)} - \sum_{j=1}^{M-1}\sum_{i=1}^{n}\alpha_{i,j}\kappa(X_j^{(k)}, X^*_{i,j})\Delta X_j^{(k)}\right)^2 \tag{3.5}$$

The formula (3.5) is an L_2-norm optimization problem with unknowns V_0, D_0 and n parameters $\alpha_{i,j}$, $1 \le i \le n$, in the kernel function at each hedging time t_j. We denote the number of unknowns in total as N and $N = n \times (M-1) + 2$, where n is the number of reference price values and M represents the total number of the hedging times. To show the optimization problem more clearly, we can express the formula (3.5) in the matrix version. We define $\mathbf{z}$ as the vector containing all unknowns, $\mathbf{b}$ as the vector of the option payoffs from all the simulations and $\mathbf{A}$ as the parameter matrix.

$$\mathbf{z} = \begin{bmatrix} V_0 \\ \xi_0 \\ \alpha_{1,1} \\ \alpha_{2,1} \\ \cdots \\ \alpha_{n-1,M-1} \\ \alpha_{n,M-1} \end{bmatrix}_{\mathbf{N}\times\mathbf{1}}, \mathbf{b} = \begin{bmatrix} H^{(1)} \\ H^{(2)} \\ H^{(3)} \\ H^{(4)} \\ \cdots \\ H^{(L-1)} \\ H^{(L)} \end{bmatrix}_{\mathbf{L}\times\mathbf{1}}$$

$$\mathbf{A} = \begin{bmatrix} 1 & \Delta X_0^{(1)} & \kappa(X_1^{(1)}, X_{1,1}^*)\Delta X_1^{(1)} & \cdots & \kappa(X_{M-1}^{(1)}, X_{n,M-1}^*)\Delta X_{M-1}^{(1)} \\ 1 & \Delta X_0^{(2)} & \kappa(X_1^{(2)}, X_{1,1}^*)\Delta X_1^{(2)} & \cdots & \kappa(X_{M-1}^{(2)}, X_{n,M-1}^*)\Delta X_{M-1}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \Delta X_0^{(L-1)} & \kappa(X_1^{(L-1)}, X_{1,1}^*)\Delta X_1^{(L-1)} & \cdots & \kappa(X_{M-1}^{(L-1)}, X_{n,M-1}^*)\Delta X_{M-1}^{(L-1)} \\ 1 & \Delta X_0^{(L)} & \kappa(X_1^{(L)}, X_{1,1}^*)\Delta X_1^{(L)} & \cdots & \kappa(X_{M-1}^{(L)}, X_{n,M-1}^*)\Delta X_{M-1}^{(L)} \end{bmatrix}_{\mathbf{L}\times\mathbf{N}}$$

Hence, $\mathbf{z}$, including all unknowns, has the size $N \times 1$, where N is the number of all the parameters in the kernel function plus two constant unknowns. The vector $\mathbf{b}$ has the size $L \times 1$, where L is the total number of the Monte Carlo simulations, and the matrix $\mathbf{A}$ has the size $L \times N$.

Then the quadratic total risk minimization problem can be rewritten as the following formula (3.6).

$$\min \ \frac{1}{L}(\mathbf{Az} - \mathbf{b})^{\mathbf{T}}(\mathbf{Az} - \mathbf{b}) \tag{3.6}$$

By taking the first derivative of the objective function (3.6), we can obtain the optimal solution for this optimization problem as shown in (3.7).

$$\mathbf{z} = (\mathbf{A^T A})^{-\mathbf{1}}\mathbf{A^T b} \tag{3.7}$$

We then consider another assumption that the holding at each hedging time is the function of all the current and past discounted stock prices since the hedging positions may also be affected by the past discounted stock prices. The assumption (2.8) is used to define the relationship between the hedging positions and the discounted stock prices. By assuming the holding at each hedging time to be a function of the current and past discounted stock prices, the quadratic total risk minimization problem becomes:

$$\min_{V_0, D_j, \tilde{D}_j} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - \sum_{j=0}^{M-1} \left(D_j(X_j^{(k)}) + \sum_{l=0}^{j-1} \tilde{D}_l(X_l^{(k)}) \frac{\Delta X_l^{(k)}}{X_j^{(k)}} \right) \Delta X_j^{(k)} \right)^2 \tag{3.8}$$

where

$$\tilde{D}_j(X) = \sum_{i=1}^{n} \tilde{\alpha}_{i,j} \kappa(X, X_{i,j}^*)$$

The functions $D_j(\cdot)$ and $\tilde{D}_j(\cdot)$ are different functions with n reference points for all $1 \leq j \leq M-1$ and $1 \leq j \leq M-2$ respectively while D_0 and $\tilde{D}_0$ are constants. The values of all coefficients $\alpha_{i,j}$ and $\tilde{\alpha}_{i,j}$ need to be determined. Thus, the size of unknowns becomes $N = n \times (M-1) + n \times (M-2) + 3$. Since the information of all the discounted stock prices X_j, $0 \leq j \leq M$, is known, the problem 3.8 can also be solved as a linear least squares problem.

In the reality, the hedging strategy is generated by mathematical models and then applied to the real financial market, which means the data used to create the model is usually different from the realized market data. Thus, we decide to use cross validation and generate two data sets to determine the best model, which are training set and testing set. After obtaining the optimal strategy, we can apply it to the new testing data set to see its hedging performance.

In the spline kernel function, we need to determine the number of the reference points n. This parameter is determined by solving the optimization problems (3.6) and (3.8) to get the smallest total risk for the testing set by cross validation.

3.2 Regularization

When we learn a model by solving an optimization problem, we need to avoid overfitting. Over fitting can be a common problem when computing the best parameters when solving the optimization problems. To obtain the optimal value of the target function, the best model usually can introduce more complexity since it is more likely to obtain a better fitting when the number of parameters increases. However, such an estimation is not the best fitting since it causes the problem that the model over explains the given data set.

In other words, although the estimated function is close to the data points, it can be too complex and sensitive to errors.

Therefore, to solve the over fitting problem, we can introduce an L_2-regularization. For the total risk minimization problem, since reference point placement does not affect spline knots and their placement, spline kernel function is easier for introducing regularization than the cubic spline function, which is another reason that we use the spline kernel functio in this project.

We then combine the objective function (3.5) with the regularization to avoid over-fitting of the function estimation. Therefore, by the method of L_2-regularization, we add the sum of squared values of parameters in the spline kernel function to the objective function (3.5) and multiply the sum with a regularization penalty parameter λ. Then the regularized quadratic total risk minimization problem becomes the formula (3.9).

$$\min_{V_0, D_0, \alpha_{i,j}} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - D_0 \Delta X_0^{(k)} - \sum_{j=1}^{M-1} \sum_{i=1}^{n} \alpha_{i,j} \kappa(X_j^{(k)}, X_{i,j}^{*}) \Delta X_j^{(k)} \right)^2 + \lambda \sum_{j=1}^{M-1} \sum_{i=1}^{n} \alpha_{i,j}^2 \tag{3.9}$$

For the other assumption 2.8, we also introduce the regularization term to the objective function to restrict the complexity for the both two functions, which are $D_j(\cdot)$ and $\tilde{D}_j(\cdot)$. With the new assumption, the number of parameters is larger than the assumption 2.7 and the hedging results may be affected more after introducing the regularization. The optimization problem to be solved is shown as the following formula.

$$\begin{aligned} \min_{V_0, D_j, \tilde{D}_j} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - \sum_{j=0}^{M-1} \left(D_j(X_j^{(k)}) + \sum_{l=0}^{j-1} \tilde{D}_l(X_l^{(k)}) \frac{\Delta X_l^{(k)}}{X_j^{(k)}} \right) \Delta X_j^{(k)} \right)^2 \\ + \lambda \left(\sum_{j=1}^{M-1} \sum_{i=1}^{n} \alpha_{i,j}^2 + \sum_{j=1}^{M-2} \sum_{i=1}^{n} \tilde{\alpha}_{i,j}^2 \right) \end{aligned} \tag{3.10}$$

In this project, we examine four quadratic total risk minimization formulations. The four criteria consider whether the regularization is used and the different assumptions of the relationship between the holding and the discounted stock price. The two relationships are that holding only depends on the current discounted stock prices (assumption (2.7)) at the hedging time and depends both on the current and past discounted stock prices (assumption (2.8)) respectively. The four criteria are shown as the following:

- Strategy 1: quadratic total risk minimization using spline kernel with assumption (2.7);
- Strategy 2: quadratic total risk minimization using spline kernel with assumption (2.8);
- Strategy 3: quadratic total risk minimization using spline kernel with regularization and assumption (2.7);
- Strategy 4: quadratic total risk minimization using spline kernel with regularization and assumption (2.8).

We get the optimal parameter values by generating and fitting data through Monte Carlo simulation. With the optimal strategy computed from each strategy, we report its hedging performance at different hedging frequencies and different discounted strike prices. The evaluation is based on the following two criteria.

- Total risk:
$$E(\mid H - V_M \mid) \tag{3.11}$$
This is the expected difference between the option payoff and the final value of the hedging portfolio.
- Total cost:
$$H - \sum_{k=0}^{M-1} \xi_k \Delta X_k \tag{3.12}$$
This is the total amount of money necessary for the writer to implement the self-financing hedging strategy and honor the option payoff at expiry.

With these two criteria, we need to find the optimal number of reference points and the best regularization penalty parameter that lead to the smallest total risk for the quadratic total risk minimization problems shown above. With the assumption that the holding of the underlying asset in the hedging portfolio is a function of the current and past discounted stock prices, we then use cross validation to find the best parameters in the holding functions for both assumptions.

Cross validation is used to examine the hedging performance for the out-of-sample data. Solving the quadratic total risk optimization problem always tends to minimize the squared sum of the residuals on the given data set. However, the optimal strategy obtained by achieving the smallest total risk on training set may lead to a relatively bad performance when applying the model to a new data set. By cross validation, the optimal parameters are obtained by minimizing the total risk of the testing set.

We show the hedging performances for the above four quadratic total risk minimization criteria in the next chapter. In addition, we show the similarity and difference of their performances for the discrete hedging.

Chapter 4

Simulation Results

4.1 Simulation Settings

In this chapter, we use Monte Carlo simulation to solve the quadratic total risk minimization problem. To find the optimal parameters in the holding functions, we generate two data sets, which are the training set and the testing set. The two data sets contain two different sets of the discounted stock price paths. The time horizon is $[0, T]$, where $T > 0$. The whole time horizon is divided into 600 time steps and the stock price process $\{S_t\}$ is generated at each time spot, which follows the stochastic differential equation below:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ_t \tag{4.1}$$

where Z_t follows a Brownian motion.

The stock price path is generated at each time step during the whole time horizon based on the above stochastic differential equation (4.1). In other words, the whole stock price path is generated at the total 600 time steps. Since we focus on discrete hedging, we only have a few hedging opportunities. We then choose several number of hedging times to show the hedging performance under each hedging frequency. The number of the rebalancing opportunities during the whole time horizon is denoted by M. It also can be represented as the number of the total time steps, which is 600 according to the above settings, divided by the number of time steps per rebalancing time as shown below.

$$M = \frac{600}{\text{\# of timesteps per rebalancing time}}$$

The discrete hedging times are denoted by $0 = t_0 < t_1 < t_2 < ... < t_M = T$. For example, when hedging only once initially, we only hedge on t_1 to see its performance. The best parameters are calculated for several time steps per rebalancing time. If there are 25 time steps per rebalancing time, the number of hedging opportunities $m = 24$. In this report, we examine several number of hedging times at each discounted strike price $K = 90, 95, 100, 105, 110$. Each optimal strategy is used to calculate the average total risk and the average total cost.

Suppose we implement the discrete hedging based on a European put with the discounted strike price K and the option payoff H. The two data sets contain two different sets of the discounted stock price paths, which are calculated by the original stock price paths $\{S_t\}$ and the continuously compounded bond price processes $\{B_t\}$. The basic assumptions for the European put option payoff H and the discounted stock price process $\{X_t\}$ are as the following.

- Continuously compounded bond price: $B_t = e^{rt}$
- Discounted stock price: $X_t = \frac{S_t}{B_t}$
- Option payoff: $H = (K - X_T)_+$

To generate the Monte Carlo simulations, we give some specific settings to the parameter values in the stochastic differential equation and the number of simulations as shown in Table 4.1.

Maturity T	1
Risk-free rate r	0.04
Initial stock price S_0	100
The instantaneous expected return of the stock price μ	0.15
Volatility σ	0.2
# of simulations L	40000

Table 4.1: Basic settings for the Monte Carlo simulation

With the above settings, we compute the optimal holding positions for the quadratic total risk minimization problem with the spline kernel function by cross validation. To obtain the best strategy, two sets of the discounted stock price paths are generated, which are the training set S_1 and the testing set S_2. The two sets are created under the same stochastic differential equation (4.1) but with different random seeds. The parameters of the holding functions obtained from the training set are applied to the testing set to see its testing performance. The number of the reference points n and the regularizing penalty parameter λ are treated as the parameters to be determined. The best parameters n^*, λ^* are those that could minimize the total risk of the testing set. The total risk and total cost are reported at five different discounted strike price from 90 to 110 to see strategy performances for both in-the-money and out-of-the-money situations.

4.2 Comparison With Delta Hedging

As we mentioned in Chapter 1, delta hedging provides a perfect hedging performance in the complete market under the Black-Scholes framework. In the incomplete market, many practitioners are still using delta hedging even if the hedging is discrete. We then make a comparison between the delta hedging and the hedging method used in our project to see their hedging performance under some measures.

4.2.1 Convergence

Under the Black-Scholes framework, delta hedging provide the smallest cost for the continuous hedging. In practice, we can only hedge at discrete times. In addition, we would like to hedge as infrequently as possible due to the transaction cost. However, as the hedging frequency becomes higher, the cost of the hedge should converge to the Black-Scholes price. We then examine the convergence to the Black-Scholes price of our hedging strategy. The Black-Scholes price for the option can be obtained by the function **blsprice** in MATLAB. We use Strategy 1 as an illustration to see the convergence.

Strike	M (# of rebalancing times during the whole time horizon)								
	Delta	600	120	60	24	12	6	2	1
90	2.5315	2.5455	2.5291	2.5194	2.4860	2.4336	2.3372	2.0581	1.7423
95	4.0325	4.0368	4.0217	4.0112	3.9704	3.9135	3.7889	3.4130	2.9704
100	6.0040	6.0208	6.0003	5.9869	5.9415	5.8681	5.7204	5.2631	4.6870
105	8.4499	8.4515	8.4340	8.4209	8.3748	8.3019	8.1433	7.6238	6.9308
110	11.3456	11.3473	11.3227	11.3065	11.2611	11.1927	11.0330	10.4794	9.7067

Table 4.2: Convergence to the Black Scholes price

From the above table, we can see that the cost of Strategy 1 converges to the Black-Scholes price at each strike price as the number of hedging times M becomes large. In addition, for very frequent rebalancing, the cost for implementing Strategy 1 is larger than the Black-Scholes price, which means that the Strategy 1 can't beat Black-Scholes delta hedging when the number of hedging times M is very large. Thus, there is no need to use other hedging strategies except the Black-Scholes delta hedging when the hedging frequency is very high. We then focus on examine the hedging performance at some lower level of hedging frequency. When hedging at a lower level of rebalancing frequency, the cost for Strategy 1 becomes smaller.

Since we focus on the discrete hedging, we then examine the hedging performance of our strategy compared with the discrete Black-Scholes delta hedging. Instead of hedging continuously, the discrete Black-Scholes delta hedging only rebalances at some specific times.

4.2.2 Normalized P&L

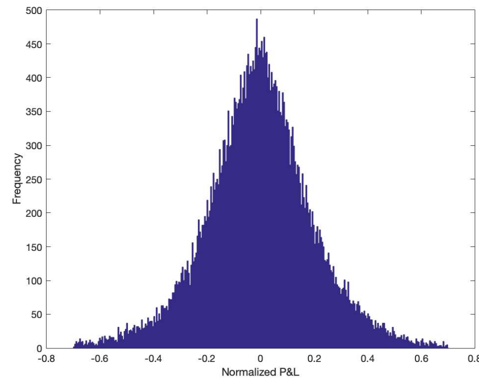
To examine the hedging performance more clearly, we introduce the density plot of the normalized P&L, which is the actual P&L divided by the Black-Scholes price as shown below.

$$\text{Normalized P\&L} = \frac{\text{Actual P\&L}}{\text{Black-Scholes price}}$$

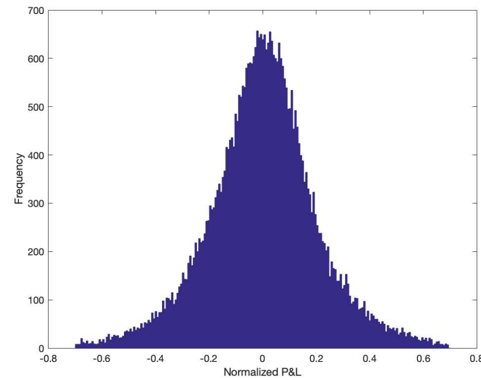
The density plots of the normalized P&L for Black-Scholes discrete delta hedging and the hedging method Strategy 1 when hedging 24 times, 60 tines and 600 times are shown in Figure 4.1.

From the plot, we can see that the patterns of the density plots of Strategy 1 are similar with the patterns of the delta hedging. When hedging more frequently, the volatility of the normalized P&L is getting smaller. This means that we can make effective risk management if we hedge at a high frequency. However, in reality we need to consider the transaction cost and try to achieve a balance between the hedging effectiveness and the hedging cost. In this project, we would like to focus on hedging at a lower level of frequency.

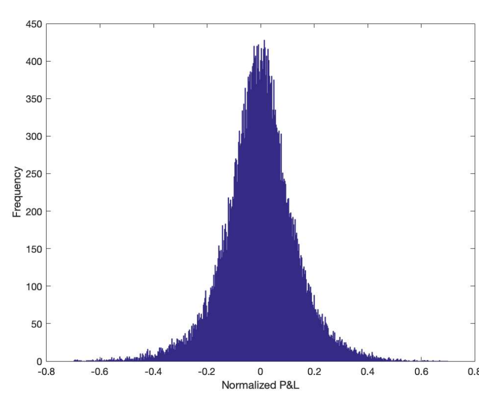
The density plot shows the stability of the hedging performance. To compare the hedging performance between the Black-Scholes discrete delta hedging and our hedging strategy thoroughly, we also introduce some other risk measures.



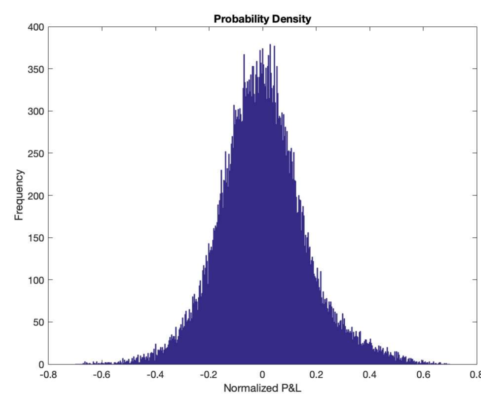
(a) Delta hedging (hedging 24 times)



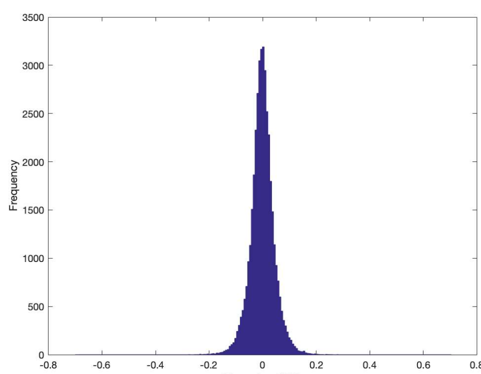
(b) Strategy 1 (hedging 24 times)



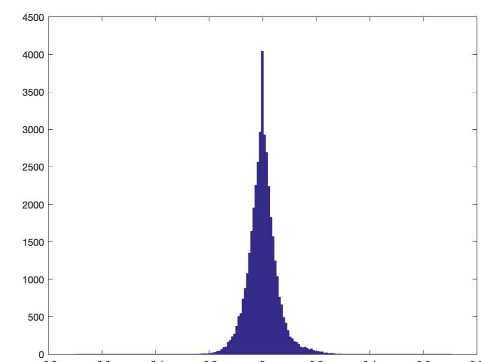
(c) Delta hedging (hedging 60 times)



(d) Strategy 1 (hedging 60 times)



(e) Delta hedging (hedging 600 times)



(f) Strategy 1 (hedging 600 times)

Figure 4.1: Frequency versus normalized P&L with $K = 100$

4.2.3 Other Risk Measures

The hedging performance can also be compared based on some other risk measures like VaR and CVaR. Value at risk (VaR) and Conditional Value at Risk (CVaR) (Rockafellar and Uryasev, 2000) are statistical measures of the riskiness of financial portfolios.

VaR is defined as the maximum dollar amount expected to be lost over a given time horizon, at a pre-defined confidence level, which is

$$VaR_\alpha(X) = \min\{z \mid F_X(z) \geq \alpha\}.$$

The Conditional Value at Risk (CVaR) is the conditional expectation of X, given that $X \geq VaR_\alpha(X)$, i.e.

$$CVaR_\alpha(X) = E[X \mid X \geq VaR_\alpha(X)].$$

In this project, we use the confidence level $\alpha = 95\%$ to examine the VaR and CVaR for the Black-Scholes discrete delta hedging and our hedging strategy. The VaR and CVaR values when strike price $K = 100$ are used as illustrations to make comparison. The values of VaR and CVaR for the P&L when hedging 24 times and 600 times are shown below.

Strike	Mean	Standard deviation	VaR	CVaR
90	-0.0111	0.3582	-0.5960	-0.8942
95	-0.0104	0.2726	-0.4594	-0.6644
100	-0.0070	0.2114	-0.3550	-0.5012
105	-0.0063	0.1653	-0.2783	-0.3899
110	-0.0052	0.1290	-0.2194	-0.3056

Table 4.3: Statistics for the P&L of Delta hedging (hedging 24 times)

Strike	Mean	Standard deviation	VaR	CVaR
90	-1.66E-10	0.3448	-0.5888	-0.8858
95	-1.22E-10	0.2647	-0.4496	-0.6396
100	-9.94E-11	0.2021	-0.3362	-0.4752
105	-8.17E-11	0.1589	-0.2632	-0.3633
110	-7.29E-11	0.1234	-0.2049	-0.2869

Table 4.4: Statistics for the P&L of Strategy 1 (hedging 24 times)

Strike	Mean	Standard deviation	VaR	CVaR
90	-0.0018	0.0745	-0.1181	-0.1757
95	-0.0012	0.0559	-0.0903	-0.1299
100	-0.0009	0.0435	-0.0701	-0.0991
105	-0.0005	0.0337	-0.0547	-0.0751
110	-0.0001	0.0268	-0.0429	-0.0592

Table 4.5: Statistics for the P&L of Delta hedging (hedging 600 times)

Strike	Mean	Standard deviation	VaR	CVaR
90	-1.89E-11	0.2115	-0.3068	-0.4177
95	-4.61E-12	0.1264	-0.1888	-0.2325
100	-2.68E-12	0.0910	-0.1379	-0.1857
105	-2.21E-12	0.0884	-0.1284	-0.1653
110	-9.02E-13	0.0730	-0.1029	-0.1238

Table 4.6: Statistics for the P&L of Strategy 1 (hedging 600 times)

Since the total risk minimization strategy aims to minimize the difference between the option payoff and the final value of the hedging portfolio, the mean value of P&L for this strategy is naturally very small. Therefore, we do not focus on the mean values of P&L. Instead, the hedging performance can be compared based on the standard deviation, VaR and CVaR.

From the results, we can see that Strategy 1 shows smaller standard deviation than the Black-Scholes discrete delta hedging when hedging 24 times, while it shows larger volatility when hedging 600 times. This result indicates that when hedging at a high level of frequency, the hedging performance of Strategy 1 is not as stable as that of the Black-Scholes discrete delta hedging.

In terms of VaR and CVaR, Strategy 1 has smaller absolute values than the Black-Scholes discrete delta hedging when hedging 24 times, which indicate a better hedging performance at such a level of hedging frequency. However, for very frequent rebalancing like hedging 600 times during the time horizon, Strategy 1 shows larger absolute values of VaR and CVaR than the Black-Scholes discrete delta hedging. In this case, Strategy 1 can't beat Black-Scholes discrete delta hedging as it may lead to larger loss with a certain confidence level.

Thus, for discrete rebalancing, when the number of hedging times M is large, the strategy introduced in this project will not be able to exceed the Black-Scholes discrete delta hedging while when M is small, the strategy could beat the delta hedging. Since we are interested in the discrete hedging at lower frequency due to the transaction cost in reality, we focus on examine the hedging performance at a lower level of rebalancing frequency in this project. The hedging performances of our four strategies are examined when the number of hedging times $M = 24, 12, 6, 2, 1$ at each discounted strike price $K = 90, 95, 100, 105, 110$. The optimal parameters for each strategy are obtained to calculate the average total risk and the average total cost to see the hedging performance.

4.3 Spline Kernel Without Regularization

In this chapter, four strategies are examined. Strategy 1 and Strategy 2 using the assumption (2.7) and (2.8) respectively without regularization. Strategy 3 and 4 introduce the regularization method to see if there is any improvement in reducing the total risk and the total cost after adding the regularization term to the optimization objective function. We first consider the cases without regularization.

4.3.1 Benchmark

Coleman et al. (2008) give the average total risk and the average total cost of the Monte Carlo simulation by solving the quadratic total risk minimization with cubic spline function. We treat the results as the benchmark. The total risk and the total cost are defined in Chapter 3, and the average values among all the simulations for both criteria are reported in this project. The results for the benchmark given by Coleman et al. (2008) with the assumption 2.7 are shown in Table 4.7 and 4.8 while the results with the assumption 2.8 are in Table 4.13 and 4.14.

The results of the benchmark come from the situation that the cross validation is not used. Therefore, we compare the total risk and total cost obtained from the training set with the values from the benchmark. Here we do not add the regularization term to the optimization problem, and use strategy 1 and strategy 2 as illustrations.

4.3.2 Strategy 1

We first consider Strategy 1, whose assumption is that the holding is a linear combination of spline kernel values with respect to some reference points which are selected among the training discounted stock prices. With Strategy 1, we assume there is no regularization and we obtain the optimal strategy by solving the following optimization problem for each discounted strike price K and each hedging frequency.

$$\min_{V_0, D_0, \ldots, D_{M-1}} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - D_0 \Delta X_0^{(k)} - \sum_{j=1}^{M-1} D_j(X_j^{(k)}) \Delta X_j^{(k)} \right)^2 \tag{4.2}$$

To solve the quadratic total risk minimization problem, we need to determine the number of the reference points. We randomly choose the number of reference points $n = 60$ to see its performance and make the comparison with the results using the cubic spline function implemented in Coleman et al. (2008).

By using the spline kernel representation, we can see the average total risk and the average total cost at different discounted strike price and different number of time steps per rebalancing time in Table 4.7 and 4.8 respectively. Compared with the benchmark implemented in Coleman et al. (2008), using spline kernel function provides smaller total risk and total cost than using the cubic spline function, which indicates that the relationship between the holding and the underlying asset price could be better described by using the spline kernel function with infinite number of knots. The result that using the cubic spline does not perform as well as the spline kernel does may come from the inappropriate choices of the number and the placement of the knots in the cubic spline function.

In the real financial market, we cannot have the information of the future discounted stock price to create the best model for hedging. For the optimal hedging strategies obtained from the training set, we would like to know if the hedging strategy could perform well in a new series of discounted stock prices. Usually, the optimal strategy for the training set is not the best for the testing set due to the problem of overfitting. Therefore, we need to get the optimal strategy model using cross validation.

The reference discounted stock price values $X_{i,j}^*$ are assumed to be evenly distributed at each hedging time. We consider the number of the reference points from 1 to 100 to see the total risk patterns for both training set and testing set. By cross validation, the best choice for the number of the reference points n is the one that minimizes the total risk for the testing set.

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Benchmark	0.6312	0.8410	1.1212	1.5727	1.7707
	Strategy 1	0.5896	0.8147	1.1109	1.5674	1.7707
95	Benchmark	0.7918	1.0771	1.4687	2.1945	2.6222
	Strategy 1	0.7587	1.0546	1.4667	2.1892	2.6231
100	Benchmark	0.9877	1.3144	1.7784	2.7944	3.5117
	Strategy 1	0.9076	1.2675	1.7718	2.7733	3.5157
105	Benchmark	1.1068	1.4677	2.0051	3.2892	4.3184
	Strategy 1	1.0123	1.4254	1.9971	3.2544	4.3107
110	Benchmark	1.1308	1.5344	2.1240	3.6189	4.9366
	Strategy 1	1.0706	1.5200	2.1217	3.5800	4.9328

Table 4.7: Average total risk for the Benchmark and Strategy 1 with $n = 60$

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Benchmark	2.4540	2.4033	2.3155	2.0400	1.7421
	Strategy 1	2.4183	2.3884	2.2963	2.0351	1.7423
95	Benchmark	3.9512	3.8830	3.7647	3.4006	2.9735
	Strategy 1	3.9066	3.8647	3.7472	3.3919	2.9704
100	Benchmark	5.9183	5.8396	5.6983	5.2566	4.6948
	Strategy 1	5.8663	5.8138	5.6826	5.2412	4.6870
105	Benchmark	8.3613	8.2809	8.1280	7.6307	6.9449
	Strategy 1	8.3086	8.2453	8.1128	7.6066	6.9308
110	Benchmark	11.2566	11.1789	11.0264	10.4994	9.7148
	Strategy 1	11.2136	11.1459	11.0189	10.4794	9.7067

Table 4.8: Average total cost for the Benchmark and Strategy 1 with $n = 60$

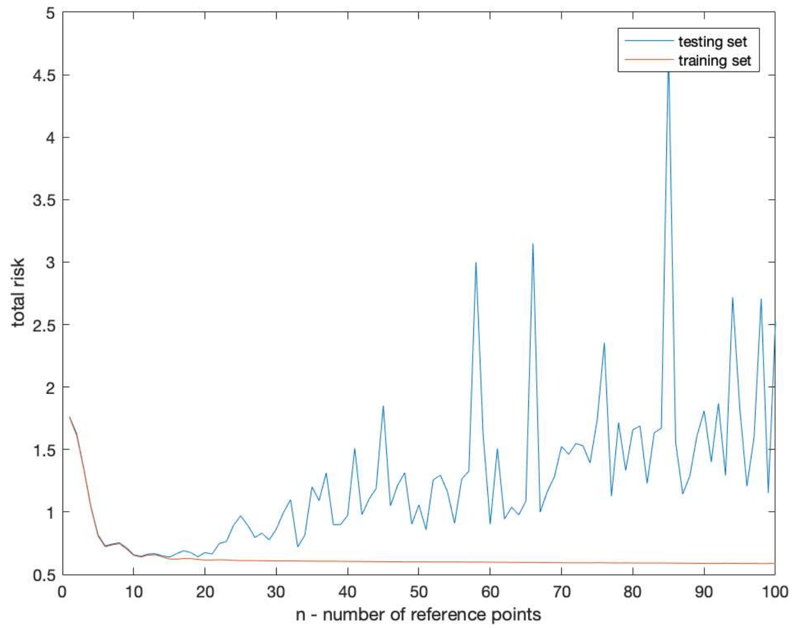


Figure 4.2: Average total risk for strategy 1 with cross validation at $K = 90$

Figure 4.2 shows the total risks with different number of the reference points for $K = 90$ and hedging 24 times, where $n^* = 15$ in this case. From the above figure, we can see that the values of total risk for the testing set first increases when n increases and then starts to decrease when n reaches certain value and keeps going large, which is common to see in the cross validation patterns. The total risk reaches its minimum when the number of reference points is 15. Therefore, we determine it as the best choice for Strategy 1 when $K = 90$ and hedging 24 times, and repeat the process to report the optimal average total risk and total cost with the best choice of the number of the reference points for other discounted strike prices and hedging frequencies. When the hedging frequency is high, the best number of the reference points is around 15 while when the hedging frequency is low, like hedging twice, the value of n^* is around 8, which is a bit smaller since there is no need to introduce much complexity when the number of features in the problem is relatively small. For different discounted strike price, the best value of n does not show a large difference.

By cross validation, the best choice for the number of the reference points n for Strategy 1 is the one that makes the total risk for the testing set be minimum. Thus, we focus on examining the total risk and the total cost for the testing set. The values of the average total risk and the average total cost for the testing set using Strategy 1 at each discounted strike price K and hedging frequency are reported as follows.

Strike	M (# of hedging times)				
	24	12	6	2	1
90	0.6371	0.8543	1.1338	1.5720	1.7945
95	0.8118	1.0965	1.4803	2.1831	2.6380
100	0.9476	1.3002	1.7820	2.7714	3.5225
105	1.0708	1.4554	2.0070	3.2588	4.3300
110	1.1256	1.5388	2.1410	3.5868	4.9525

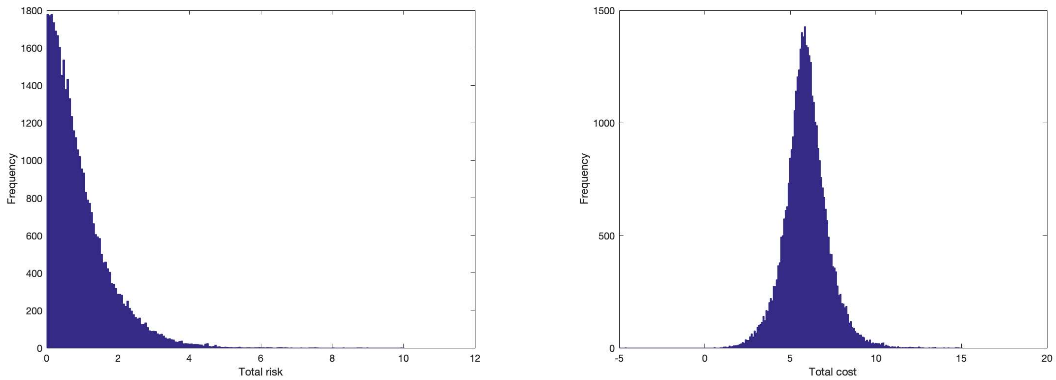
Table 4.9: Average total risk for testing set using Strategy 1

Strike	M (# of hedging times)				
	24	12	6	2	1
90	2.4620	2.4202	2.3342	2.0576	1.7423
95	3.9477	3.8946	3.7783	3.4081	2.9704
100	5.9089	5.8438	5.7155	5.2604	4.6870
105	8.3503	8.2765	8.1421	7.6218	6.9308
110	11.2522	11.1759	11.0358	10.4785	9.7067

Table 4.10: Average total cost for testing set using Strategy 1

We then analyze the distributions of the total risk and the total cost. The density plots of the total risk and the total cost when hedging 24 times at strike price $K = 100$ are shown in the Figure 4.3. As shown in the Table 4.9 and 4.10, the average total risk and total cost in this case are 0.9476 and 5.9089 respectively.

From the plot, we can see that the distribution of the total risk is asymmetric. The total risk has a large probability to have small values and a small probability to have large values. The total cost has a more symmetric distribution. The values in the neighborhood of the mean are most possible to achieve.



(a) Total risk ($M = 24$) (b) Total cost ($M = 24$)

Figure 4.3: Density plot of total risk and total cost for Strategy 1

We can also examine of values of VaR and CVaR for the total risk and total cost to see its hedging performance in the extreme cases. Given the confidence level $\alpha = 95\%$, the values of VaR and CVaR for the total risk and total cost at the right tail when using Strategy 1 are shown in the tables below.

Strike	Risk Measure	M (# of hedging times)				
		24	12	6	2	1
90	VaR	1.9789	2.6940	3.4896	4.3930	4.8491
	CVaR	2.7875	3.6619	4.9306	8.0370	10.3693
95	VaR	2.3657	3.1827	4.2031	6.0468	7.5196
	CVaR	3.1871	4.1863	5.6074	9.4070	12.7732
100	VaR	2.6060	3.5588	4.7723	7.2628	9.6788
	CVaR	3.5143	4.5954	6.1826	10.2829	14.4387
105	VaR	2.8829	3.9006	5.2730	7.9769	11.2036
	CVaR	3.8801	4.9556	6.6993	10.9249	15.5210
110	VaR	3.0069	4.0936	5.5014	8.3088	12.2248
	CVaR	4.1129	5.2398	7.1485	11.4839	16.2611

Table 4.11: VaR and CVaR of total risk for testing set using Strategy 1

Strike	Risk Measure	M (# of hedging times)				
		24	12	6	2	1
90	VaR	4.0199	4.5942	5.3230	6.4505	6.6256
	CVaR	4.9075	5.7834	7.0703	10.0945	12.1458
95	VaR	5.8540	6.5272	7.4809	9.4549	10.5241
	CVaR	6.7586	7.7301	9.1760	12.8150	15.7777
100	VaR	8.0472	8.8154	9.9271	12.5231	14.4052
	CVaR	9.0273	10.0691	11.6430	15.5432	19.1651
105	VaR	10.7025	11.5162	12.7112	15.5987	18.1699
	CVaR	11.7740	12.8145	14.5284	18.5467	22.4873
110	VaR	13.6992	14.5899	15.8118	18.7874	21.9496
	CVaR	14.9304	15.9968	17.8127	21.9625	25.9859

Table 4.12: VaR and CVaR of total cost for testing set using Strategy 1

4.3.3 Strategy 2

In Strategy 2, we use a different assumption to model the relationship between the holding and the discounted stock price from Strategy 1. In this strategy, the holding of the underlying asset at each hedging date is the function of the current and past discounted stock prices. In addition, this strategy assumes that the effect of the current discounted stock prices is different from the effect of the past discounted stock prices. Thus, we assume that the holding of the underlying asset follows different functions of the current and past discounted stock prices, which are denoted as $D_j(\cdot)$, where $0 \leq j \leq M-1$, and $\tilde{D_j}(\cdot)$, where $0 \leq j \leq M-2$.

After introducing the new relationship between the holdings and the discounted stock price, the quadratic total risk optimization problem becomes the formula recalled as the following.

$$\min_{V_0, D_j, \tilde{D}_j} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - \sum_{j=0}^{M-1} \left(D_j(X_j^{(k)}) + \sum_{l=0}^{j-1} \tilde{D}_l(X_l^{(k)}) \frac{\Delta X_l^{(k)}}{X_j^{(k)}} \right) \Delta X_j^{(k)} \right)^2 \tag{4.3}$$

The average total risk and cost of using Strategy 2 with the number of reference points $n = 60$ compared with the benchmark is shown in Table 4.13 and 4.14. Again we can see the total risk and the total cost of using Strategy 2 are smaller than those of the benchmark, which confirms the better performance by using the spline kernel function.

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Benchmark	0.5450	0.7497	1.0325	1.5722	1.7707
	Strategy 2	0.4933	0.7104	1.0182	1.5674	1.7707
95	Benchmark	0.6952	0.9662	1.3551	2.1908	2.6222
	Strategy 2	0.6463	0.9308	1.3474	2.1892	2.6231
100	Benchmark	0.8563	1.1789	1.6518	2.7843	3.5117
	Strategy 2	0.7868	1.1340	1.6357	2.7733	3.5157
105	Benchmark	0.9722	1.3319	1.8802	3.2738	4.3184
	Strategy 2	0.8940	1.2906	1.8551	3.2543	4.3107
110	Benchmark	1.0460	1.4279	2.0079	3.6025	4.9366
	Strategy 2	0.9623	1.3938	1.9956	3.5800	4.9328

Table 4.13: Average total risk for the Benchmark and Strategy 2 with $n = 60$

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Benchmark	2.4504	2.4086	2.3224	2.0388	1.7421
	Strategy 2	2.3609	2.3706	2.2867	2.0347	1.7423
95	Benchmark	3.9443	3.8885	3.7741	3.3983	2.9735
	Strategy 2	3.8494	3.8470	3.7395	3.3915	2.9704
100	Benchmark	5.9118	5.8455	5.7119	5.2530	4.6948
	Strategy 2	5.8114	5.7969	5.6770	5.2407	4.6870
105	Benchmark	8.3584	8.2882	8.1412	7.6261	6.9449
	Strategy 2	8.2579	8.2295	8.1095	7.6061	6.9308
110	Benchmark	11.2569	11.1881	11.0413	10.4945	9.7148
	Strategy 2	11.1690	11.1317	11.0176	10.4790	9.7067

Table 4.14: Average total cost for the Benchmark and Strategy 2 with $n = 60$

The number of the parameters in the above formula is more than that of Strategy 1, which introduces more complexity to the optimization problem. We then find the best n for Strategy 2 at each discounted strike price K and each hedging frequency by using cross validation.

With the number of the reference points n from 1 to 100, the total risks for the training set and the testing set when the discounted strike price $K = 90$ and hedging 24 times are shown in Figure 4.4, where we can find that the best choice of the number of the reference points at this discounted strike price and hedging frequency is around 12. This is the number of the reference points at each hedging time for both the holding functions $D_j(\cdot)$ and $\tilde{D}_j(\cdot)$. The figure has a similar pattern to Strategy 1 that also does not include the regularization term for the optimization problem. We then find the best number of reference points for each discounted strike price and hedging frequency and report the optimal average total risk and the average total cost for the testing set.

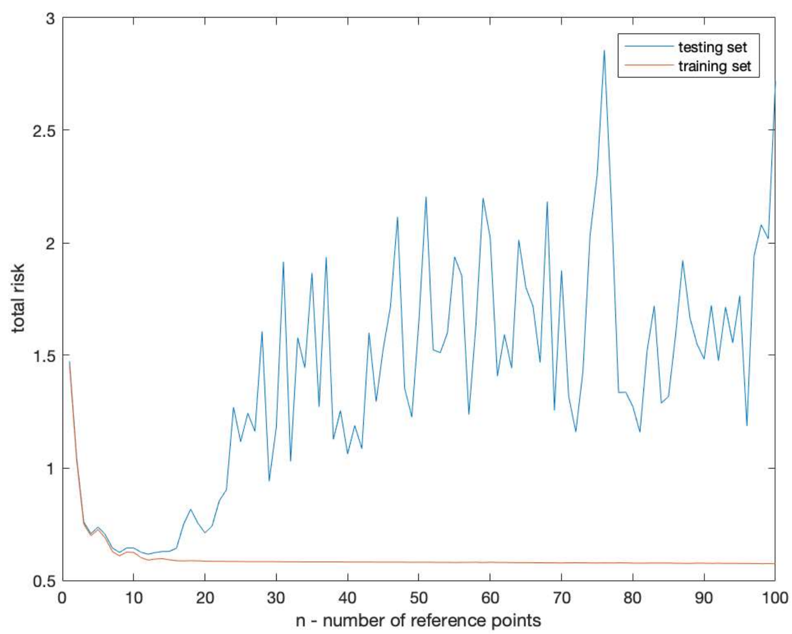


Figure 4.4: Average total risk for strategy 2 with cross validation at $K = 90$

By using the best choice of the number of the reference points, the average total risk and the average total cost for the testing set is shown in Table 4.15 and Table 4.16 respectively.

Strike	M (# of hedging times)				
	24	12	6	2	1
90	0.5641	0.7668	1.0498	1.5715	1.7945
95	0.7100	0.9802	1.3647	2.1823	2.6380
100	0.8566	1.1853	1.6513	2.7702	3.5225
105	0.9708	1.3421	1.8745	3.2580	4.3300
110	1.0362	1.4272	2.0111	3.5858	4.9525

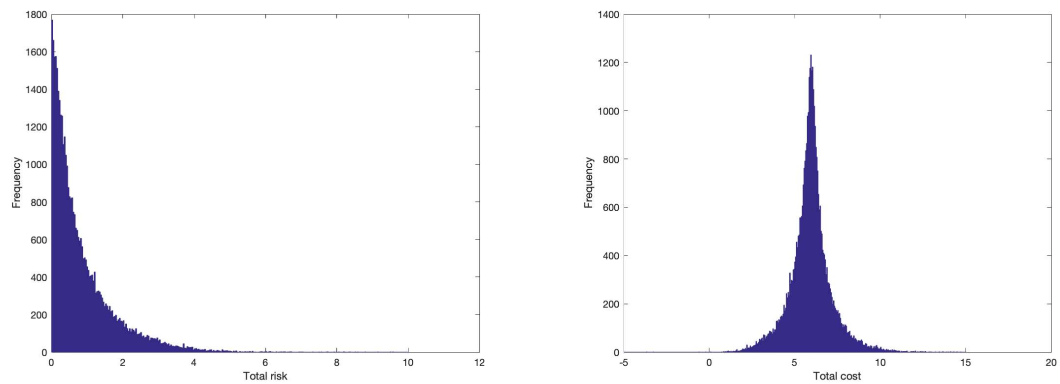
Table 4.15: Average total risk for testing set using Strategy 2

Strike	M (# of hedging times)				
	24	12	6	2	1
90	2.4630	2.4213	2.3270	2.0556	1.7423
95	3.9505	3.8987	3.7813	3.4092	2.9704
100	5.9130	5.8504	5.7177	5.2609	4.6870
105	8.3555	8.2844	8.1474	7.6256	6.9308
110	11.2582	11.1835	11.0327	10.4801	9.7067

Table 4.16: Average total cost for testing set using Strategy 2

Using the different assumptions of the relationship between the holding and the underlying asset price, we can see that the total risks at different discounted strike prices and different numbers of time steps per rebalancing time for the testing set all show reduction compared with Strategy 1. However, for the total cost, the values obtained from Strategy 2 do not show an obvious improvement compared with Strategy 1.

The distributions of the total cost and total risk for Strategy 2 are in the following density plot. Compared with the density plot of Strategy 1 in Figure 4.3, Strategy 2 seems to have more values located near the neighborhood of the mean value.



(a) Total risk ($M = 24$)

(b) Total cost ($M = 24$)

Figure 4.5: Density plot of total risk and total cost for Strategy 2

The following tables show the VaR and CVaR of total risk and total cost when using Strategy 2. Compared with the values shown in Table 4.11 and 4.12 for Strategy 1, we notice that the values of VaR and CVaR for Strategy 2 are smaller than the values for Strategy 1. This result indicates that in terms of the VaR and CVaR, Strategy 2 makes better risk management than Strategy 1, which is consistent with the comparison in terms of the total risk.

Strike	Risk Measure	M (# of hedging times)				
		24	12	6	2	1
90	VaR	2.0202	2.7151	3.5053	4.4020	4.8491
	CVaR	2.8824	3.7708	4.9850	8.0460	10.3693
95	VaR	2.3798	3.1897	4.2121	6.0480	7.5196
	CVaR	3.2346	4.3274	5.7325	9.4167	12.7732
100	VaR	2.7097	3.5941	4.8101	7.2715	9.6788
	CVaR	3.5850	4.7640	6.3534	10.2859	14.4387
105	VaR	2.8996	3.9082	5.2896	7.9890	11.2036
	CVaR	4.0084	5.0713	6.8603	10.9313	15.5210
110	VaR	3.0139	4.1018	5.6150	8.3142	12.2248
	CVaR	4.1371	5.3112	7.2807	11.4920	16.2611

Table 4.17: VaR and CVaR of total risk for testing set using Strategy 2

Strike	Risk Measure	M (# of hedging times)				
		24	12	6	2	1
90	VaR	3.9278	4.4552	5.2185	6.4496	6.6256
	CVaR	4.8931	5.7782	7.0544	10.0937	12.1458
95	VaR	5.7657	6.4179	7.3257	9.4501	10.5241
	CVaR	6.7512	7.7111	9.1547	12.8148	15.7777
100	VaR	7.9995	8.7184	9.8085	12.5218	14.4052
	CVaR	9.0147	10.0374	11.5671	15.5402	19.1651
105	VaR	10.6258	11.4332	12.4996	15.5909	18.1699
	CVaR	11.7433	12.8077	14.4696	18.5432	22.4873
110	VaR	13.6301	14.4807	15.6188	18.7729	21.9496
	CVaR	14.8543	15.0730	17.7426	21.9606	25.9859

Table 4.18: VaR and CVaR of total cost for testing set using Strategy 2

4.4 Spline Kernel With Regularization

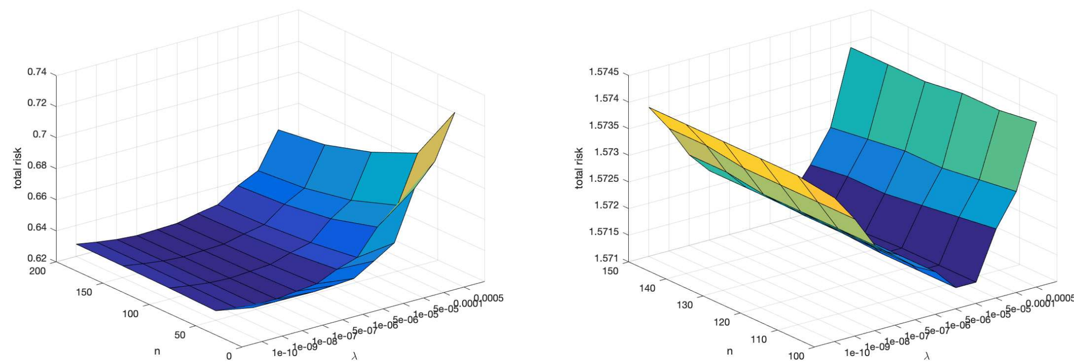
4.4.1 Strategy 3

For Strategy 3, we would like to see how the optimal strategies perform after we add the regularization term to the objective function of the quadratic total risk optimization problem.

$$\min_{V_0,D_0,\ldots,D_{M-1}} \frac{1}{L}\sum_{k=1}^{L}\left(H^{(k)} - V_0 - D_0\Delta X_0^{(k)} - \sum_{j=1}^{M-1} D_j(X_j^{(k)})\Delta X_j^{(k)}\right)^2 + \lambda\sum_{j=1}^{M-1}\sum_{i=1}^{n}\alpha_{i,j}^2 \quad (4.4)$$

We need to determine the best regularizing penalty parameter λ and the number of the reference points n for each discounted strike price K and each hedging frequency. Again we get the optimal values for λ and n by cross validation. Since the objective function always tends to minimize its value, the best number of λ on the training set is 0 all the time. Thus, we find the best parameters based on the pattern of the total risk for the testing set.

For Strategy 1 and Strategy 2, with the range of n in $[1, 100]$, the total risk achieves its minimum at a relatively small number of reference points. In Strategy 3, with the introduction of the regularization parameter, we can choose a wider range for the number of reference points, since the regularization parameter λ can be used to restrict the complexity of the model. For each discounted strike price K and hedging frequency, we find the optimal strategy by considering the different pairs of λ and n. The range of λ is set to be from 10^{-10} to 10^{-4} and the number of the reference points n are selected from 10 to 200.



(a) Hedging twenty-four times

(b) Hedging twice

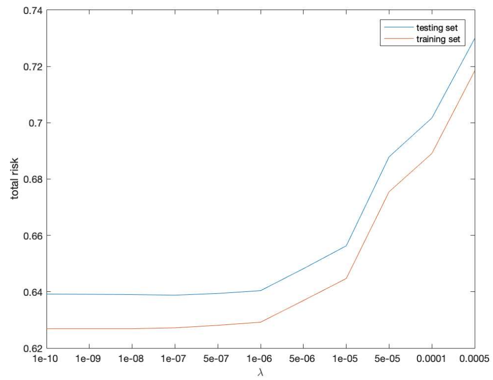
Figure 4.6: Average total risk for testing set with pairs of λ and n using Strategy 3

Using the discounted strike price $K = 90$ as an example, Figure 4.6 shows the total risk obtained from the optimal strategy with the different pairs of n and λ when hedging 24 times and hedging twice respectively for Strategy 3. The optimal hedging strategy for Strategy 3 can be found by comparing the minimum total risk for each pair of n and λ. We can see that the total risk reaches its minimum at a specific pair of n and λ, and begins to increase when going away from that pair. For example, the optimal choices in the case of hedging 24 times with $K = 90$ are $n^* = 160$ and $\lambda^* = 10^{-7}$. By implementing at each strike price and hedging frequency, we find that the optimal choice for n is in the range of $[50, 160]$ while the optimal choice for λ is in the range of $[10^{-7}, 10^{-5}]$. If we fix one parameter and examine the relationship between the total risk and the other parameter, we could also see that total risk first decreases and then increases after the parameter reaches its optimal value.

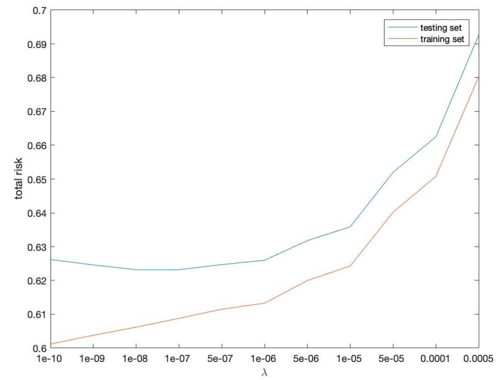
In addition, from Figure 4.6, we could find that the total risk seems to change more with respect to λ than n. For a fixed value of λ, the total risk seems to change slightly when the number of reference points n changes. However, with a fixed value of n, the total risk provides a obvious change with the change of λ.

To show it clearly, we can examine the change of the total risk with respect to the change of parameter λ when fixing the number of the reference points n. The results can be seen in Figure 4.7, where the number of the reference points is chosen to be $n = 10$, 50, 100, 200. On the other hand, we can compare the total risk with the fixed value of λ for different numbers of the reference points n. Using the discounted strike price $K = 90$ as an example, Figure 4.8 shows the relationship between the total risk and the number of the reference points n with the fixed value of λ when hedging 24 times for 4.8(a) and hedging twice for 4.8(b).

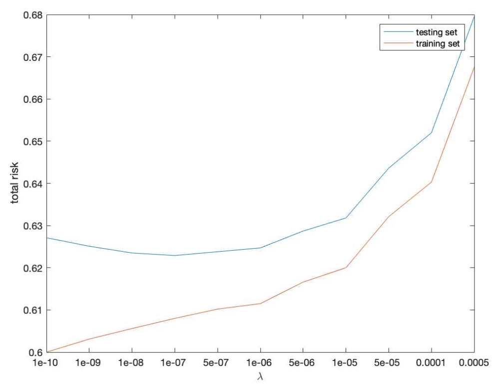
From the results in the following figures, we notice that when choosing different λ for a fixed n, the total risk changes in a wide range. However, the value of the total risk changes in a relatively smaller range when changing n for a fixed λ. When the value of λ becomes much larger, the total risk increases quickly for each fixed number of the reference points n shown in Figure 4.7. In addition, for different choices of n, the optimal values of λ tend to be the same. From Figure 4.8, we can also see that when the hedging frequency gets lower, the optimal strategy tends to have smaller n at certain λ, although the difference for the value of the total risk with different n is relatively small. Therefore, we may conclude that in this optimization problem, the value of the total risk is affected more by the value of λ.



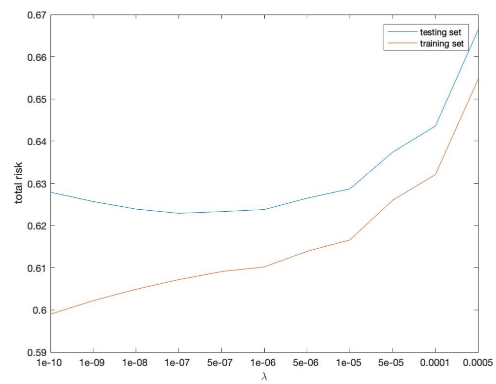
(a) Total risk versus λ for $n = 10$



(b) Total risk versus λ for $n = 50$

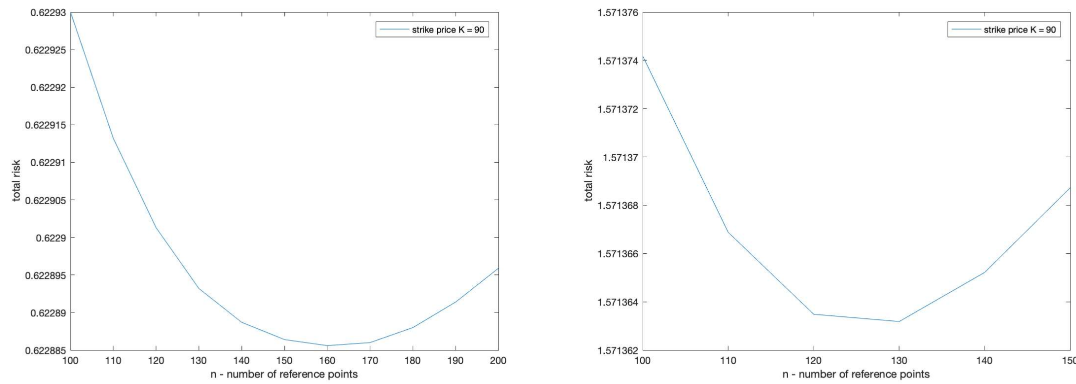


(c) Total risk versus λ for $n = 100$



(d) Total risk versus λ for $n = 200$

Figure 4.7: Change of total risk with respect to λ for fixed n



(a) Total risk versus n for $\lambda = 10^{-7}$ (with twenty-four hedging times)

(b) Total risk versus n for $\lambda = 10^{-5}$ (with two hedging times)

Figure 4.8: Change of total risk with respect to n for fixed λ

After obtaining the optimal choice of the parameter pairs (λ^*, n^*) for each discounted strike price K and hedging frequency with the two-dimensional (2D) cross validation, we report the value of the minimum total risk and the total cost for the testing set in Table 4.19 and 4.20 respectively.

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Strategy 1	0.6371	0.8543	1.1338	1.5720	1.7945
	Strategy 3	0.6229	0.8505	1.1332	1.5715	1.7945
95	Strategy 1	0.8118	1.0965	1.4803	2.1831	2.6380
	Strategy 3	0.7909	1.0901	1.4782	2.1827	2.6380
100	Strategy 1	0.9476	1.3002	1.7820	2.7714	3.5225
	Strategy 3	0.9352	1.2983	1.7817	2.7710	3.5225
105	Strategy 1	1.0708	1.4554	2.0070	3.2588	4.3300
	Strategy 3	1.0398	1.4482	2.0054	3.2581	4.3300
110	Strategy 1	1.1256	1.5388	2.1410	3.5868	4.9525
	Strategy 3	1.1070	1.5369	2.1388	3.5859	4.9525

Table 4.19: Average total risk using Strategy 3 with 2D cross validation

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Strategy 1	2.4620	2.4202	2.3342	2.0576	1.7423
	Strategy 3	2.4627	2.4198	2.3362	2.0574	1.7423
95	Strategy 1	3.9477	3.8946	3.7783	3.4081	2.9704
	Strategy 3	3.9509	3.8969	3.7785	3.4085	2.9704
100	Strategy 1	5.9089	5.8438	5.7155	5.2604	4.6870
	Strategy 3	5.9107	5.8460	5.7160	5.2602	4.6870
105	Strategy 1	8.3503	8.2765	8.1421	7.6218	6.9308
	Strategy 3	8.3506	8.2768	8.1422	7.6213	6.9308
110	Strategy 1	11.2522	11.1759	11.0358	10.4785	9.7067
	Strategy 3	11.2526	11.1751	11.0366	10.4780	9.7067

Table 4.20: Average total cost using Strategy 3 with 2D cross validation

From the total risk for the testing set shown in Table 4.19, we can see that the total risks for Strategy 3 become smaller than those for Strategy 1, especially when the number of time steps per rebalancing time is relatively small. Table 4.20 shows the total cost for Strategy 3 compared with Strategy 1, which is using the assumption 2.7 without the regularization. From the results shown in the above tables, we can find that the total costs obtained from Strategy 1 and Strategy 3 do not show an obvious difference when using the optimal choice of the parameters to implement both hedging strategies. By comparing the results obtained from Strategy 1 and Strategy 3, we find that adding the regularization term to the objective function of the optimization problem could reduce the total risk by maintaining the similar values for the total cost, which possibly gives a better hedging performance.

However, with the observations from the above figure, we may consider that for Strategy 3, the parameter λ possibly plays a more important role than the number of the reference points n. Thus, to obtain an optimal strategy, we may choose a relatively large number of n to set the complexity of the holding function to a certain level, and then only focus on changing the value of λ to get the minimum total risk. Then it becomes much easier and more efficient to compute the optimal strategy. To show the reliability of this way, we fix the number of reference points $n = 100$ and use the one-dimensional (1D) cross validation by only changing the value of λ. After obtaining the best value of the regularization parameter, which is λ^*, at each discounted strike price and hedging frequency, we can get the optimal total risk.

Strike	Cross validation	M (# of hedging times)				
		24	12	6	2	1
90	2D	0.6229	0.8505	1.1332	1.5715	1.7945
	1D	0.6229	0.8506	1.1336	1.5716	1.7945
95	2D	0.7909	1.0901	1.4782	2.1827	2.6380
	1D	0.7911	1.0903	1.4785	2.1828	2.6380
100	2D	0.9352	1.2983	1.7817	2.7710	3.5225
	1D	0.9355	1.2985	1.7819	2.7711	3.5225
105	2D	1.0398	1.4482	2.0054	3.2581	4.3300
	1D	1.0399	1.4484	2.0058	3.2582	4.3300
110	2D	1.1070	1.5369	2.1388	3.5859	4.9525
	1D	1.1071	1.5371	2.1393	3.5859	4.9525

Table 4.21: Comparison of average total risk using 2D and 1D cross validation

The average total risk for Strategy 3 using 1D cross validation with fixed number of reference points $n = 100$ is reported in Table 4.21. With the results, we make a comparison between the best results obtained from the 2D cross validation with the best pairs of (λ^*, n^*) and the 1D cross validation with the best λ^* at $n = 100$. We can find that the values of total risk obtained from the 2D cross validation and the 1D cross validation do not show obvious differences. In some cases, it is even hard to tell the difference of the total risk values between the two types of cross validation.

This result suggests that regularization using λ has effect of choosing the number of reference points. Since the number of the reference points $n = 100$, which is relatively large, some of the reference points are unnecessary. The introduction of the regularization penalty parameter λ could force the coefficients before this kind of reference points to be 0 or very close to 0, which makes the number of reference points that actually work be smaller than n.

Thus, instead of using the 2D cross validation to determine the best pairs of (λ^*, n^*), we may only need to use the 1D cross validation to determine the optimal value of the regularization penalty parameter λ^* with fixed n when solving the total risk minimization problem with regularization.

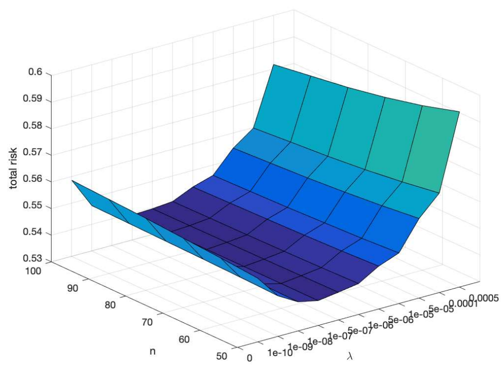
4.4.2 Strategy 4

Finally, we try Strategy 4, which is using the assumption 2.8 and introducing the regularization term to the objective function. The optimization problem to be solved is shown below:

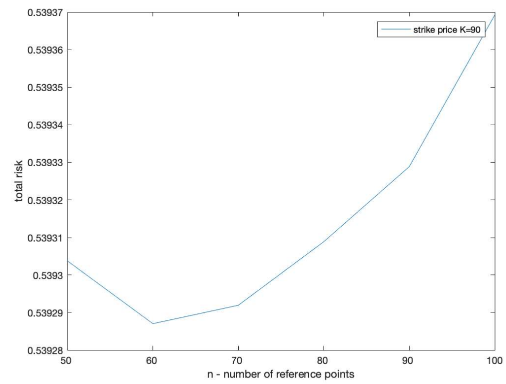
$$\begin{aligned}\min_{V_0, D_j, \tilde{D}_j} \frac{1}{L} \sum_{k=1}^{L} \left(H^{(k)} - V_0 - \sum_{j=0}^{M-1} \left(D_j(X_j^{(k)}) + \sum_{l=0}^{j-1} \tilde{D}_l(X_l^{(k)}) \frac{\Delta X_l^{(k)}}{X_j^{(k)}} \right) \Delta X_j^{(k)} \right)^2 \\ + \lambda \left(\sum_{j=1}^{M-1} \sum_{i=1}^{n} \alpha_{i,j}^2 + \sum_{j=1}^{M-2} \sum_{i=1}^{n} \tilde{\alpha}_{i,j}^2 \right)\end{aligned} \tag{4.5}$$

The settings for Strategy 4 is similar to Strategy 3, where the reference points are evenly located. The range of λ is set to be from 10^{-10} to 10^{-4} and the number of reference points is in the range from 10 to 200. We plot the total risk with respect to different pairs of λ and n to find the best choice of the parameters. Figure 4.9(a) shows the total risk of different pairs of λ and n. For Strategy 4, a smaller number of reference points is needed while the optimal value of λ remains at the same magnitude compared with Strategy 3. The optimal value of n is in the range of $[20, 60]$ while the optimal range for λ is still in $[10^{-5}, 10^{-7}]$. In addition, the optimal value of λ tends to decrease a bit as the number of reference points increases. When the number of reference points is big enough, the value of λ remains at the magnitude of 10^{-7}, which is the same optimal value of the regularization parameter in Strategy 3 when n is large.

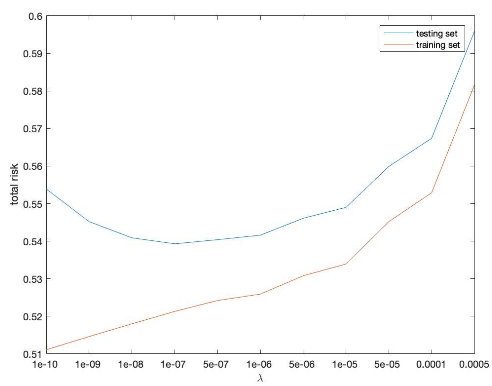
The total risk versus different n for a fixed value of λ is shown in Figure 4.9(a) and the total risk versus different λ when $n = 50, 60$ are shown in Figure 4.9(c) and 4.9(d) respectively.



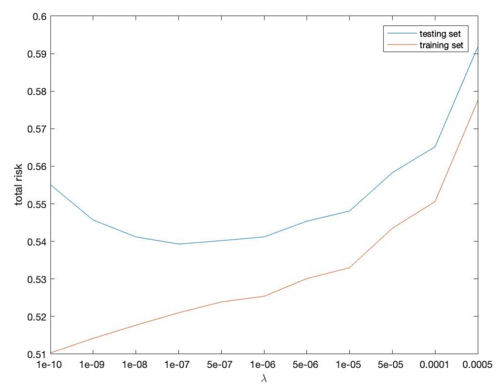
(a) Total risk versus pairs of λ and n



(b) Total risk versus n for $\lambda = 10^{-7}$



(c) Total risk versus λ for $n = 50$



(d) Total risk versus λ for $n = 60$

Figure 4.9: Average total risk for testing set using Strategy 4

Similar to the Strategy 3, the value of the total risk is affected more by the value of λ while the change of n has a much smaller influence. It is hard to tell the difference of the total risk between Figure 4.9(c) and 4.9(d) using $n = 50$ and 60 respectively. However, for Strategy 4, the optimal hedging strategy can be obtained with a smaller number of reference points n compared with Strategy 3. If we take the complexity of computation into consideration, we may still need to roughly choose a number of n based on the complexity of the model although the value of total risk does not increase much as n increases.

With the best choice of the parameter pairs (λ^*, n^*) for each discounted strike price and hedging frequency, we then report the average total risk and total cost for Strategy 4.

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Strategy 2	0.5641	0.7668	1.0498	1.5715	1.7945
	Strategy 4	0.5393	0.7528	1.0446	1.5712	1.7945
95	Strategy 2	0.7100	0.9802	1.3647	2.1823	2.6380
	Strategy 4	0.6939	0.9749	1.3628	2.1814	2.6380
100	Strategy 2	0.8566	1.1853	1.6513	2.7702	3.5225
	Strategy 4	0.8310	1.1743	1.6482	2.7694	3.5225
105	Strategy 2	0.9708	1.3421	1.8745	3.2580	4.3300
	Strategy 4	0.9382	1.3253	1.8680	3.2564	4.3300
110	Strategy 2	1.0362	1.4272	2.0111	3.5858	4.9525
	Strategy 4	1.0153	1.4219	2.0096	3.5854	4.9525

Table 4.22: Average total risk for Strategy 4 with testing data set

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	Strategy 2	2.4630	2.4213	2.3270	2.0556	1.7423
	Strategy 4	2.4613	2.4223	2.3259	2.0559	1.7423
95	Strategy 2	3.9505	3.8987	3.7813	3.4092	2.9704
	Strategy 4	3.9508	3.9006	3.7805	3.4095	2.9704
100	Strategy 2	5.9130	5.8504	5.7177	5.2609	4.6870
	Strategy 4	5.9125	5.8513	5.7183	5.2611	4.6870
105	Strategy 2	8.3555	8.2844	8.1474	7.6256	6.9308
	Strategy 4	8.3539	8.2834	8.1495	7.6258	6.9308
110	Strategy 2	11.2582	11.1835	11.0327	10.4801	9.7067
	Strategy 4	11.2564	11.1820	11.0347	10.4801	9.7067

Table 4.23: Average total cost for Strategy 4 with testing data set

Table 4.22 shows the total risk using Strategy 4 compared with Strategy 2 with the testing data set. From the results, we can see that the total risk obtained from solving the optimization problem with regularization is smaller when the hedging frequency is high. In this case, the hedging strategy with regularization could perform better.

Table 4.23 shows the total cost for Strategy 4 compared with Strategy 2. From the results shown in the above table, we can see that the total costs obtained from Strategy 2 without the regularization and Strategy 4 with the regularization do not show much difference, which leads to a similar conclusion when comparing the total cost from Strategy 3 and Strategy 1.

4.4.3 New Placement of Reference Points

Using the Monte Carlo simulation with the Black Scholes model, the range of the discounted stock price is relatively small when the time t_j is close to 0. However, the best number of the reference points n for Strategy 3 is over 100 and n for Strategy 4 is around 50, which is very large for a narrow range of the discounted stock price. Using a large n when the range of X_j is small may introduce unnecessary complexity and leads to a negative effect for the minimization of the total risk.

In addition, the quadratic total risk minimization problem is robust when changing the number of reference points. Since the value of the total risk is not very sensitive to the number of reference points n, we then see if the total risk is sensitive to the placement of the reference points.

We then try another placement method, where the number of the reference points n follows a increasing trend as the hedging date gets closer to maturity and the width of the discounted stock price range increases. Thus, we put less reference points at the beginning of the stock path while more points near the maturity. By this assumption, we assume the number of the reference points n follows a function of the ith hedging date before the maturity. The number of reference points at t_j is set to be $n = f(j) = kj + 1$, where $k = 1, 2, ...$ representing k times of a certain variable while j is the jth hedging date.

We define this new placement method for the reference points as method b, where we have less reference points at the beginning of the stock path while more near the end, and the former method as method a, where we have a fixed number of reference points at each hedging time.

For Strategy 1 and Strategy 2 that are without regularization, the best choices of k are always $k = 1$. Without regularization, these two strategies only allow very few complexity.

This result is consistent with the best choice obtained before. For Strategy 3 and Strategy 4, we consider different pairs of λ and k. Here, we use Strategy 3 as an illustration. The best choice of k for Strategy 3 when the discounted strike price $K = 90$ and hedging 24 times is around 10, which is $n = f(j) = 10j + 1$. The best λ is 10^{-7}. The results are shown in Figure 4.10.

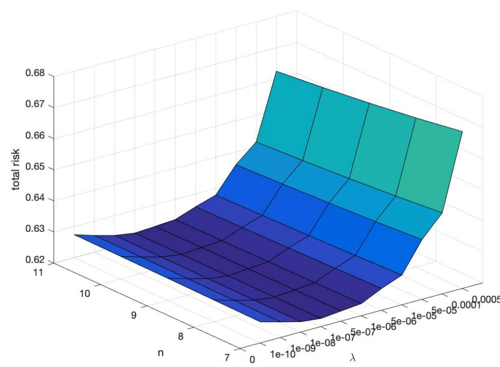
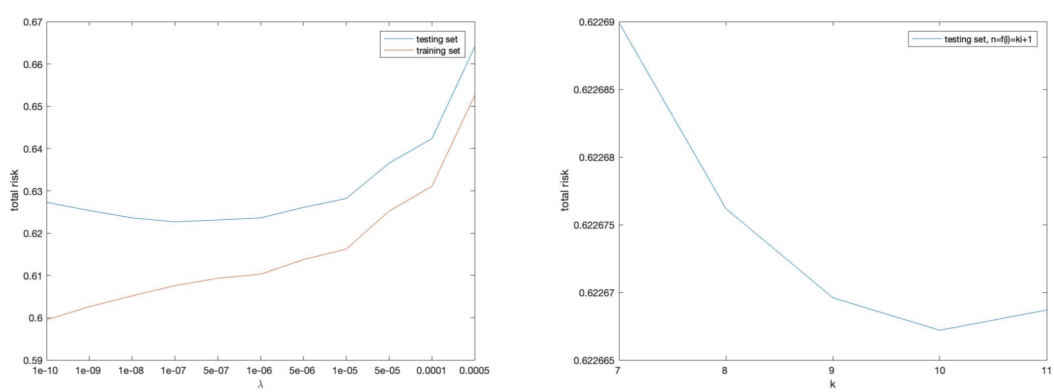


Figure 4.10: Average total risk for testing set with pairs of λ and n using Strategy 3-b



(a) Total risk versus λ for $k = 10$

(b) Total risk versus k for $\lambda = 10^{-7}$

Figure 4.11: Average total risk for testing set using Strategy 3-b

The best results of the total risk and the total cost obtained from this method for Strategy 3 are shown in Table 4.24 and Table 4.25 respectively.

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	3-a	0.6229	0.8505	1.1332	1.5715	1.7945
	3-b	0.6227	0.8503	1.1330	1.5714	1.7945
95	3-a	0.7909	1.0901	1.4782	2.1827	2.6380
	3-b	0.7908	1.0899	1.4781	2.1824	2.6380
100	3-a	0.9352	1.2983	1.7817	2.7710	3.5225
	3-b	0.9351	1.2979	1.7815	2.7709	3.5225
105	3-a	1.0398	1.4482	2.0054	3.2581	4.3300
	3-b	1.0396	1.4477	2.0051	3.2580	4.3300
110	3-a	1.1070	1.5369	2.1388	3.5859	4.9525
	3-b	1.1069	1.5363	2.1384	3.5857	4.9525

Table 4.24: Average total risk for Strategy 3 with new placement of n

Strike	Strategy	M (# of hedging times)				
		24	12	6	2	1
90	3-a	2.4627	2.4198	2.3362	2.0574	1.7423
	3-b	2.4629	2.4205	2.3380	2.0577	1.7423
95	3-a	3.9509	3.8969	3.7785	3.4085	2.9704
	3-b	3.9511	3.8977	3.7796	3.4089	2.9704
100	3-a	5.9107	5.8460	5.7160	5.2602	4.6870
	3-b	5.9108	5.8468	5.7167	5.2607	4.6870
105	3-a	8.3506	8.2768	8.1422	7.6213	6.9308
	3-b	8.3509	8.2777	8.1439	7.6218	6.9308
110	3-a	11.2526	11.1751	11.0366	10.4780	9.7067
	3-b	11.2529	11.1759	11.0383	10.4785	9.7067

Table 4.25: Average total cost for Strategy 3 with new placement of n

For the total risk of testing set which is shown in the above table, we can see that the method b with changing number of the reference points at each hedging time performs very closely to the method a with the fixed number of reference points. This result confirms that the total risk is relatively robust to the reference points when using the spline kernel function. The robustness makes it easier to solve the total risk minimization problem using the spline kernel function by 1D cross validation instead of 2D cross validation. To obtain the optimal hedging strategy, we could introduce the regularization term to the optimization problem and get the minimum total risk by only adjusting the regularization penalty parameter λ.

In terms of the total cost, by using the new placement of the reference points, the method b also have very similar total cost to the method a. A possible advantage to use the method b is that it introduces less complexity than the method a by using less number of the reference points in total and could result in a similar hedging performance with the method a.

Chapter 5

Conclusion

In this project, we use the method proposed by Coleman et al. (2003) that model the relationship between the holding and the underlying asset price with a specific function to solve the quadratic total risk minimization problem. Instead of a cubic spline function, we use spline kernel function to model the relationship and introduce the regularization term to improve the hedging performance. To investigate the out-of-sample performance, we use cross validation where two data sets are generated. The regularization parameter and the number of reference points are chosen by minimizing the total risk of the testing set and the evaluation is based on the average total risk and the average total cost.

By using cross validation, we examine the performance on the testing set, which needs to be considered when implementing in practice. The results obtained in Chapter 4 indicate that using spline kernel function performs better than using cubic spline function. Introduction of regularization allows more model complexity and may lead to a relatively smaller total risk. In addition, the total risk shows robustness with respect to the number of reference points to some extent. After adding the regularization term, the optimal hedging strategy can be obtained by only changing the value of the regularization parameter, which is easier to implement. However, there is no obvious improvement for the total cost after introducing the regularization term. Sometimes, the reduction of total risk leads to some rise in the total cost.

In this chapter, we would like to discuss some technical issues when implementing the method and some future work that can be further examined for the total risk minimization problem.

5.1 Technical Issues

When solving the quadratic total risk hedging optimization problem, there may exist some technical issues related to the implementation.

One problem is that after introducing the regularization term to the total risk minimization problem, we use a relatively large number of the reference points. Since we use 40000 simulations of the stock price paths, rise in the number of the reference points leads to a significant increase in the dimension of the optimization problem. Solving such a problem is very time-consuming and may not be efficient with a large size of data.

Another problem is that we use MATLAB to solve the optimization problem that the result is sometimes limited by the calculation ability of MATLAB. In practice, the frequency of hedging may be higher than what we use in this project. If we would like to increase the number of simulations and the hedging frequency, the size of the input matrix may exceed the limit of the matrix size in this programming environment.

5.2 Future Work

For the total risk minimization problem with the spline kernel function discussed in this project, there is some potential work that can be examined in the future.

In this project, we assume that the hedging position is a function of the discounted stock price. However, in the real market, the stock price may not be the only factor that affects the hedging position. Other features like the volatility of the underlying price may also play a important role when hedging the intrinsic risk of an option. We could also use other models to simulate the relationship between the holding and some features. When introducing more features, we may use the neural network or other machine learning methods. In addition, the framework discussed in this project can be applied to the piecewise linear risk minimization problem and the L_2 regularization can also be changed to other common used methods like the L_1 regularization.

On the other hand, the way to generate the data set can be changed. The Black Scholes model used in this project is very basic to model the asset price. Other models that considers the jump risk or changing volatility may bring a different hedging performance. In addition, it is possible to consider historical market price as training and testing data sets.

References

[1] Dimitris Bertsimas, Leonid Kogan, and Andrew W Lo. Hedging derivative securities and incomplete markets: an -arbitrage approach. *Operations research*, 49(3):372–397, 2001.

[2] Thomas F Coleman, Yuying Li, and Maria-Cristina Patron. Discrete hedging under piecewise linear risk-minimization. Technical report, Cornell University, 2003.

[3] Thomas F Coleman, Yuying Li, and Maria-Cristina Patron. Total risk minimization using monte carlo simulations. *Handbooks in Operations Research and Management Science*, 15:593–635, 2008.

[4] Thomas F Coleman, Yuying Li, and Cheng Wang. Stable local volatility function calibration using spline kernel. *Computational optimization and applications*, 55(3):675–702, 2012.

[5] Nicole El Karoui and Marie-Claire Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM journal on Control and Optimization*, 33(1):29–66, 1995.

[6] Hans Föllmer and Martin Schweizer. Hedging by sequential regression: An introduction to the mathematics of option trading. *ASTIN Bulletin: The Journal of the IAA*, 18(2):147–160, 1989.

[7] Rüdiger Frey. Derivative asset analysis in models with level-dependent and stochastic volatility. In *CWI quarterly*. Citeseer, 1997.

[8] David Heath, Eckhard Platen, and Martin Schweizer. A comparison of two quadratic approaches to hedging in incomplete markets. *Mathematical Finance*, 11(4):385–413, 2001.

[9] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[10] Manfred Schäl. On quadratic cost criteria for option hedging. *Mathematics of operations research*, 19(1):121–131, 1994.

[11] Martin Schweizer. Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20(1):1–32, 1995.

[12] Martin Schweizer. A guided tour through quadratic hedging approaches. Technical report, Discussion Papers, Interdisciplinary Research Project 373: Quantification , 2001.

[13] Vladimir N Vapnic. Statistical learning theory. *A Wiley-Interscience Publication*, 1998.