

# STUDIES ON OPTIMAL TRADE EXECUTION

TARDU SELIM SEPIN

A DISSERTATION

PRESENTED TO THE FACULTY  
OF PRINCETON UNIVERSITY  
IN CANDIDACY FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF  
OPERATIONS RESEARCH & FINANCIAL ENGINEERING  
ADVISER: PROFESSOR PATRICK CHERIDITO

JANUARY 2015

UMI Number: 3682773

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3682773

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

© Copyright by Tardu Selim Sepin, 2015.

All rights reserved.

# Abstract

This dissertation deals with the question of how to optimally execute orders for financial assets that are subject to transaction costs. We study the problem in a discrete-time model where the asset price processes of interest are subject to stochastic volatility and liquidity.

First, we consider the case for the execution of a single asset. We find predictable strategies that minimize the expectation, mean-variance and expected exponential of the implementation cost.

Second, we extend the single asset case to incorporate a dark pool where the orders can be crossed at the mid-price depending on the liquidity available. The orders submitted to the dark pool face execution uncertainty and are assumed to be subject to adverse selection risk. We find strategies that minimize the expectation and the expected exponential of the implementation shortfall and show that one can incur less costs by also making use of the dark pool.

Next chapter studies a multi asset setting in the presence of a dark pool. We find strategies that minimize the expectation and expected exponential of a cost functional that consists of the implementation shortfall and an aversion term that penalizes the orders crossed in the dark pool. In the expected exponential of the cost case, the dimensionality of the problem does not allow for the numerical computation of optimal strategies. Therefore, we first solve the expected exponential case for a second order Taylor approximation and then provide a framework via a duality argument which can be used to generate approximate strategies.

Lastly, we treat the case where the single asset execution problem exhibits ambiguity for the distribution of stochastic liquidity and volatility. We see the implementation cost as the sum of risk terms arising at each execution period. We consider the problem obtained from aggregating worst case expectations of these risk terms, by penalizing the distributions used with dynamic indicator, relative entropy and

Gini indices. Next, we formulate the problem as the multi-prior first order certainty equivalent of the exponential cost and lastly we consider a second order certainty equivalence formulation.

## Acknowledgements

I am indebted to my adviser Professor Patrick Cheridito. He has been instrumental in my academic endeavors at Princeton with the support he has given me. Without his help this thesis would not have been possible. I have been very lucky to have had the chance to work with a scholar of his academic stature and to learn from him. He has been a true role model with the example he has set.

I am very thankful to Michael Sotiropoulos for many fruitful discussions about optimal execution. I had the privilege of enjoying the friendship of many brilliant people during my doctoral studies. I am thankful to Aykut Ahlatçioğlu, Çağın Ararat, Emre Barut, Quentin Berthet, Andres Maggi, Chintan Mehta, Onur and Özlem Özyeşil, Nicolas Pègard, Patrick Rebeschini, Lorenzo Reus, Juan Sagredo, Xin Tong, Firdevs Ulus, Kevin Webster, Alexander Wugalter, and Zhikai Xu for their friendship.

Most of all, I am indebted to my parents Sibel and Yavuz, and to my sister Tuğçe.

To Sibel, Tuğçe and Yavuz.

# Contents

Abstract . . . . .	iii
Acknowledgements . . . . .	v
List of Tables . . . . .	ix
List of Figures . . . . .	x
<b>1 Introduction</b>	<b>1</b>
1.1 Macro level execution models . . . . .	3
1.2 The Almgren–Chriss model . . . . .	6
<b>2 Single–asset execution</b>	<b>9</b>
2.1 The model . . . . .	11
2.2 Risk neutral objective . . . . .	14
2.3 Expected exponential cost . . . . .	16
2.4 Mean-variance criterion . . . . .	21
2.4.1 The restricted mean-variance problem . . . . .	23
2.4.2 The full mean-variance problem . . . . .	28
2.5 Conclusion . . . . .	30
<b>3 Single–asset execution with a dark pool</b>	<b>33</b>
3.1 The model . . . . .	36
3.2 Risk neutral objective . . . . .	37
3.3 Expected exponential cost . . . . .	41



3.4	Conclusion . . . . .	43
<b>4</b>	<b>Multi-asset execution with a dark pool</b>	<b>47</b>
4.1	The model . . . . .	49
4.2	Parameter set for simulations . . . . .	52
4.3	Risk neutral objective . . . . .	55
4.4	Expected exponential cost . . . . .	58
4.4.1	Second order Taylor approximation . . . . .	61
4.4.2	An approximation via duality . . . . .	64
4.5	Conclusion . . . . .	70
<b>5</b>	<b>Robust execution of a single-asset</b>	<b>72</b>
5.1	The model . . . . .	74
5.2	Aggregation with dynamic ambiguity indices . . . . .	75
5.2.1	Indicator index . . . . .	75
5.2.2	Relative entropy index . . . . .	77
5.2.3	Gini index . . . . .	79
5.3	First order certainty equivalent . . . . .	80
5.4	Second order certainty equivalent . . . . .	81
<b>A</b>		<b>83</b>
A.1	Proofs for Chapter 2 . . . . .	83
A.2	Proofs for Chapter 3 . . . . .	87
A.3	Proofs for Chapter 4 . . . . .	89
A.4	Proofs for Chapter 5 . . . . .	97
	<b>Bibliography</b>	<b>107</b>

# List of Tables

2.1	Parameter values . . . . .	14
4.1	Parameter values . . . . .	55

# List of Figures

1.1	Unaffected price process . . . . .	7
1.2	Price dynamics under permanent and temporary impact . . . . .	7
2.1	Average and standard deviation of the optimal positions for different scenarios in the risk neutral case. . . . .	16
2.2	Optimal strategies for three particular scenarios in the risk neutral case compared to the constant speed strategy. . . . .	17
2.3	Histograms of $C(x)$ for the optimal strategy and the constant speed strategy in the risk neutral case. . . . .	18
2.4	Average and standard deviation of optimal positions for the expected exponential cost criterion with different absolute risk aversions $\alpha$ . . . . .	20
2.5	Three realizations of the optimal strategy for the expected exponential criterion with parameter $\alpha = 2 \times 10^{-5}$ compared to the Almgren–Chriss strategy (1.2.4) corresponding to $\bar{\sigma}$ , $\bar{\eta}$ and parameter $\lambda = 10^{-5}$ . . . . .	21
2.6	Histogram of $C(x)$ for the optimal strategy and the Almgren–Chriss strategy (1.2.4) corresponding to $\bar{\sigma}$ and $\bar{\eta}$ in the expected exponential cost case with absolute risk aversion $\alpha = 2 \times 10^{-5}$ . . . . .	22
2.7	Average and standard deviation of $\mathcal{R}$ –optimal positions for the mean-variance criterion with different values of $\lambda$ . . . . .	26

2.8	Three realizations of the optimal strategy for the mean-variance criterion with parameter $\lambda = 10^{-5}$ compared to the Almgren–Chriss strategy (1.2.4) corresponding to $\bar{\sigma}$ and $\bar{\eta}$ . . . . .	26
2.9	Mean-variance criterion for optimal solutions to $P(\lambda, \mu)$ for $\lambda = 10^{-5}$ and different values of $\mu$ . The minimum is attained for $\mu^* = 0.968$ , and $1 - 2\lambda\mathbb{E}[Q(x^{\mu^*})] = 0.997$ . . . . .	27
2.10	Histogram of $C(x)$ for the optimal restricted strategy in the mean-variance case for $\lambda = 10^{-5}$ . . . . .	27
2.11	Mean-variance criterion for optimal solutions to $\hat{P}(\lambda, \mu)$ for $\lambda = 10^{-5}$ and different values of $\mu$ . The minimum is attained for $\mu^* = 1.192$ , and $1 - 2\lambda\mathbb{E}[Q(x^{\mu^*})] = 0.998$ . . . . .	30
2.12	Histograms of $C(x)$ for the optimal strategy and the Almgren–Chriss strategy (1.2.4) in the mean-variance case with constant volatility $\bar{\sigma}$ and liquidity $\bar{\eta}$ for $\lambda = 10^{-5}$ . . . . .	31
3.1	Figures for average positions $\bar{x}_n, \bar{y}_n$ over different scenarios in the risk neutral case for the optimal strategies of Theorem 3.2.1, 2.2.1 and constant speed strategy. . . . .	40
3.2	Long term averages of $d_{n-1} + e_{n-1}$ and trajectory of three different strategies under a realization in the risk neutral case. . . . .	40
3.3	Histograms of $C(y, z)$ for the optimal strategy and the strategy in the risk neutral case of Chapter 2. . . . .	44
3.4	Figure on top shows average positions $\bar{x}_n$ over different scenarios in the expected exponential case for the optimal strategies of Theorem 3.3.1, of Theorem 2.3.1 with $\alpha = 10^{-5}$ and risk neutral optimal strategy of Theorem 3.2.1. Figure at the bottom shows average positions $\bar{x}_n$ for the optimal strategy of Theorem 3.3.1 for three different parameter values of $\alpha$ . . . . .	45

3.5	Histograms of $C(y, z)$ for the optimal strategies in the expected exponential cost case under the restricted optimal strategy of Theorem 3.3.1 and Theorem 2.3.1 with $\alpha = 10^{-5}$ . . . . .	46
4.1	Average of net optimal positions for different scenarios and a realization of execution trajectories in the risk neutral case with $\alpha = 0$ . . . . .	59
4.2	Detailed analysis of the negatively correlated pair Company A and B shows the net average stock positions, and average of order sizes submitted into the regular exchange and the dark pool as percentages of initial positions in the assets for the risk neutral case with $\alpha = 0$ . . . . .	60
4.3	Average size of orders submitted to the regular exchange and to the dark pool for Company A as a function of the parameter $\alpha$ in the risk neutral case. . . . .	60
4.4	Histograms of $C(x)$ computed for the risk neutral objective under the optimal strategy in the multi-asset case with a dark pool where $\alpha = 0$ and the aggregation of two individual single-asset cases from from Theorem 2.2.1 without a dark pool. . . . .	61
4.5	Average of net optimal positions for different scenarios as a function of risk aversion parameter $\lambda$ , and a realization of execution trajectories in the second order Taylor approximation case with $\alpha = 0$ . . . . .	65
4.6	Detailed analysis of the negatively correlated pair Company A and B shows the net average stock positions, and average of order sizes submitted into the regular exchange and the dark pool as percentages of initial positions in the assets for the second order Taylor approximation of the expected exponential cost case with $\alpha = 0$ and $\lambda = 10^{-4}$ . . . . .	66

4.7	Standard deviation from the average execution trajectories and average size of orders submitted to the dark pool for Company A as a function of the parameter $\alpha$ in the second order Taylor approximation case with $\lambda = 10^{-4}$ . . . . .	67
4.8	First histogram of $C(x)$ is computed for the strategy minimizing the expectation of the second order exponential of the cost in the multi-asset case with a dark pool where $\lambda = 10^{-4}$ and $\alpha = 0$ . Second histogram of $C(x)$ is computed by separately finding two individually optimal strategies minimizing the expected exponential of the cost in the single-asset case without a dark pool and where $\lambda = 10^{-4}$ . . . . .	68

# Chapter 1

## Introduction

The foundation of modern day equities trading has been floor trading, a system which has been established about a century ago. A typical trading process would start by a customer placing an order to his/her brokers order taker which would then be passed on to a runner who would find the brokers floor trader on the trading floor. If the floor trader can get a fill for the order then this information would be rushed back to the customer and the clearing and settlement process would start. In todays markets there is no need for runners and floor traders anymore. The shift from the physical form of trading to electronic trading has replaced runners and floor traders with electronic communication networks. These networks made trading substantially easier and more transparent. Anyone can have access to all the bid and ask information, order sizes and to the history of all such quantities. These networks ensure that one gets feedback for his/her orders in fractions of seconds.

The technological progression of markets allowed participants to be more quantitative about how one should trade in an exchange hence it also changed the way the brokers deal in the markets. Although, the mechanism of trading introduced by floor trading fundamentally still remains the same, today the brokers are like an interface that allows customers to interact with markets through algorithms which

are designed to execute trades on behalf of the customer. It is common to measure the performance of the broker against some benchmark performance criteria. At the core of most performance criteria lies the transaction costs of trading. When speaking of transaction costs we understand the unfavorable price movement for the trade that is being executed. The transaction costs arise from shifts in the demand and supply. When large orders are placed into the order book, they reveal information to other market participants and therefore the prices of securities move unfavorably for the broker and hence the customer. A typical practice for the broker would start by receiving an execution order from an institutional investor. The deadline until which the execution must be completed would be specified which typically is not longer than a day and is usually around 1–2 hours. Then a benchmark performance criterion would be selected or specified by the institutional investor. At this point the broker or namely the execution services provider first selects a macro level algorithm, known as the slicer, to decide on a trading schedule. The slicer decides how many assets one should trade in a certain amount of time with respect to the performance criteria that the broker is benchmarked against. Secondly, the broker has a tactical algorithm which works on the limit order book level. The tactical algorithm uses the output of the slicer and tries to get the best price for the assets being traded by deciding how to place limit and market orders across different venues.

This dissertation focuses on the question of how to optimally execute orders for financial assets at a macro algorithm level. In the rest of this chapter, we first briefly introduce some of the most commonly used execution models by reviewing the available literature. Then we discuss the Almgren–Chriss execution model in detail. Chapter 2 studies the optimal execution of a single–asset in a market with stochastic liquidity and volatility. Chapter 3 extends on the model for the single asset market by including a dark pool; an exchange that hides the available liquidity and where the orders can be crossed at the mid–price. Chapter 4 treats the case when one executes



orders for a portfolio of assets in a setting with stochastic liquidity, volatility and a dark pool. Chapter 5 deals with the question of finding a robust execution strategy that is optimal for a single-asset when there is ambiguity for the distributions of liquidity and volatility processes. All proofs are given in the Appendix.

## 1.1 Macro level execution models

The transaction costs have mostly been associated with the rate of trading. When one trades a large volume in a short amount of time, it translates into consuming the liquidity and it may result in high transaction costs. Therefore, in an execution model one major component is how one models the impact of submitted orders on the asset price. We should note that in the execution models that work on the macro level it is assumed that price impact is exogenous, where as in the models that work on the micro level the price impact is observed to be the natural consequence of evolution of arriving orders on both the bid and ask side of the limit order book. One strand of literature works with models which assume the market impact has two components. One of these component is assumed to have a temporary impact and the other a permanent impact on the asset price. In a fundamental paper, Bertsimas and Lo (1998) address an optimal execution problem using a model where the asset price process follows a discrete-time random walk and transactions in the asset cause a permanent price impact that is linear in the amount of shares traded. In an other fundamental paper, Almgren and Chriss (2001) add a temporary price impact to this discrete-time model from which the asset recovers in the next period. They assume that the temporary impact is also linear in the amount of transactions in the asset. We classify these type of models as linear temporary impact models. In the same setting Almgren (2003) studies the optimal execution problem for nonlinear impact functions. The other strand of literature assumes that price impact is transient, which

means that the impact of submitted orders will decay over time. Such a framework was first proposed by Obizhaeva and Wang (2013). When using such a framework one usually assumes a continuous density for the depth of order book, which helps to model the impact of trading, and then a decay kernel dictating the time it takes for the order book to recover from the price impact is specified. Obizhaeva and Wang (2013) assume a constant density for the depth of the order book and use an exponential decay kernel to model the order book recovery. Alfonsi et al. (2010) generalized this framework by allowing the depth of the order book to vary as a function of the price. Alfonsi et al. (2012) and Gatheral et al. (2012) have generalized it further by using different decay kernels and order book densities.

Perold (1988) defines the cost of execution through implementation shortfall. Implementation shortfall is the difference between the initial monetary value of the position that one holds in a financial asset and the money collected from the sale of the asset that is being liquidated. Therefore, it measures how costly the execution has been due to exposure to adverse selection. Cost functionals used in the literature are mostly built around the implementation shortfall. Bertsimas and Lo (1998) is an example where the expected implementation shortfall is minimized. On the other hand, Almgren and Chriss (2001) minimize a mean-variance criterion of the implementation shortfall. Both of these models assume constant volatility and price impact in a setting where price process has an additive market innovation component. Therefore, in the case of Bertsimas and Lo (1998) one loses the noise term as the innovations have mean zero; where as, in the case of Almgren and Chriss (2001) the volatility risk is captured in the objective function that is being minimized because of the variance term. It turns out the cost minimizing strategy of Bertsimas and Lo (1998) is a constant speed sell strategy where one sells an equal amount of shares in the asset at each period until the end of the execution. However, the deterministic mean-variance minimizing strategy of Almgren and Chriss (2001) sells faster than

the constant speed strategy because of the aversion to volatility risk. Schied and Schöneborn (2009) have associated the speed of the optimal execution strategy to the absolute risk aversion of the objective function that is being optimized. They have shown that the speed of selling increases as the absolute risk aversion increases. Forsyth (2011) and Gatheral and Schied (2011) are examples of a continuous-time model in which the unaffected price process is assumed to be a geometric Brownian motion. One rather surprising result is by Schied et al. (2010) who studied the optimal execution problem for an agent with expected exponential utility of the implementation shortfall in a continuous-time model with Lévy noise. They found that in the analog of the Almgren-Chriss model the optimal strategy is of the same form as the deterministic best mean-variance strategy even if one looks for a solution in the class of all adapted strategies. A deterministic strategy usually turns out to be optimal in the class of all adapted strategies if one works with a price process that consists of additive noise and non-random price impact, and the cost functional does not have a risk term that depends on the price process.

Among the models that account for stochastic price impact and volatility is Walia (2006) where it is assumed that the volatility and liquidity processes follow a coupled Markov chain and they minimize one-step mean-variance criteria. In another contribution, Almgren (2012) uses the continuous-time analog of this model where the volatility and liquidity follow correlated diffusion processes. To incorporate stochastic impact Bayraktar and Ludkovski (2011) and Moazeni et al. (2013) model the price impact with a Poisson process. When one assumes stochastic dynamics for the liquidity process the dynamically optimal strategy will no longer be deterministic. Makimoto and Sugihara (2010) have studied a portfolio execution problem in a setting by allowing the decay kernel used by Obizhaeva and Wang (2013) to be stochastic. We refer to Gatheral and Schied (2013) for a more detailed survey of models used in optimal execution.

## 1.2 The Almgren–Chriss model

The market impact model proposed by Almgren and Chriss (2001) is the model we extend on in the following chapters, therefore, we present it in detail. We consider the problem of liquidating an asset position of  $X \in \mathbb{R}_+$  shares until a given time  $T \in \mathbb{R}_+$ . We divide the interval  $[0, T]$  into  $N$  subintervals of length  $\Delta t$  and decide at every time  $t_{n-1} = (n-1)\Delta t$  how many shares  $y_n$  to sell in the interval  $(t_{n-1}, t_n]$ . We assume that the unaffected price process of the asset is given by

$$S_n^0 = S_{n-1}^0 + \sigma\sqrt{\Delta t}\xi_n,$$

$(\xi_n)$  is a sequence of standard normals and  $\sigma$  is the volatility of the unaffected stock price process. When a sell order of  $y_n$  is submitted it is assumed that the asset price is depressed linearly in the amount of shares  $y_n$ . We further assume that there are two components of the price impact that  $y_n$  has on the asset. One component depresses the price process permanently hence for a constant  $c \in \mathbb{R}_+$  describing a permanent price impact, the price process while executing an order is described by

$$\begin{aligned} S_n &= S_n^0 - c \sum_{i=1}^n y_i \\ &= S_{n-1} + \sigma\sqrt{\Delta t}\xi_n - cy_i. \end{aligned} \tag{1.2.1}$$

since  $S_0^0 = S_0$ . We can think of  $S_n$  as a fixed convex combination of the bid and ask price:  $S_n = \lambda S_n^b + (1-\lambda)S_n^a$ . For  $\lambda = 1/2$ ,  $S_n$  is the mid price. The temporary component affects the execution price that one gets for the order  $y_n$  and its effect is assumed to vanish in the next period. We assume the resulting execution price is given by

$$\tilde{S}_n = S_{n-1} - \eta y_n, \tag{1.2.2}$$

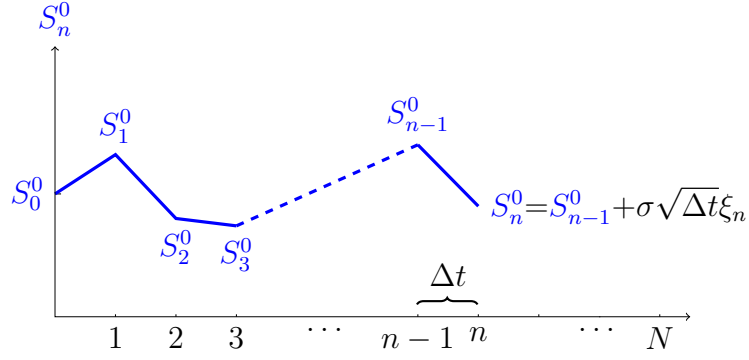


Figure 1.1: Unaffected price process

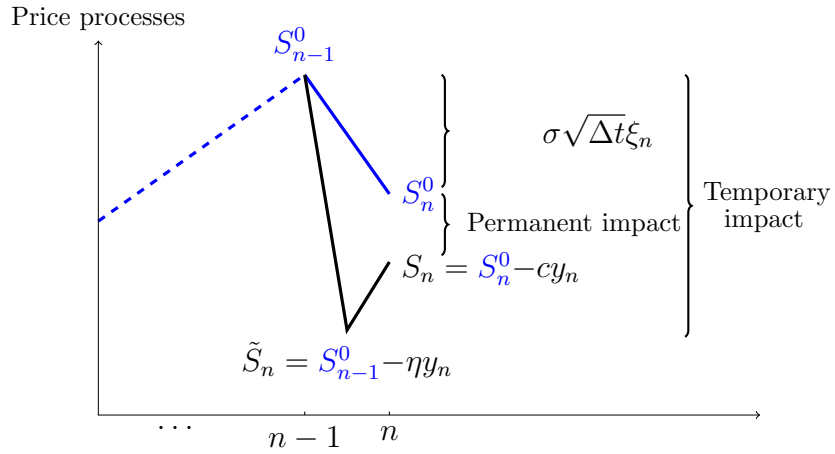


Figure 1.2: Price dynamics under permanent and temporary impact

where  $\eta \in \mathbb{R}_+$  is a constant modeling the temporary price impact. Figures 1.1 and 1.2 depict the price processes we described. In Figure 1.2 it is assumed that  $y_i = 0$  for all  $i \leq n - 1$  and  $y_n > 0$ .

Then the proceeds from selling the asset shares will be  $\sum_{n=1}^N y_n \tilde{S}_n$ . We can describe an execution strategy also in terms of remaining shares  $x_n = X - \sum_{i=1}^n y_i$ . Furthermore, we require:

$$x_0 = X, \quad x_{n-1} \geq x_n, \quad x_N = 0,$$

This means that each  $y_n$  must be non-negative and  $\sum_{n=1}^N y_n = X$ . An execution strategy is completely specified by  $x_n$  for  $1 \leq n \leq N - 1$ . Let us denote the set of all

such strategies by  $\mathcal{D}$ . The implementation cost of a strategy  $x \in \mathcal{D}$  is the difference between the initial value and the proceeds:  $C(x) := XS_0 - \sum_{n=1}^N y_n \tilde{S}_n$ . Since  $x_N = 0$ , it can be written as

$$C(x) = \frac{cX^2}{2} + \sum_{n=1}^N (x_{n-1} - x_n)^2 \left( \eta - \frac{c}{2} \right) - x_n \sigma \sqrt{\Delta t} \xi_n.$$

The goal in Almgren and Chriss (2001) is to find  $x \in \mathcal{D}$  such that for a parameter  $\lambda \in \mathbb{R}_+$  the mean–variance trade–off given by

$$\mathbb{E}[C(x)] + \lambda \text{Var}(C(x)) = \sum_{n=1}^N (x_{n-1} - x_n)^2 \left( \eta - \frac{c}{2} \right) + \lambda \sigma^2 \Delta t x_n^2, \quad (1.2.3)$$

is minimized. Assume that  $\eta - c/2 > 0$ , then one can find the mean–variance optimal deterministic strategy simply by a first order condition since the expression in (1.2.3) is convex in  $x$  and the solution obtained in such a way turns out to be in  $\mathcal{D}$ . The first order condition reads as

$$x_n \lambda \sigma^2 \Delta t = (x_{n-1} - 2x_n + x_{n+1})(\eta - c/2), \quad n = 1, \dots, N - 1,$$

and the solution to these set of of linear difference equations computes as,

$$x_n^* = X \frac{\sinh(\kappa(T - n\Delta t))}{\sinh(\kappa T)}, \quad (1.2.4)$$

for the unique  $\kappa > 0$  satisfying  $\cosh(\kappa \Delta t) - 1 = \frac{\lambda \sigma^2 \Delta t}{2\eta - c}$ .

# Chapter 2

## Single–asset execution

Using the execution model proposed by Almgren and Chriss (2001) we consider an execution problem under stochastic volatility and temporary price impact by following Cheridito and Sepin (2014). We consider three different objective functions. First is the expectation of the implementation cost. The second one is the expected exponential and the last one is the mean–variance trade–off of the implementation cost. Our focus is on finding predictable strategies that minimize the objective functions we consider. For finding predictable strategies we assume that prices, volatility and liquidity are observable. In the frameworks we consider, volatility and liquidity are allowed to be dependent. Moreover, the strategy that minimizes the expectation and a mean–variance of the implementation cost is also allowed to depend on the market innovations of the price process. We use a Markovian setting. This assumption allows us to focus on the last observed state of the process instead of the whole history and hence makes the models we consider tractable.

The predictable strategy minimizing the expected implementation cost is the solution of a forward–backward system of stochastic equations. The solutions we obtain in this case generalize the constant speed strategy of Bertsimas and Lo (1998) by changing the position in the asset at every step by a fraction that depends on condi-

tional expectations of future liquidity terms. Since the term that consists of stochastic volatility and market innovations disappears when one takes the expectation of the implementation cost, they are observed to affect the optimal strategy only through their co-dependence with the liquidity. In the case of the expected exponential cost and the mean-variance criterion, we assume that the volatility and liquidity follow a coupled Markov chain that is independent of the price innovations. For the expected exponential case we deduce a Bellman equation that can be solved numerically. We find that it is enough for the optimal strategy to observe the realizations of volatility and liquidity. In the case of a mean-variance criterion the dynamic programming principle cannot be directly applied to obtain a predictable strategy since the criterion does not break into sub-problems that optimally connect to each other. Moreover, the optimal strategy depends on past realizations of volatility, liquidity and price innovations. To simplify the problem, we first consider strategies that are restricted to observing volatility and liquidity only. After that we solve the full problem in which strategies can react to changing volatility, liquidity and prices. In both cases we relate the execution problem to a quadratic cost minimization problem and derive a Bellman equation that can be solved by discretizing the control space from which the mean-variance optimal strategy can be obtained. The numerical experiments show that the mean-variance optimal strategy produces implementation costs that have a lower sample mean and standard deviation than the Almgren-Chriss strategy corresponding to the long term time-averages of volatility and liquidity. The mean-variance formulation we study is related to the optimal strategies found in Walia (2006) and Almgren (2012). But our strategy minimizes a mean-variance objective of the total implementation cost as opposed to a local mean-variance criterion. In addition, the mean-variance optimal strategy that we characterize by observing the prices generalizes Almgren and Lorenz (2007, 2011), where they investigate strategies that react to price changes, to a stochastic volatility and liquidity setting. In the numerical



experiments we do we observe that the optimal strategy for the expected exponential criterion generates outcomes that are virtually indistinguishable from those of the best mean-variance strategy if the risk aversion parameters are chosen accordingly. Moreover, it is simpler to solve than the mean-variance problem.

The structure of this chapter is as follows. In Section 2.1 we introduce our model of price behavior and trading impact. In Section 2.2 we study the risk neutral case. Section 2.3 treats the expected exponential case, and the mean-variance case is investigated in Section 2.4. Section 2.5 concludes. We give all the proofs in Appendix A.1.

## 2.1 The model

We use the same notation as in Section 1.2 and consider the same problem of liquidating an asset position of  $X \in \mathbb{R}_+$  shares on or before a given time  $T \in \mathbb{R}_+$ . A sell order of size  $y_n$  results in an execution price of

$$\tilde{S}_n = S_{n-1} - \eta_n y_n, \tag{2.1.1}$$

where  $S_n$  follows the dynamics

$$S_n = S_{n-1} + \sigma_n \sqrt{\Delta t} \xi_n - c y_n. \tag{2.1.2}$$

$(\xi_n)$  is again a sequence of standard normals, but here  $(\sigma_n)$  is a stochastic volatility and  $(\eta_n)$  is a stochastic liquidity process modeling the stochastic temporary price impact. We suppose that  $S_n, \sigma_n, \eta_n$  are observable and define the filtration that they generate as

$$\mathcal{F}_n := \sigma(S_i, \sigma_i, \eta_i : -\infty < i \leq n).$$

Notice that  $S_n$  can be observed directly by looking at the quoted price in the exchange and  $\tilde{S}_n$  is observable to the user of the execution program since the proceeds from the sale are observable. Therefore, if  $y_n > 0$ ,  $\eta_n$  can be deduced from (2.1.1). In the case when  $y_n = 0$ , it has to be inferred by using statical procedures (i.e. by considering the observable quantities such as the bid-ask spread and order book imbalance).  $\sigma_n$  has to be estimated from (2.1.2) or by a separate statistical procedure that uses more frequent observations.

We describe an execution strategy in terms of remaining shares ( $x_n$ ) and say that it is admissible if:

$$x_0 = X, \quad x_{n-1} \geq x_n, \quad x_N = 0, \quad \text{and } (x_n) \text{ is predictable with respect to } (\mathcal{F}_n).$$

This means that we require the strategy to be non-increasing hence each  $y_n$  must be non-negative. Moreover, asset position  $x_n$  and  $y_n$  are  $\mathcal{F}_{n-1}$ -measurable as well as  $\sum_{n=1}^N y_n = X$ . An admissible strategy can be completely specified by  $x_n$  for  $1 \leq n \leq N - 1$  and by  $\mathcal{A}$  we denote the set of all admissible strategies. The implementation shortfall of an admissible strategy  $x$  reads as

$$C(x) = \frac{cX^2}{2} + \sum_{n=1}^N (x_{n-1} - x_n)^2 \left( \eta_n - \frac{c}{2} \right) - x_n \sigma_n \sqrt{\Delta t} \xi_n.$$

Since dropping the constant  $cX^2/2$  from  $C(x)$  does not change the optimality of the solutions we obtain in the objective functions we consider, we work with the quadratic form

$$Q(x) = \sum_{n=1}^N (x_{n-1} - x_n)^2 \left( \eta_n - \frac{c}{2} \right) - x_n \sigma_n \sqrt{\Delta t} \xi_n. \quad (2.1.3)$$

We always assume that  $\xi_n$  is independent of  $\sigma(\mathcal{F}_{n-1}, \sigma_n, \eta_n)$  and  $\tilde{\eta}_n := \eta_n - c/2 > 0$ . In the case when we minimize the expectation of the cost  $Q(x)$  we can assume slightly generalized dynamics for  $(S_n, \sigma_n, \eta_n)$ . To do this we fix a non-negative integer  $k$

and for some measurable function  $\varphi : \mathbb{R} \rightarrow \{1, \dots, k\}$  we define  $\varphi_n := \varphi(\sigma_n \xi_n)$ . This process allows us to incorporate information about the market innovations. For example one can set  $\varphi(s) = 1$  for  $s < 0$  and  $\varphi(s) = 2$  for  $s \geq 0$  to account for the sign of innovation terms. We suppose that  $(\sigma_n, \eta_n)$  takes finitely many and different values in  $V \subset \mathbb{R}_+^2$ , and conditional on  $\mathcal{F}_{n-1}$ , its distribution only depends on  $(\sigma_{n-1}, \eta_{n-1}, \varphi_{n-1})$ . Therefore,  $(\sigma_n, \eta_n, \varphi_n)$  is a Markov chain with the finite state space  $V^k \subseteq \mathbb{R}_+^2 \times \{1, \dots, k\}$ . We assume that it has time-dependent transition probabilities given by

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n, \varphi_n) = w \mid (\sigma_{n-1}, \eta_{n-1}, \varphi_{n-1}) = v], \quad v, w \in V^k.$$

These dynamics can be used in models where  $(\sigma_n, \eta_n)$  may depend on the last price innovation. In Sections 2.3 and 2.4 we assume that the couple  $(\sigma_n, \eta_n)$  takes values in the state space  $V$  which does not depend on the price innovations. Hence we define the transition probabilities for this case as,

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v], \quad v, w \in V.$$

For our numerical simulations we use the parameter values of Table 1 and assume that  $(\sigma_n)$ ,  $(\eta_n)$ ,  $(\xi_n)$  are independent, and  $(\sigma_n)$ ,  $(\eta_n)$  are time-homogeneous Markov chains with transition matrices given by

$$p^\eta = \begin{pmatrix} 0.50 & 0.30 & 0.20 \\ 0.15 & 0.80 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix} \quad \text{and} \quad p^\sigma = \begin{pmatrix} 0.9349 & 0.0434 & 0.0217 \\ 0.7164 & 0.2239 & 0.0597 \\ 0.4400 & 0.4800 & 0.0800 \end{pmatrix}.$$

We estimated the volatility parameters from 20 days of TAQ data for Panera Bread. The liquidity parameters are inspired by the ones used in Almgren and Chriss (2001). We always use  $\sigma_{low}$  and  $\eta_{low}$  as starting points for the chains  $(\sigma_n)$  and  $(\eta_n)$ .

Parameter	Symbol	Value
Initial stock price	$S_0$	172 \$/share
Initial position	$X$	35,000 shares
Duration	$T$	100 minutes
Number of subintervals	$N$	100
Length of subintervals	$\Delta t$	1 minute
Permanent impact	$c$	$2.5 \times 10^{-7}$
Volatility states	$\sigma_{low}, \sigma_{med}, \sigma_{high}$	$3.51 \times 10^{-3}, 3.3 \times 10^{-2}, 1.172 \times 10^{-1}$
Liquidity states	$\eta_{low}, \eta_{med}, \eta_{high}$	$10^{-6}, 5 \times 10^{-6}, 25 \times 10^{-6}$

Table 2.1: Parameter values

## 2.2 Risk neutral objective

In the risk neutral case  $(\sigma_n, \eta_n, \varphi_n)$  is assumed to be a Markov chain with finite state space  $V^k := V \times \{1, \dots, k\} \subseteq \mathbb{R}_+^2 \times \mathbb{N}$  and transition probabilities

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n, \varphi_n) = w \mid (\sigma_{n-1}, \eta_{n-1}, \varphi_{n-1}) = v], \quad v, w \in V^k.$$

$\mathbb{E}_n^v$  denotes the conditional expectation  $\mathbb{E}[\cdot \mid (\sigma_n, \eta_n, \varphi_n) = v]$ . We want to find  $x \in \mathcal{A}$  that minimizes the conditional expectation  $\mathbb{E}_0^v[Q(x)]$  for a given initial state  $v$  of the Markov chain. Notice that  $Q(x)$  does not contain  $S_n$ ; therefore, it is enough for us to observe the process  $(\sigma_n, \eta_n, \varphi_n)$ ,  $n \geq 0$  to find an optimal strategy. The following theorem characterizes the optimal strategy and gives the optimal value for this problem.

**Theorem 2.2.1.** *One has*

$$\min_{x \in \mathcal{A}} \mathbb{E}_0^v[Q(x)] = X^2 a_0^v, \quad (2.2.1)$$

and the unique optimal execution strategy is given by

$$x_n^* |_{x_{n-1}^*, (\sigma_{n-1}, \eta_{n-1}, \varphi_{n-1})=v} = x_{n-1}^* \frac{\mathbb{E}_{n-1}^v[\tilde{\eta}_n]}{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \sum_{w \in V^k} p_{n-1}^{vw} a_n^w}, \quad n = 1, \dots, N-1, \quad (2.2.2)$$

where the coefficients  $a_n^v$  satisfy the backwards recursion:

$$a_{N-1}^v = \mathbb{E}_{N-1}^v[\tilde{\eta}_N], \quad a_{n-1}^v = \frac{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] \sum_{w \in V^k} p_{n-1}^{vw} a_n^w}{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \sum_{w \in V^k} p_{n-1}^{vw} a_n^w}, \quad n \leq N-1. \quad (2.2.3)$$

In the forward-backward system given by equations (2.2.2)–(2.2.3), the optimal strategy is defined as a function of the liquidity available in the market. One decreases the position in the asset in every step by a fraction that depends on the expected liquidity in the next period and a weighted sum of future liquidity terms. It is seen by inspecting (2.2.2) that if the expected cost of liquidity in the next period is large, hence the fraction in (2.2.2) is large, then there is an incentive to execute less orders and wait for a more favorable state.

**Remark 2.2.2.** When one assumes constant liquidity;  $\eta_n \equiv \eta$ , Theorem 2.2.1 yields

$$a_n^v = \frac{\eta - c/2}{N - n}, \quad x_n^* = x_{n-1}^* \frac{N - n}{N - n + 1},$$

which gives the constant speed strategy

$$x_n^* = X \frac{N - n}{N} \quad (2.2.4)$$

and this is consistent with the findings of Bertsimas and Lo (1998). In the case of constant liquidity one sees that there will be no incentive to sell faster or slower as there will be no favorable state to take advantage of; therefore, the order is sliced into equal amounts.

## Simulation

We simulated 50,000 paths of the Markov chain  $(\sigma_n, \eta_n, \xi_n)$  and we have calculated the optimal strategy on each of these paths. On the left side of Figure 2.1 we see the average optimal position  $\bar{x}_n$  in percent of the initial number of shares  $X$ , and on

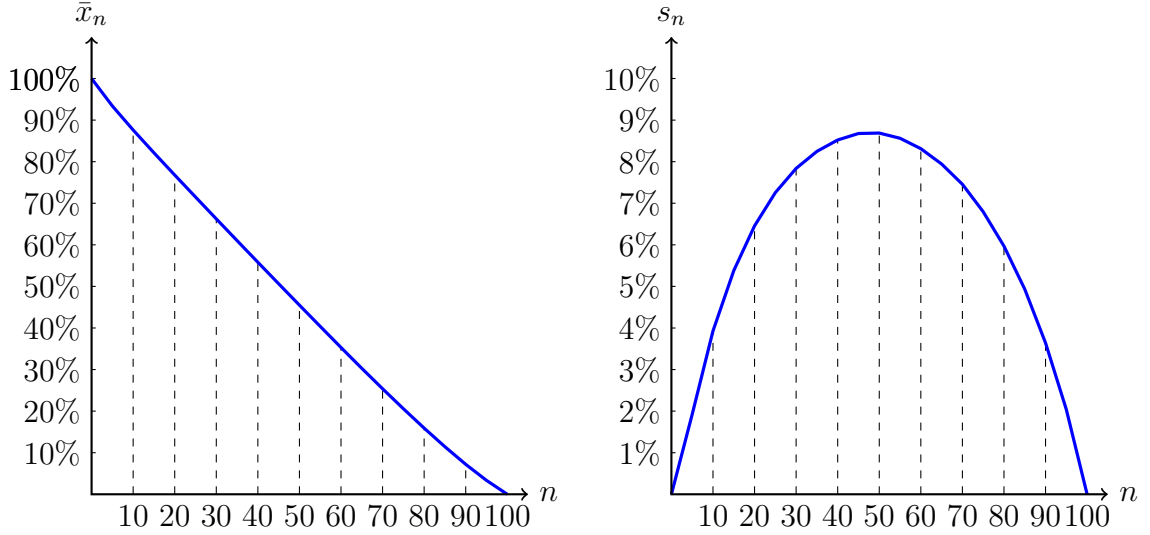


Figure 2.1: Average and standard deviation of the optimal positions for different scenarios in the risk neutral case.

the right side is the corresponding standard deviation  $s_n$  of the liquidation trajectory around this mean. Figure 2.2 shows three realizations of the optimal strategy on different paths and they are compared to the constant speed strategy (2.2.4). The histograms of Figure 2.3 show realizations of the implementation cost  $C(x)$ . It can be seen, that the optimal admissible strategy produces a lower sample mean and standard deviation than the constant speed strategy.

## 2.3 Expected exponential cost

This section assumes that  $(\sigma_n, \eta_n)$  is a Markov chain with finite state space  $V \subseteq \mathbb{R}_+^2$  and transition probabilities

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v], \quad v, w \in V,$$

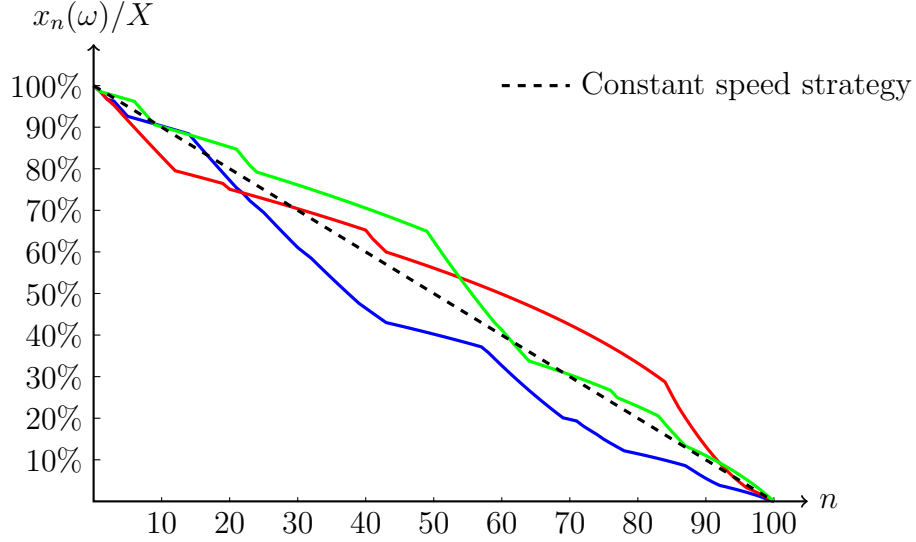


Figure 2.2: Optimal strategies for three particular scenarios in the risk neutral case compared to the constant speed strategy.

which does not depend on the sequence of innovations  $(\xi_n)$ . Our goal is to find  $x \in \mathcal{A}$  that minimizes the expectation of the exponential cost

$$\mathbb{E}_0^v[\exp(\alpha Q(x))] \quad (2.3.1)$$

for a parameter of absolute risk aversion  $\alpha > 0$ , where  $\mathbb{E}_n^v$  is the conditional expectation  $\mathbb{E}[\cdot \mid (\sigma_n, \eta_n) = v]$ . Let  $\mathcal{A}_n(z)$  be the set of  $(\mathcal{F}_n)$ -predictable strategies  $(x_i)_{i=n}^N$  satisfying  $x_n = z$ ,  $x_{i-1} \geq x_i$ ,  $x_N = 0$ . Notice that  $Q(x)$  is a sum of local terms and the exponential function factorizes; therefore, it is enough to observe  $(\sigma_n, \eta_n)$ ,  $n \geq 0$  to find the optimal strategy. Define

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \mathbb{E}_n^v[\exp(\alpha Q_n(x))],$$

The following theorem gives the optimal solution for 2.3.1.

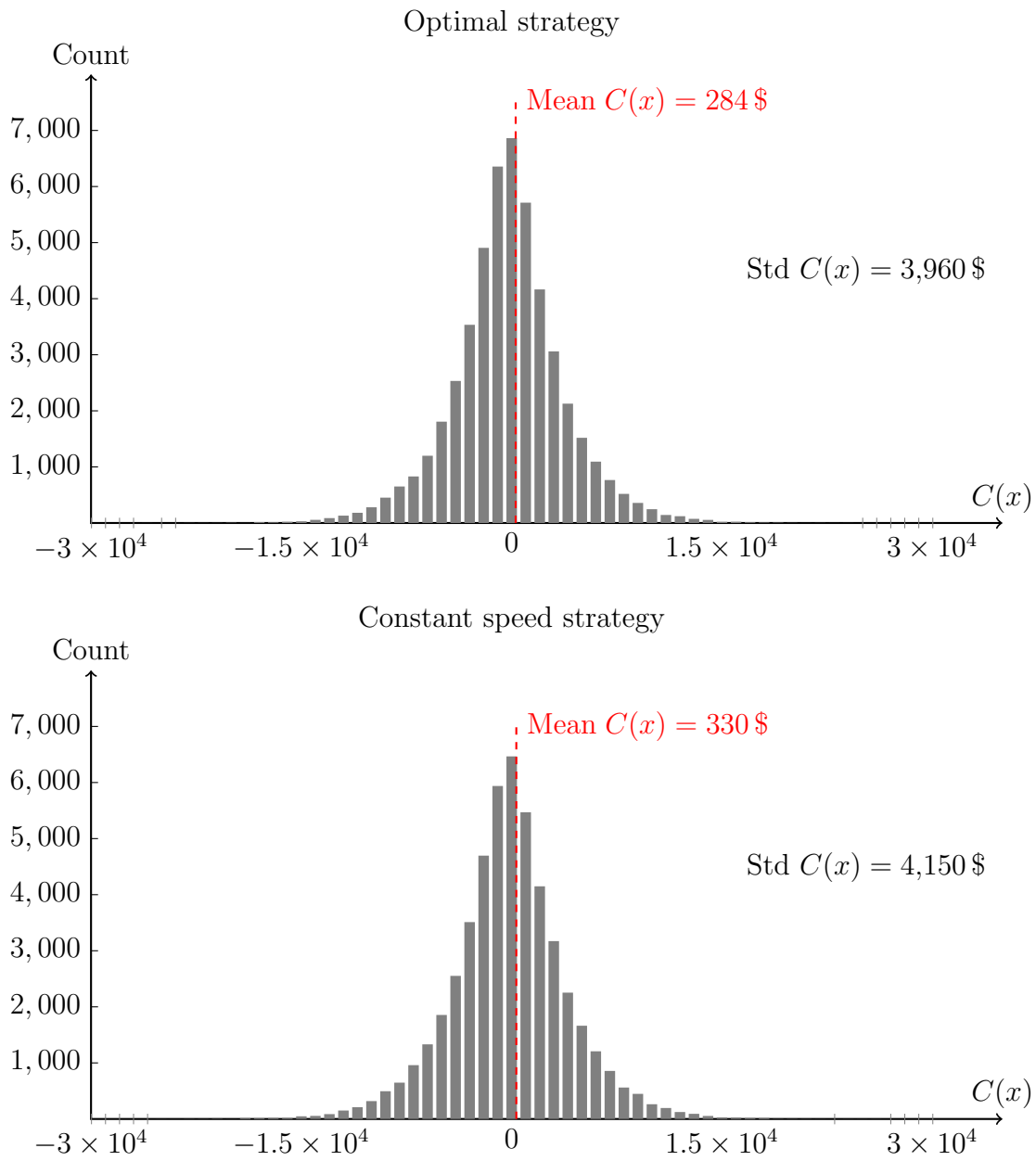


Figure 2.3: Histograms of  $C(x)$  for the optimal strategy and the constant speed strategy in the risk neutral case.



**Theorem 2.3.1.** *The value function  $J$  satisfies the Bellman equation*

$$J_{N-1}^v(x_{N-1}) = \sum_{w \in V} p_{N-1}^{vw} \exp(\alpha x_{N-1}^2 (w_2 - c/2))$$

$$J_{n-1}^v(x_{n-1}) = \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in V} p_{n-1}^{vw} \exp\left(\alpha(x_{n-1} - x_n)^2 (w_2 - c/2) + \frac{1}{2} \alpha^2 x_n^2 w_1^2 \Delta t\right) J_n^w(x_n),$$

for  $n \leq N - 1$ , and the minimizing  $x_n^*$  form the unique optimal strategy for problem (2.3.1).

**Remark 2.3.2.** In the special case of constant volatility  $\sigma_n \equiv \sigma$  and liquidity  $\eta_n \equiv \eta$ , the minimization of the expected exponential (2.3.1) reduces to the deterministic problem

$$\min_{x_{n-1} \geq x_n \geq 0} \exp \sum_{n=1}^N \alpha (x_n - x_{n-1})^2 (\eta - c/2) + \frac{1}{2} \alpha^2 \Delta t \sigma^2 x_n^2,$$

which is equivalent to the deterministic mean–variance problem (1.2.3) with  $\lambda = \alpha/2$ , hence the solution is given as in (1.2.4). This observation is in line with Schied et al. (2010).

It should be pointed out that in the case of having constant volatility and liquidity, (1.2.4) is the best  $(\mathcal{F}_n)$ –predictable strategy for the exponential problem; however, the optimal  $(\mathcal{F}_n)$ –predictable mean–variance strategy is not deterministic; see Almgren and Lorenz (2007), Lorenz and Almgren (2011) and Section 2.4.

For stochastic volatility and liquidity, the Bellman equation of Theorem 2.3.1 can be solved numerically on a discrete grid of controls. The computational cost is of order  $\mathcal{O}(X^2)$ , when one allows trading in quantities of single shares. Because in every step,  $J_{n-1}^v(x_{n-1})$  has to be computed for all  $0 \leq x_{n-1} \leq X$ , and for fixed  $x_{n-1}$ , the calculation of  $J_{n-1}^v(x_{n-1})$  requires an evaluation for every  $x_n = 0, \dots, x_{n-1}$ . For large  $X$ , the computational complexity can be reduced by restricting the sale of the asset to larger lots of shares.

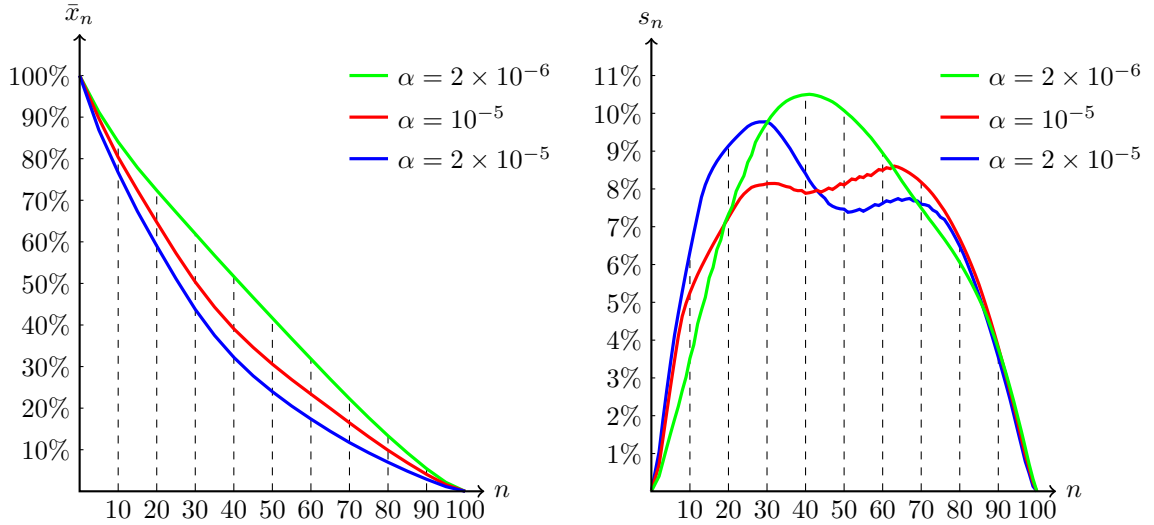


Figure 2.4: Average and standard deviation of optimal positions for the expected exponential cost criterion with different absolute risk aversions  $\alpha$ .

### Simulation

In this section's simulations we allowed assets to be sold in lots of 350 shares which amounts to 1% of the initial position. The optimal strategies have been computed along 50,000 simulated paths of  $(\sigma_n, \eta_n, \xi_n)$ . Figure 2.4 shows the average  $\bar{x}_n$  of the optimal positions and the standard deviations  $s_n$  for three different values of absolute risk aversion  $\alpha$ . It can be seen that the average liquidation speed is increasing in  $\alpha$ , an observation that makes sense intuitively because the expected exponential of the cost also penalizes the exposure of the remaining asset positions to volatility risk. In Figure 2.5 we see that three different realizations of the optimal admissible strategy are compared to the Almgren–Chriss strategy (1.2.4) corresponding to  $\lambda = \alpha/2$  and the mean volatility  $\bar{\sigma}$  and liquidity  $\bar{\eta}$  with respect to long run averages of  $(\sigma_n)$  and  $(\eta_n)$ . Figure 3.5 shows histograms of realized implementation costs  $C(x)$  produced by the optimal strategy of Theorem 2.3.1 and the Almgren–Chriss strategy (1.2.4). It can be seen that the sample mean and variance of the optimal admissible strategy are significantly lower.

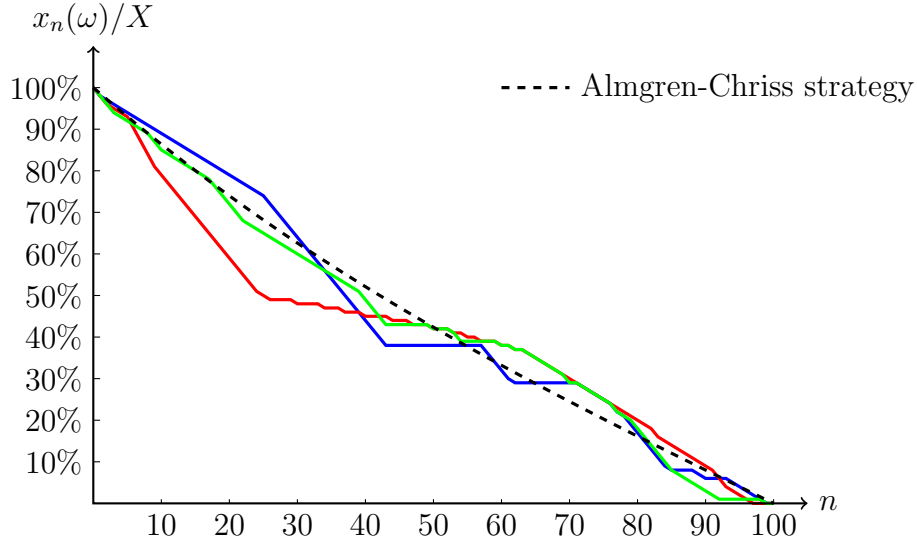


Figure 2.5: Three realizations of the optimal strategy for the expected exponential criterion with parameter  $\alpha = 2 \times 10^{-5}$  compared to the Almgren–Chriss strategy (1.2.4) corresponding to  $\bar{\sigma}, \bar{\eta}$  and parameter  $\lambda = 10^{-5}$ .

## 2.4 Mean-variance criterion

This section considers the mean-variance minimization problem of the form

$$\mathbb{E}[Q(x)] + \lambda \text{Var}(Q(x)) \tag{2.4.1}$$

for a given trade-off parameter  $\lambda > 0$ . We assume that as in Section 2.3,  $(\sigma_n, \eta_n)$  is a Markov chain with finite state space  $V \subseteq \mathbb{R}_+^2$  and transition probabilities

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v], \quad v, w \in V,$$

which is independent of innovations  $(\xi_n)$ . Notice that problem (2.4.1) is more difficult to solve for a predictable strategy than the exponential problem (2.3.1). First of all, the criterion (2.4.1) is not amenable for direct use of dynamic programming methods. Additionally, since it does not factorize, it is no longer enough for the optimal admissible strategy to only observe  $(\sigma_n, \eta_n)$ . We solve the mean-variance

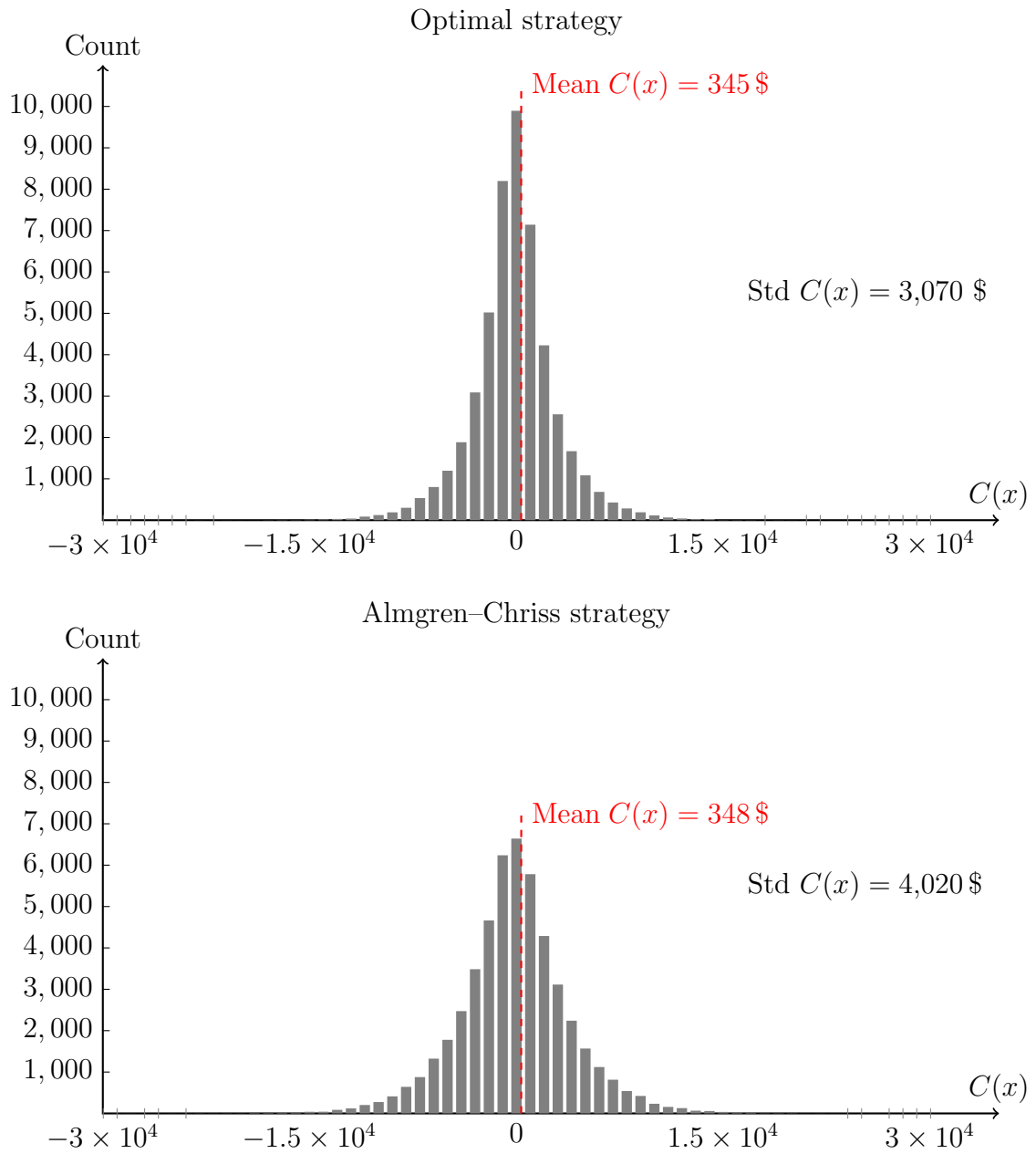


Figure 2.6: Histogram of  $C(x)$  for the optimal strategy and the Almgren-Chriss strategy (1.2.4) corresponding to  $\bar{\sigma}$  and  $\bar{\eta}$  in the expected exponential cost case with absolute risk aversion  $\alpha = 2 \times 10^{-5}$ .

problem for two different cases. In Subsection 2.4.1 we restrict the class of admissible trading strategies to strategies that only observe  $(\sigma_n, \eta_n)$  and find the optimal one among those. In Subsection 2.4.2 we compute the fully optimal admissible strategy. Our simulation results suggest that the restricted solution is very close to being fully optimal. Moreover, it showed a performance that is virtually indistinguishable from the optimal strategy for the exponential criterion (2.3.1) with  $\alpha = 2\lambda$ .

### 2.4.1 The restricted mean-variance problem

Let us define the filtration  $\mathcal{G}_n := \sigma(\sigma_i, \eta_i : 0 \leq i \leq n)$ , and denote by  $\mathcal{R}$  the subset of admissible strategies  $x \in \mathcal{A}$  that are predictable with respect to  $(\mathcal{G}_n)$ . For a given  $\lambda > 0$  the restricted mean–variance problem  $P(\lambda)$  reads as

$$P(\lambda) = \min_{x \in \mathcal{R}} \mathbb{E}[Q(x) \mid \mathcal{G}_0] + \lambda \text{Var}(Q(x) \mid \mathcal{G}_0).$$

Since conditional variances are not recursively determined, problem  $P(\lambda)$  cannot be solved recursively by using dynamic programming methods. But we show in Proposition A.1.1 that a solution  $x^*$  of  $P(\lambda)$  also solves the problem  $P(\lambda, \mu)$ ,

$$P(\lambda, \mu) = \min_{x \in \mathcal{A}} \mathbb{E} [\mu Q(x) + \lambda Q(x)^2 \mid \mathcal{G}_0]$$

for  $\mu = 1 - 2\lambda \mathbb{E}[Q(x^*)]$ . Therefore, it is possible to relate  $P(\lambda)$  and  $P(\lambda, \mu)$ . Furthermore, we can derive a Bellman equation for problem  $P(\lambda, \mu)$ . However, the subtlety is that we do not know  $\mu$  before knowing  $x^*$ . Therefore, we compute solutions to  $P(\lambda, \mu)$  for different values of  $\mu$  and check which one minimizes  $P(\lambda)$ . One other subtlety is whether the solution to  $P(\lambda)$  exists. Notice that  $(\mathcal{G}_n)$  is discrete, and  $\mathcal{R}$  is a compact subset of a finite–dimensional space. Therefore, since the objective function of  $P(\lambda)$  is continuous in  $x$  by standard arguments it follows that  $P(\lambda)$  admits an optimal solution  $x^* \in \mathcal{R}$ . To solve  $P(\lambda, \mu)$ , we only need to know  $(\sigma_0, \eta_0)$ . Let us denote by

$\mathcal{R}_n(z)$  the set of  $\mathcal{R}$ -predictable strategies  $(x_i)_{i=n}^N$  such that  $x_n = z$ ,  $x_{i-1} \geq x_i$ ,  $x_N = 0$  and define an optimal value function

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \mathbb{E}_n^v [\mu Q_n(x) + \lambda Q_n(x)^2],$$

where  $Q_n(x) := \sum_{i=n+1}^N (x_{i-1} - x_i)^2 \tilde{\eta}_i - x_i \sigma_i \sqrt{\Delta t} \xi_i$ . Then the following theorem gives the optimal solution.

**Theorem 2.4.1.** *The value function  $J$  satisfies the Bellman equation*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \mu x_{N-1}^2 \mathbb{E}_{N-1}^v [\tilde{\eta}_N] + \lambda x_{N-1}^4 \mathbb{E}_{N-1}^v [\tilde{\eta}_N^2], \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \mu (x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v [\tilde{\eta}_n] + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v [\sigma_n^2] \\ &\quad + \lambda (x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v [\tilde{\eta}_n^2] + 2\lambda (x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v \left[ \tilde{\eta}_n \sum_{i=n+1}^N (x_{i-1} - x_i)^2 \tilde{\eta}_i \right] \\ &\quad + \sum_{w \in V} p_{n-1}^{vw} J_n^w(x_n), \quad n \leq N-1, \end{aligned} \tag{2.4.2}$$

and any strategy  $x^* \in \mathcal{R}$  minimizing (2.4.2) for all  $n = 1, \dots, N-1$ , is an optimal solution to problem  $P(\lambda, \mu)$ . Moreover, if  $\mu \geq 0$ , the optimal strategy is unique.

**Remark 2.4.2.** For constant volatility  $\sigma_n \equiv \sigma$  and liquidity  $\eta_n \equiv \eta$ , problem  $P(\lambda)$  becomes the Almgren–Chriss problem (1.2.3) with  $\lambda = \alpha/2$ . (1.2.4) is the optimal deterministic solution, but it is sub-optimal among all admissible strategies  $\mathcal{A}$ ; see Almgren and Chriss (2001), Almgren and Lorenz (2007), Lorenz and Almgren (2011) and Subsection 2.4.2.

In the case of having stochastic volatility and liquidity, the Bellman equation of Theorem 2.4.1 can be solved numerically by discretizing the space of controls. If one assumes that the asset can be sold in single units, the complexity of the numerical procedure is  $\mathcal{O}(X^2)$  as in the expected exponential of the cost case. If the initial

position  $X$  is large, then one may reduce the complexity by restricting the sale of the asset to larger lots.

### Simulation

When computing the optimal strategies along 50,000 simulated realizations of  $(\sigma_n, \eta_n, \xi_n)$  by using a numerical procedure, we allowed trading in lots of 350 shares, which amounts to 1% of the initial position of 35,000 shares. The average optimal positions  $\bar{x}_n$  and the standard deviations  $s_n$  for three different values of the mean–variance trade–off parameter  $\lambda$  can be seen in Figure 2.7. One notices that the speed of liquidation is higher for larger values of  $\lambda$ . This is intuitive since larger values of  $\lambda$  penalize the exposure of the remaining asset position to volatility risk more and hence the position is liquidated faster to reduce this exposure. Figure 2.8 shows three realizations of the optimal admissible strategy compared to the Almgren–Chriss’ mean–variance optimal deterministic strategy (1.2.4) corresponding to the mean values  $\bar{\sigma}$  and  $\bar{\eta}$  under the steady state distributions of the Markov chains  $(\sigma_n)$  and  $(\eta_n)$ . Figure 2.9 shows the mean–variance criterion for optimal solutions  $x^\mu$  to problem  $P(\lambda, \mu)$  for  $\lambda = 10^{-5}$  and different values of  $\mu$ . The minimum is attained for  $\mu^* = 0.968$ , which is almost equal to  $1 - 2\lambda\mathbb{E}[Q(x^{\mu^*})] = 0.997$ . Notice that this is consistent with the theoretical result of Proposition A.1.1. Figure 2.10 shows histograms of the realized implementation cost  $C(x)$  for the optimal admissible strategy and the Almgren–Chriss strategy (1.2.4) corresponding to  $\bar{\sigma}$  and  $\bar{\eta}$ . It can be seen that the simulation sample mean and variance of the optimal admissible strategy are significantly lower. We note that in Figure 3.5 the optimal admissible strategy produces a histogram of outcomes of  $C(x)$  that is almost indistinguishable from the one produced by the optimal mean-variance strategy with  $\lambda = \alpha/2$ . This can be explained by the relation that the second order Taylor expansion of the logarithm of the expected exponential cost formulation yields a mean–variance type objective with  $\lambda = \alpha/2$ .

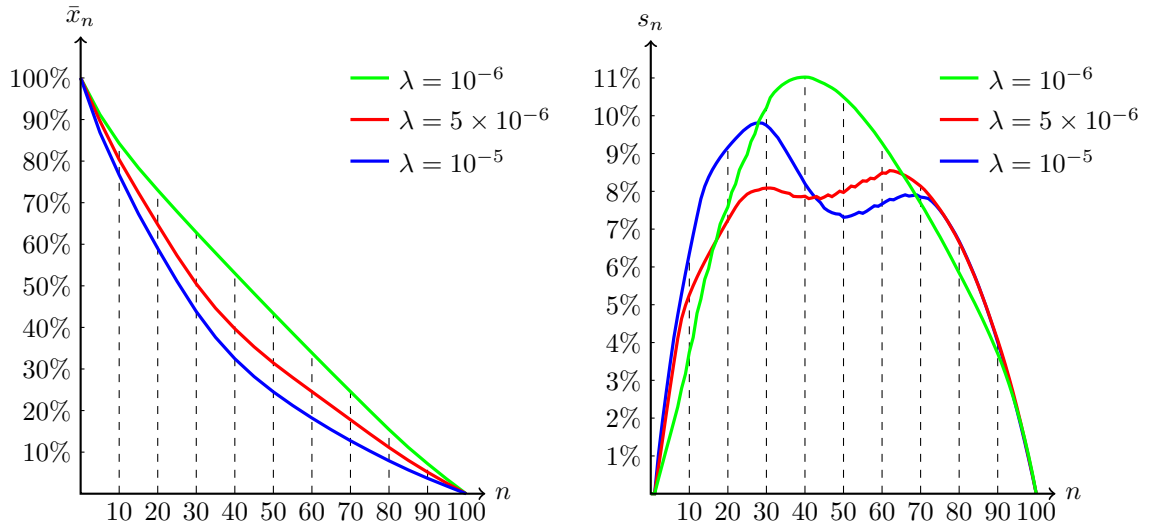


Figure 2.7: Average and standard deviation of  $\mathcal{R}$ -optimal positions for the mean-variance criterion with different values of  $\lambda$ .

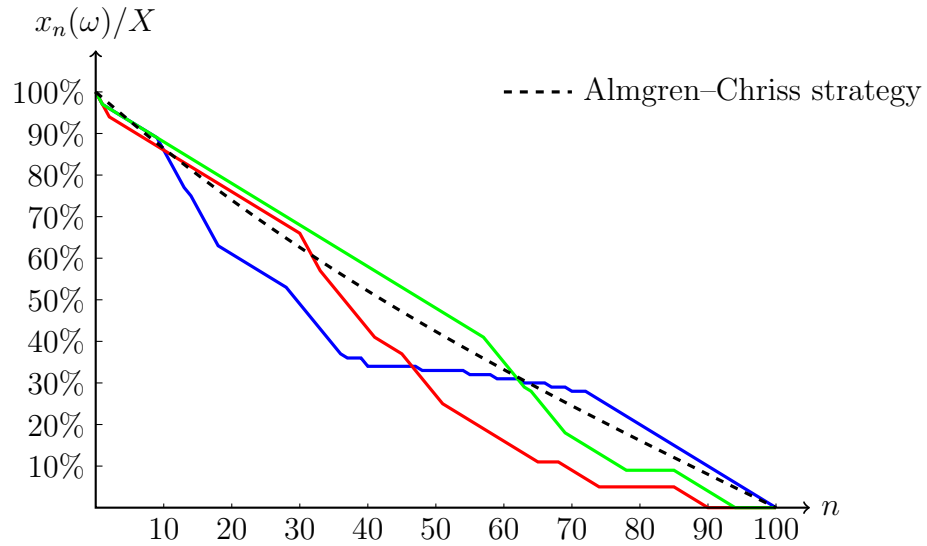


Figure 2.8: Three realizations of the optimal strategy for the mean-variance criterion with parameter  $\lambda = 10^{-5}$  compared to the Almgren-Chriss strategy (1.2.4) corresponding to  $\bar{\sigma}$  and  $\bar{\eta}$ .



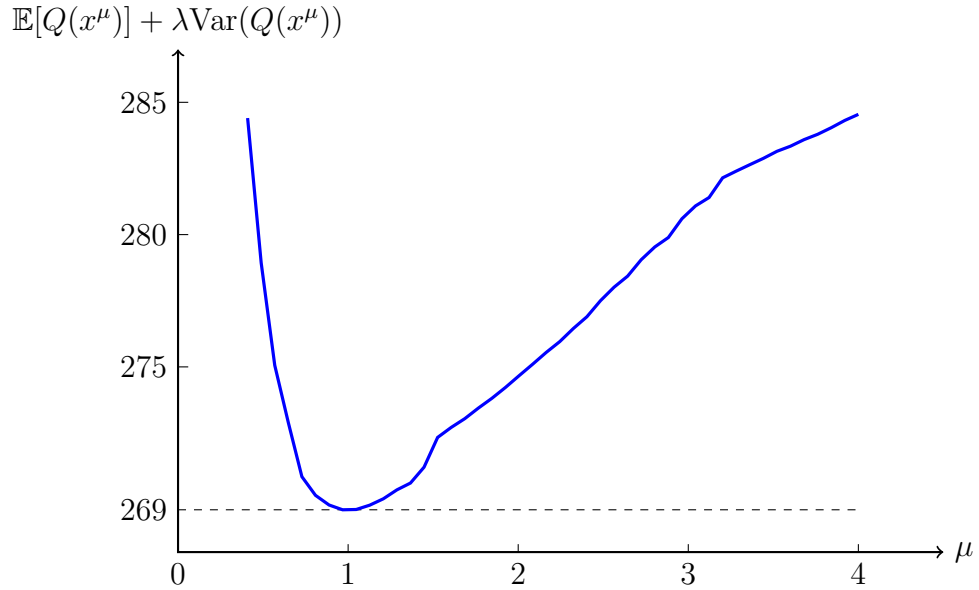


Figure 2.9: Mean–variance criterion for optimal solutions to  $P(\lambda, \mu)$  for  $\lambda = 10^{-5}$  and different values of  $\mu$ . The minimum is attained for  $\mu^* = 0.968$ , and  $1 - 2\lambda\mathbb{E}[Q(x^{\mu^*})] = 0.997$ .

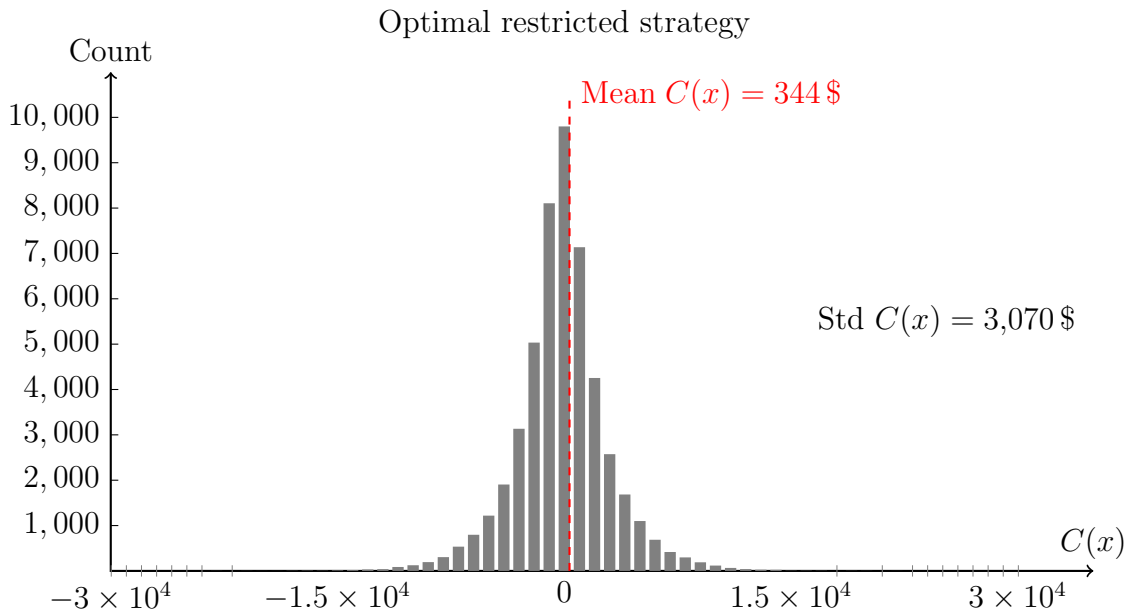


Figure 2.10: Histogram of  $C(x)$  for the optimal restricted strategy in the mean–variance case for  $\lambda = 10^{-5}$ .

## 2.4.2 The full mean-variance problem

In this subsection we are interested in the problem  $\hat{P}(\lambda)$ ,

$$\hat{P}(\lambda) = \min_{x \in \mathcal{A}} \mathbb{E}[Q(x) \mid \mathcal{F}_0] + \lambda \text{Var}(Q(x) \mid \mathcal{F}_0)$$

of finding a mean–variance optimal strategy in the set of all admissible strategies  $\mathcal{A}$ . Notice that  $\mathcal{A}$  can be viewed as a closed, bounded, and convex subset of a Hilbert space. Hence, it is weakly compact, and it follows that  $\hat{P}(\lambda)$  has an optimal solution  $x^* \in \mathcal{A}$ . As in Subsection 2.4.1, we derive a Bellman equation for the auxiliary problem  $\hat{P}(\lambda, \mu)$  which is given by,

$$\hat{P}(\lambda, \mu) = \min_{x \in \mathcal{A}} \mathbb{E} [\mu Q(x) + \lambda Q(x)^2 \mid \mathcal{F}_0].$$

We solve the auxiliary problem for different  $\mu$  and see which one gives the minimum value. Let us introduce the running cost

$$h_n(x) := \sum_{i=1}^n (x_{i-1} - x_i)^2 \tilde{\eta}_i - x_i \sigma_i \xi_i,$$

and the value function

$$J_n^v(h, z) := \min_{x \in \mathcal{A}_n(z)} \mathbb{E}_n^v [(\mu + 2\lambda h)Q_n(x) + \lambda Q_n(x)^2],$$

where  $\mathbb{E}_n^v$ ,  $\mathcal{A}_n(z)$  and  $Q_n(x)$  are defined as before in Section 2.3.  $J_0^v(0, X)$  gives the optimal value of the problem  $\hat{P}(\mu, \lambda)$  given that the initial state we are in is  $(\sigma_0, \eta_0) = v$ , and the following theorem yields a Bellman equation for  $J$ .

**Theorem 2.4.3.** *The value function  $J$  satisfies the Bellman equation*

$$J_{N-1}^v(h, x_{N-1}) = (\mu + 2\lambda h)x_{N-1}^2 \mathbb{E}_{N-1}^v[\tilde{\eta}_N] + \lambda x_{N-1}^4 \mathbb{E}_{N-1}^v[\tilde{\eta}_N^2],$$

$$\begin{aligned} J_{n-1}^v(h, x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} (\mu + 2\lambda h)(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] \\ &\quad + \lambda(x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v[\tilde{\eta}_n^2] + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v[\sigma_n^2] \\ &\quad + \sum_{w \in V} p_{n-1}^{vw} \int_{\mathbb{R}} J_n^w \left( h + (x_{n-1} - x_n)^2 (w_2 - c/2) - x_n w_1 \sqrt{\Delta t} \xi, x_n \right) \rho(\xi) d\xi, \end{aligned}$$

where  $\rho$  is the density of a standard normal random variable. A strategy  $x^* \in \mathcal{A}$  solves problem  $\hat{P}(\mu, \lambda)$  if in every step,  $x_n^*$  gives the minimum for  $h = h_{n-1}(x^*)$  and  $x_{n-1} = x_{n-1}^*$ . Moreover, if  $\mu \geq 0$ , the optimal strategy  $x^* \in \mathcal{A}$  is unique.

**Remark 2.4.4.**  $\hat{P}(\lambda)$  extends the problem considered in Lorenz and Almgren (2011) by allowing volatility and liquidity to be stochastic. To find the optimal strategy, one needs to find solutions  $x^\mu$  to the auxiliary problem  $\hat{P}(\mu, \lambda)$  for different values of  $\mu$  and check which one minimizes  $\hat{P}(\lambda)$ . However, notice that the additional variable  $h$  makes the numerical solution of  $J$  much more complicated than in Section 4.4 and Subsection 2.4.1. Therefore, we did our numerical experiments in an Almgren–Chriss model with constant volatility and liquidity by using only five time steps. The numerical results we obtained indicate that the optimal solution performs only insignificantly better than the Almgren–Chriss strategy (1.2.4).

### Simulation

The numerical results of this subsection have been obtained by restricting the simulation to a model with constant volatility and liquidity that only has five time steps. We simulated 50,000 realizations of  $(\xi_n)_{n=1}^5$  and used the long term averages of volatility and liquidity, denoted by  $\bar{\sigma}$  and  $\bar{\eta}$  as in Section 4.4. We discretized the state space of running cost using 250 evenly spaced grid points in the interval  $[150 - 2 \times 1500, 150 + 2 \times 1500]$ . The values 150 and 1500 were inspired by the mean

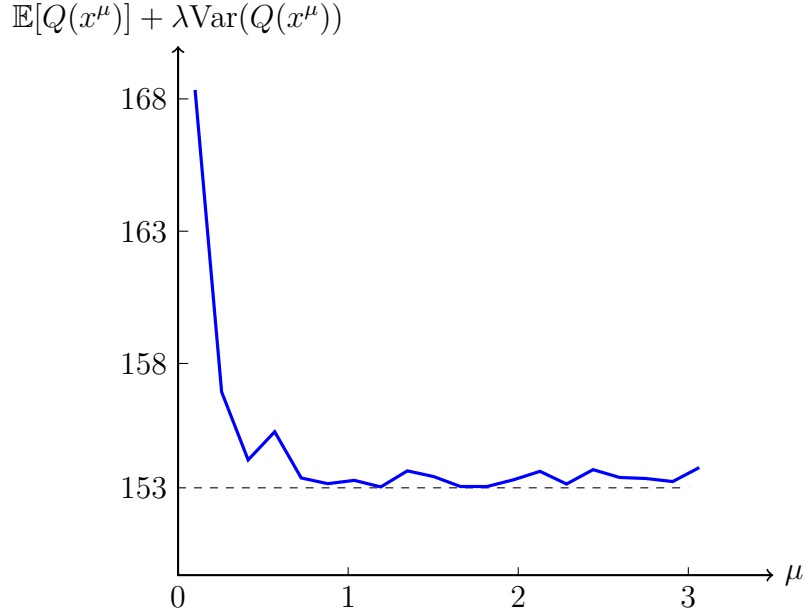


Figure 2.11: Mean-variance criterion for optimal solutions to  $\hat{P}(\lambda, \mu)$  for  $\lambda = 10^{-5}$  and different values of  $\mu$ . The minimum is attained for  $\mu^* = 1.192$ , and  $1 - 2\lambda\mathbb{E}[Q(x^{\mu^*})] = 0.998$ .

and standard deviation of  $Q(x)$  generated by the Almgren–Chriss strategy. We allow trading in multiples of 350 shares, which corresponds to 1% of the initial stock position. Monte Carlo method has been used to compute the integral in Theorem 2.4.3. Figure 2.11 shows the  $\mu$ -value producing the lowest mean-variance criterion. Figure 2.12 contains the histograms of the optimal and the Almgren–Chriss strategy for a mean-variance trade-off parameter  $\lambda = 10^{-5}$ . The optimal strategy is observed to improve the mean-variance criterion slightly over the Almgren–Chriss strategy. Notice also that while it produced a better sample mean, its standard deviation is worse.

## 2.5 Conclusion

This chapter studied an optimal execution problem in a Markovian setting where the volatility and temporary price impact are allowed to be stochastic. We used risk neutral, expected exponential and mean-variance of the cost objectives to compute

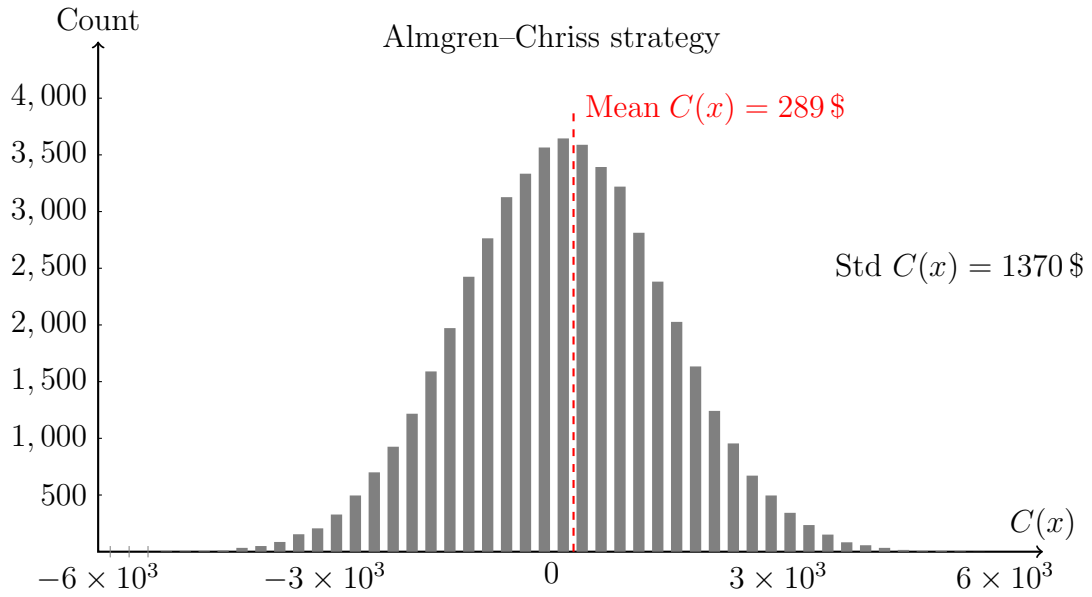
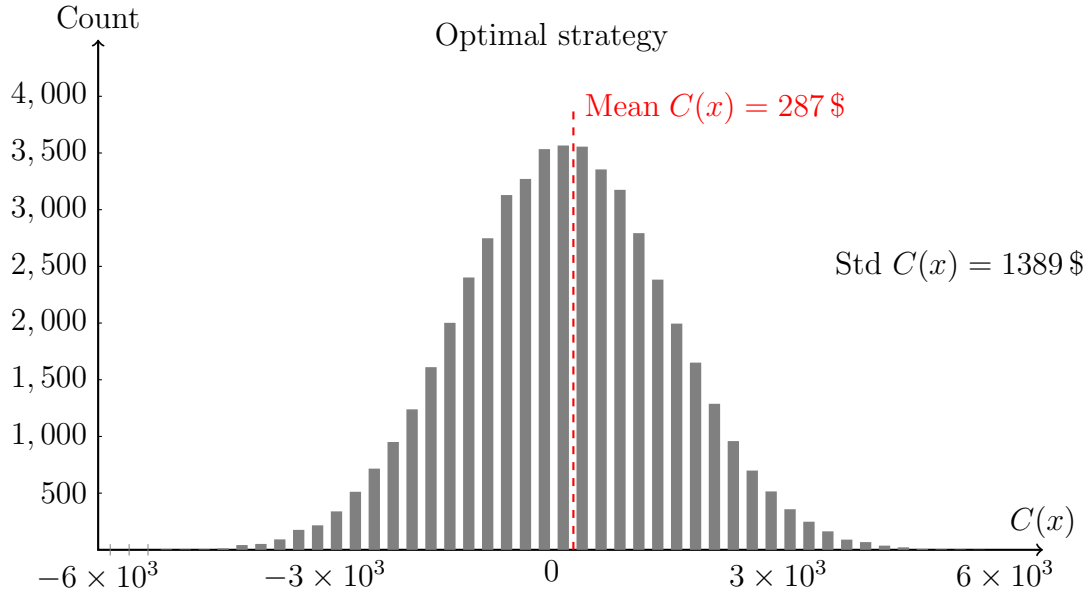


Figure 2.12: Histograms of  $C(x)$  for the optimal strategy and the Almgren-Chriss strategy (1.2.4) in the mean-variance case with constant volatility  $\bar{\sigma}$  and liquidity  $\bar{\eta}$  for  $\lambda = 10^{-5}$ .

optimal execution strategies. Notice that although we made independence assumptions for the liquidity and volatility parameter set we have specified, our formulation allows for dependence between volatility, liquidity and price innovations. The transition probabilities can also be time-dependent. This feature makes the model flexible to incorporate specific situations into the price dynamics. It is well known that price dynamics depend on the time of the day. For example, some small or medium cap stocks are traded less liquidly at certain times of the day, and volatility typically increases around scheduled news announcements (i.e. earnings and FED announcements). To apply the model in practice, one needs to estimate the parameters using real data.

# Chapter 3

## Single–asset execution with a dark pool

As it was the case in Chapter 2, the majority of the models are used to determine an optimal schedule of orders to be submitted to an exchange where the order book is observable and do not consider dark pools. Dark pools are exchanges where the liquidity dynamics are not directly observable. Although, the order crossing dynamics may vary between dark pools, it is usually the case that the orders are crossed at the mid price observed in the regular exchange. Therefore, it is assumed that the dynamics of the regular exchange create price impact on the dark pool but not vice versa. Nevertheless, the convenience of cost saving comes at the expense of execution uncertainty, since the orders submitted to the dark pool may not be filled if there is no liquidity available. One may see Degryse et al. (2009) for an overview of literature on dark pools and their operating mechanisms.

We consider an aggregated dark pool as opposed to considering different exchanges and deciding how to splits orders between them. For example, Ganchev et al. (2010) study a dark pool problem where one submits orders into several dark pools in a discrete time setting. After each order submission it is learned how many shares have

executed at each venue. By using this information they estimate the tail probability of how much extra liquidity there could be in each venue. Another example is Laurelle et al. (2011) where they study the same problem but also incorporate a stochastic crossing price which is modeled as a random variable that depends on the price process in the regular exchange. We are not concerned with discovering the liquidity across different venues and treat one dark pool as the main provider of hidden liquidity.

Examples of studies looking for an optimal execution program in the presence of a dark pool include the models of Kratz and Schöneborn (2013a), Kratz and Schöneborn (2013b), Kratz (2014). Kratz and Schöneborn (2013a) study a discrete time portfolio execution problem in the presence of a dark pool. They use an objective function that has a quadratic penalty term for the volatility of remaining stock positions and they show the existence of an optimal strategy when one only assumes temporary costs of transaction. They produce analytical solutions for Almgren–Chriss type transient impact function in a setting where all or none of the orders submitted to dark pool execute. Additionally they study a case for adverse selection. They assume that depending on the execution in dark pool the expected market innovations will result in an adverse price move in the regular exchange. This price move is assumed to be equal to the permanent costs of transaction. Kratz and Schöneborn (2013b) study the same problem in a continuous time analogue of Kratz and Schöneborn (2013a) without adverse selection effects. Furthermore, they use a multi-dimensional Poisson process to determine the time of execution for the orders submitted to the dark pool. It is assumed that the orders in the dark pool will be executed full at the jump times of the corresponding Poisson process. The jump probabilities are assumed to be independent of the order sizes. Kratz (2014), on the other hand, treats a single asset continuous time by accounting for adverse selection effects in a way that is along the lines of Kratz and Schöneborn (2013a). However, they use the Poisson dynamics proposed by Kratz and Schöneborn (2013b) for the dark pool execution times.



We formulate our problem using the model in Chapter 2 by including a dark pool where the orders submitted will be subject to adverse selection. We assume the orders submitted to the dark pool are crossed at the mid price hence do not incur temporary transaction costs. However, since the orders that transact in the regular exchange cause a permanent change in the stock price they affect the dark pool crossing price. Moreover, we make the trading program subject to adverse selection by letting the dark pool execution probabilities depend on the direction of market innovations and the orders submitted to the dark pool. Our model differs from the models of Kratz and Schöneborn (2013a), Kratz and Schöneborn (2013b) and Kratz (2014) in a few aspects. We first allow the permanent impact to affect the dynamics of the dark pool. Second, we allow the stock price volatility and liquidity to be codependent and stochastic. We also differ in the way we model the adverse selection. Instead of effecting the fundamental price process depending on dark pool executions, we choose to let the dark pool execution probabilities depend on the direction of submitted order and the market innovations. In our framework the dark pool dynamics depend on the primary exchange but not vice versa. We find the optimal execution strategy when the objective is to minimize the expectation and the expected exponential of the execution cost. When the objective is to minimize the expected cost the strategy becomes insensitive to volatility of the price process. The optimal strategy in this case can be characterized by a set of stochastic equations. In the expected exponential of the cost case we derive a backward recursion that describes the optimal strategy and solve it numerically. In this case we are able to account for volatility of remaining stock positions as the market innovations effect does not disappear from the objective function.

The structure of this chapter is as follows. In the next section we extend the model of Chapter 2 to include a dark pool. In Section 3.2 we study the risk neutral

case. Section 3.3 treats the expected exponential case and Section 3.4 concludes. We give all the proofs in Appendix A.2.

### 3.1 The model

The mid price process is assumed to behave according to (2.1.2) and the execution price that an order  $y_n$  achieves according to (2.1.1). The orders submitted to the dark pool are denoted by  $z_n$  and are assumed to be executed at price  $S_{n-1}$ . But they are not always crossed. We suppose they are of fill or kill type, that is, an order to buy  $z_n$  shares at time  $t_{n-1}$  is either fully executed at price  $S_{n-1}$ , or it is immediately canceled. The number of shares executed is  $b_n z_n$ , where  $b_n$  is a random variable taking the value 0 or 1. Trading in the dark pool is exposed to adverse selection risk, which results from the fact that orders are more likely to be crossed if the next price movement is in a favorable direction. This is modeled by making the probability of  $\{b_n = 1\}$  depend on the sign of  $z_n \xi_n$ .

We call the process  $(y_n, z_n)$  an execution strategy and denote by  $x_n = X - \sum_{i=1}^n (y_i + b_i z_i)$  the remaining orders to be executed. We suppose that  $S_n, \sigma_n, \eta_n, b_n$  are observable and denote the filtration they generate by  $(\mathcal{F}_n)$ . An execution strategy  $(y, z)$  is called admissible if it is predictable with respect to  $(\mathcal{F}_n)$  and satisfies  $x_N = z_N = 0$ .  $x_N = 0$  where the condition  $z_N = 0$  ensures that that the program achieves execution completely by forcing  $y_N = x_{N-1}$ . We denote the set of all admissible strategies by  $\mathcal{A}$ . The implementation cost  $C(y, z)$  of a strategy  $(y, z) \in \mathcal{A}$  reads as

$$C(y, z) = X S_0 - \sum_{n=1}^N \left( y_n \tilde{S}_n + b_n z_n S_{n-1} \right). \quad (3.1.1)$$

which can be expressed as

$$C(y, z) = \sum_{n=1}^N c x_n y_n + y_n^2 \eta_n - x_n \sigma_n \sqrt{\Delta t} \xi_n.$$

$\xi_n$  is independent of  $\sigma(\mathcal{F}_{n-1}, \sigma_n, \eta_n)$  and  $(\sigma_n, \eta_n)$  is a Markov chain with finite state space  $V \subseteq \mathbb{R}_+^2$  and time-dependent transition probabilities

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v], \quad v, w \in V.$$

We denote the conditional expectation  $\mathbb{E}[\cdot \mid (\sigma_n, \eta_n) = v]$  by  $\mathbb{E}_n^v$ . Moreover, we assume there are two numbers  $0 \leq q_1 < q_2 \leq 1$  such that

$$\mathbb{P}[b_n = 1 \mid \text{sign}(z_n \xi_n) = 1] = q_1 \quad \text{and} \quad \mathbb{P}[b_n = 1 \mid \text{sign}(z_n \xi_n) = -1] = q_2.$$

Furthermore it is assumed that  $\mathbb{E}_{n-1}^v \tilde{\eta}_n > 0$  for all  $n$ , where  $\tilde{\eta}_n := \eta_n - c$ .

In the numerical examples we present, we use the same parameter set of Chapter 2 and furthermore assume that the dark pool execution probabilities are assumed to be  $q_1 = 1/1000$ , and  $q_2 = 5/1000$ . The initial states  $\sigma_0$ , and  $\eta_0$  are assumed to be  $\sigma_{low}$  and  $\eta_{low}$ .

## 3.2 Risk neutral objective

We first find a strategy  $(y, z) \in \mathcal{A}$  that minimizes the expected implementation cost. The following theorem gives this strategy.

**Theorem 3.2.1.** *For  $c \geq 0$  sufficiently small, one has*

$$\min_{(y,z) \in \mathcal{A}} \mathbb{E}_0^v C(y, z) = a_0^v X^2,$$

*and the unique optimal strategy is given by*

$$y_n^* |_{x_{n-1}^*, (\sigma_{n-1}, \eta_{n-1})=v} = d_{n-1}^v x_{n-1}^*, \quad z_n^* |_{x_{n-1}^*, (\sigma_{n-1}, \eta_{n-1})=v} = e_{n-1}^v x_{n-1}^*, \quad n = 1, \dots, N-1,$$

where the coefficients  $a_n^v$  satisfy the backwards recursion

$$\begin{aligned} a_{N-1}^v &= \mathbb{E}_{N-1}^v \eta_N \\ a_{n-1}^v &= cd_{n-1}^v(1 - \bar{q}e_{n-1}^v) + (d_{n-1}^v)^2 \mathbb{E}_{n-1}^v(\eta_n - c) \\ &\quad + (1 - 2d_{n-1}^v + (d_{n-1}^v)^2 + \bar{q}e_{n-1}^v [e_{n-1}^v - 2 + 2d_{n-1}^v]) \sum_w p_{n-1}^{vw} a_n^w, \end{aligned}$$

for  $n \leq N - 1$  and

$$\begin{aligned} d_{n-1}^v &:= \frac{(1 - \bar{q}) (\sum_w p_{n-1}^{vw} a_n^w) (\sum_w p_{n-1}^{vw} a_n^w - c/2)}{(\sum_w p_{n-1}^{vw} a_n^w) (\mathbb{E}_{n-1}^v \tilde{\eta}_n + \sum_w p_{n-1}^{vw} a_n^w) - \bar{q} (\sum_w p_{n-1}^{vw} a_n^w - c/2)^2} \\ e_{n-1}^v &:= \frac{(\sum_w p_{n-1}^{vw} a_n^w) (\mathbb{E}_{n-1}^v \tilde{\eta}_n + \sum_w p_{n-1}^{vw} a_n^w) - (\sum_w p_{n-1}^{vw} a_n^w - c/2)^2}{(\sum_w p_{n-1}^{vw} a_n^w) (\mathbb{E}_{n-1}^v \tilde{\eta}_n + \sum_w p_{n-1}^{vw} a_n^w) - \bar{q} (\sum_w p_{n-1}^{vw} a_n^w - c/2)^2}. \end{aligned}$$

If the assumption on  $c$  being sufficiently small (equivalently  $a_n^v$  being sufficiently negative to make the problem concave) is violated then the optimal strategy might not exist if one does not restrict the set of admissible strategies. However, as it is usually assumed that the temporary transaction costs are greater than half the spread which is represented by  $c/2$  the problem will most of the time be convex in practical applications and can be solved in closed form.

In the case when  $c = 0$ , we see that  $y_n + z_n = x_{n-1}$ , where  $y_n = d_{n-1}^v x_{n-1}$ . Furthermore when  $q_1, q_2$  are small as the actual dark pool execution probabilities would be in practice  $d_{n-1}^v \in (0, 1)$ . This means that in a given period the optimal strategy will be to execute a fraction of the remaining position directly in the regular exchange and put the rest into the dark pool. Moreover, note that  $d_{n-1}^v$  is decreasing in  $q_1$  and  $q_2$ , therefore, one allocates more shares to the dark pool as the probability of execution increases. The special case of having  $q_1 = q_2 = 0$  produces the risk neutral optimal strategy Theorem 2.2.1. One interesting thing to note is that depending on parameter values one may have optimal strategies such that  $y_n + z_n > x_{n-1}$ . This might seem counter intuitive at first, because one might expect the strategy to decide

on a schedule  $y_n$  and choose  $z_n = x_{n-1} - y_n$  so that when the order in the dark pool executes the program will be over. However, depending on parameter values it will be optimal to put more orders into the dark pool than one needs to acquire the necessary position. When one acquires more shares than needed the direction of the execution problem reverses and the strategy will start to sell and sales will contribute positively to the execution cost as the acquisition of the extra shares came at no cost through the dark pool. The following example illustrates this idea.

**Example 3.2.2.** *Assume that  $N = 2$ , and consider a degenerate case where  $\eta_n \equiv \eta$ . From Theorem 3.2.1, it follows that  $a_1^v \equiv a_1 = \eta$ ; therefore,  $d_0^v \equiv d_0$  and  $e_0^v \equiv e_0$ . So,*

$$d_0 + e_0 = \frac{\eta^2(2 - \bar{q}) - (1 - \bar{q})\eta c/2 - c^2/4}{\eta^2(2 - \bar{q}) - (1 - \bar{q})\eta c - \bar{q}c^2/4}.$$

and  $d_0 + e_0 > 1$  if  $\eta > c/2 > 0$ .

## Simulation

For the numerical examples we present, we have simulated 50,000 paths of  $\sigma_n$ ,  $\eta_n$ ,  $b_n$ ,  $\xi_n$ . Figure 3.1 shows the simulation mean of portfolio positions;  $\bar{x}_n$ , and orders submitted to the regular exchange;  $\bar{y}_n$  for the optimal strategy of Theorem 3.2.1, risk neutral optimal strategy of Theorem 2.2.1 and constant speed strategy of Bertsimas and Lo (1998). We note that the risk neutral strategy 2.2.2 and the constant speed strategy do not take into account the existence of a dark pool. The constant speed strategy treats the liquidity to be constant; however, the strategy 2.2.2 observes the liquidity process  $\eta_n$ . It is seen that that on average the optimal strategy submits less orders to the regular exchange than Cheridito and Sepin (2014) because of the possibility of execution in the dark pool. Figure 3.2 shows  $\bar{d}_{n-1} + \bar{e}_{n-1}$ , the total ratio of orders submitted to the remaining stock position for expected  $d_{n-1} + e_{n-1}$  under the steady state distribution of  $(\sigma_n, \eta_n)$ . Notice that this ratio is slightly higher than the remaining stock position to be acquired

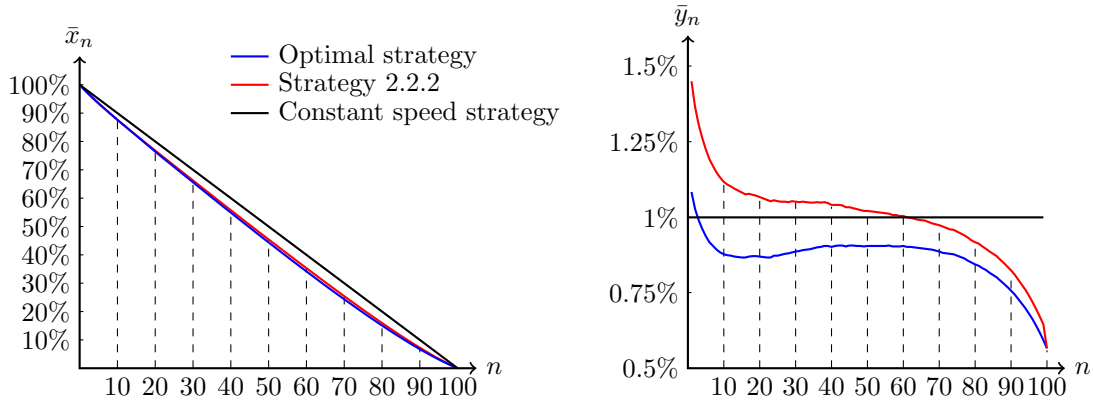


Figure 3.1: Figures for average positions  $\bar{x}_n$ ,  $\bar{y}_n$  over different scenarios in the risk neutral case for the optimal strategies of Theorem 3.2.1, 2.2.1 and constant speed strategy.

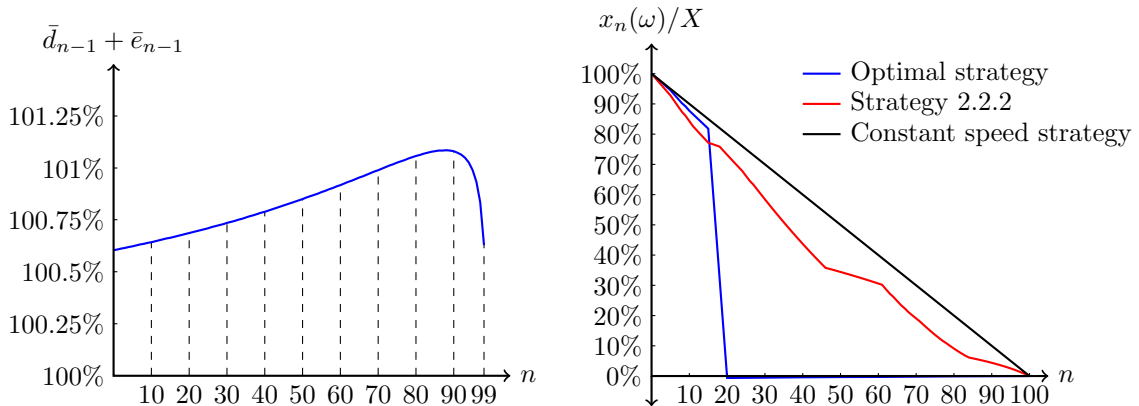


Figure 3.2: Long term averages of  $d_{n-1} + e_{n-1}$  and trajectory of three different strategies under a realization in the risk neutral case.

and is consistent with the observation in Example 3.2.2. On the right hand side of Figure 3.2 we see the three different execution strategies on the same realization of outcomes where the order submitted to the dark pool is crossed. The histograms of Figure 4.4 have been computed on the same realization of outcomes. It is seen that the optimal strategy produces a better mean of implementation shortfall than than the risk neutral optimal strategy of Chapter 2 and coincidentally a worse standard deviation.

### 3.3 Expected exponential cost

Our goal now is to minimize the expected exponential

$$\mathbb{E}_0^v \exp(\alpha C(y, z)), \quad v \in V, \quad (3.3.1)$$

for an absolute risk aversion parameter  $\alpha \geq 0$ . Denote by  $\mathcal{A}_n(x)$  the set of admissible strategies  $(y, z)$  with  $x_n = x$ , and define the value function

$$J_n^v(x) := \min_{(y,z) \in \mathcal{A}_n(x)} \mathbb{E}_n^v \exp(\alpha C_n(y, z)),$$

where  $C_n(y, z) := \sum_{i=n+1}^N cx_i y_i + y_i^2 \eta_i - x_i \sigma_i \sqrt{\Delta t} \xi_i$ .

Introduce the function  $\psi : \{0, 1\} \times \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ ,

$$\psi(b, z, u) := \begin{cases} q_2 - (q_2 - q_1)\Phi(u) & \text{if } b = 1 \text{ and } z \geq 0 \\ q_2 + (q_2 - q_1)\Phi(u) & \text{if } b = 1 \text{ and } z < 0 \\ 1 - q_2 + (q_2 - q_1)\Phi(u) & \text{if } b = 0 \text{ and } z \geq 0 \\ 1 - q_2 - (q_2 - q_1)\Phi(u) & \text{if } b = 0 \text{ and } z < 0. \end{cases}$$

where  $\Phi$  is the cumulative density function of a standard normal variable. Then the following holds:

**Theorem 3.3.1.** For  $c \geq 0$  sufficiently small, the value function  $J$  satisfies the Bellman equation

$$\begin{aligned}
J_{N-1}^v(x_{N-1}) &= \sum_{w \in V} p_{N-1}^{vw} \exp(\alpha x_{N-1}^2 w_2) \\
J_{n-1}^v(x_{n-1}) &= \min_{y_n, z_n} \sum_{w \in V, b \in \{0,1\}} p_{n-1}^{vw} \psi(b, z_n, \alpha w_1 \sqrt{\Delta t} (x_{n-1} - y_n - bz_n)) \times \\
&\quad \exp\left(\alpha c(x_{n-1} - y_n - bz_n)y_n + \alpha y_n^2 w_2 + \frac{\alpha^2}{2} w_1^2 \Delta t (x_{n-1} - y_n - bz_n)^2\right) \times \\
&\quad J_n^w(x_{n-1} - y_n - bz_n)
\end{aligned}$$

for  $n \leq N - 1$  and the minimizing  $(y_n^*, z_n^*)$  form an optimal strategy that minimizes the objective function (3.3.1).

The Bellman equations of Theorem 3.3.1 can be solved by discretizing the space of admissible controls  $\mathcal{A}$ .

## Simulation

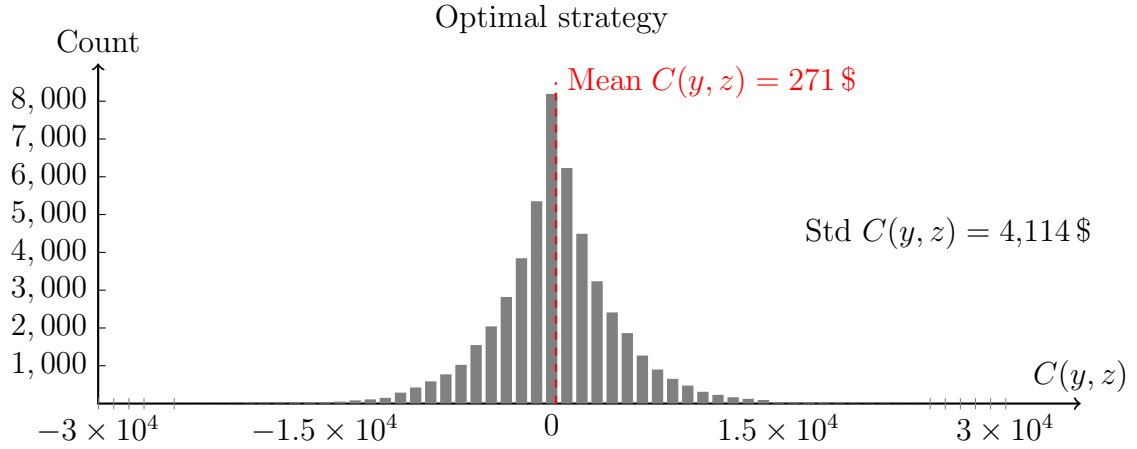
This sections numerical examples have also been produced by simulating 50,000 paths of  $\sigma_n, \eta_n, b_n, \xi_n$ . In the simulations the strategy of Theorem 3.3.1 is restricted to orders such that  $y_n + z_n = x_{n-1}$  for computational convenience. Figure 3.4 shows the simulation mean of portfolio positions  $\bar{x}_n$ , for the restricted optimal strategy of Theorem 3.3.1, the strategy that minimizes the expected exponential in Chapter 2.3.1 and the strategy given in Theorem 3.2.1 when  $\alpha = 10^{-5}$ . As in the previous section the strategy of Theorem 2.3.1 does not submit orders to the dark pool. On the right hand side of Figure 3.4 we see the restricted optimal strategy of Theorem 3.3.1 for different absolute risk aversion parameter values. Notice that the speed of execution is increasing in the absolute risk aversion parameter. The histograms of Figure 3.5 have been computed on the same realization of outcomes as of the histograms of Figure 4.4 for strategies when  $\alpha = 10^{-5}$ . It is seen that the optimal strategy produces a better mean of implementation shortfall than the optimal strategy of Theorem 2.3.1



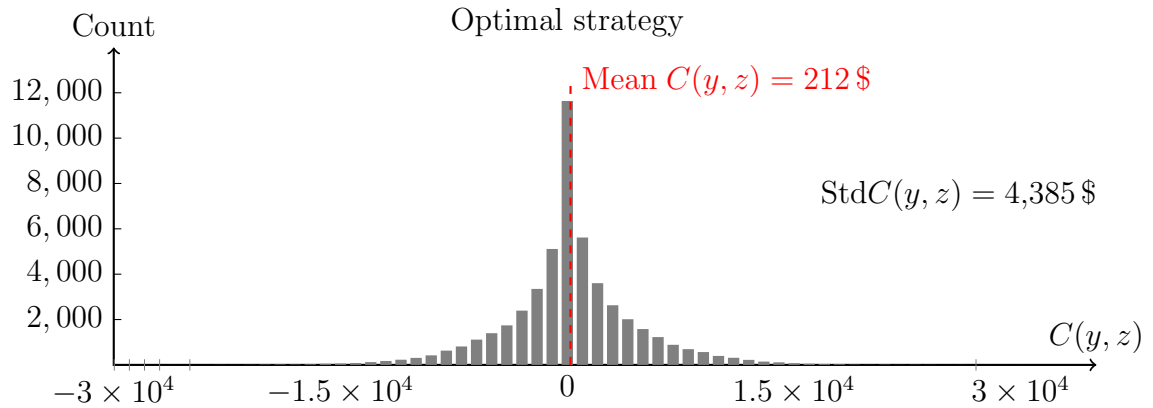
for the expected exponential case and a worse standard deviation. Both strategies on the other hand have produced less standard deviation than the risk neutral optimal strategy.

## **3.4 Conclusion**

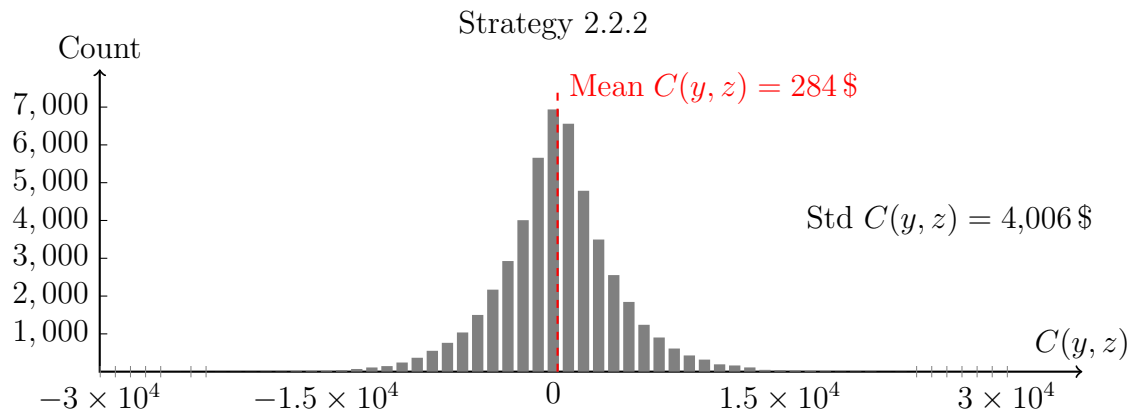
In this chapter we have extended the model in Chapter 2 to include a dark pool. In the dark pool, it is assumed that the orders are executed at the current mid-price of the primary exchange. Therefore trading in the dark pool is affected by the permanent price impact of orders executing at the standard exchange. On the other hand, not all orders submitted to the dark pool are crossed. Moreover, trading in the dark pool is exposed to adverse selection risk. We modeled this effect by making the dark pool execution probability of an order dependent on the next price innovation. The simulations showed that costs can be reduced by submitting orders to a dark pool in addition to the standard exchange.



(a) Histogram of optimal strategy with  $q_1 = 1/1000$  and  $q_2 = 5/1000$



(b) Histogram of optimal strategy with  $q_1 = 4/1000$  and  $q_2 = 2/1000$



(c) Histogram of optimal strategy 2.2.2

Figure 3.3: Histograms of  $C(y, z)$  for the optimal strategy and the strategy in the risk neutral case of Chapter 2.

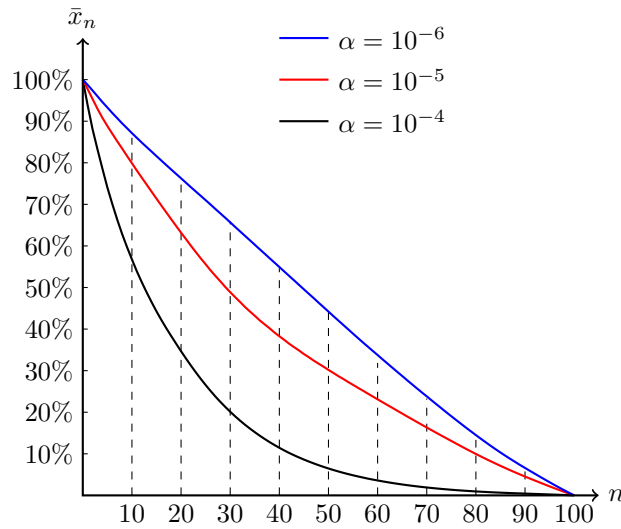
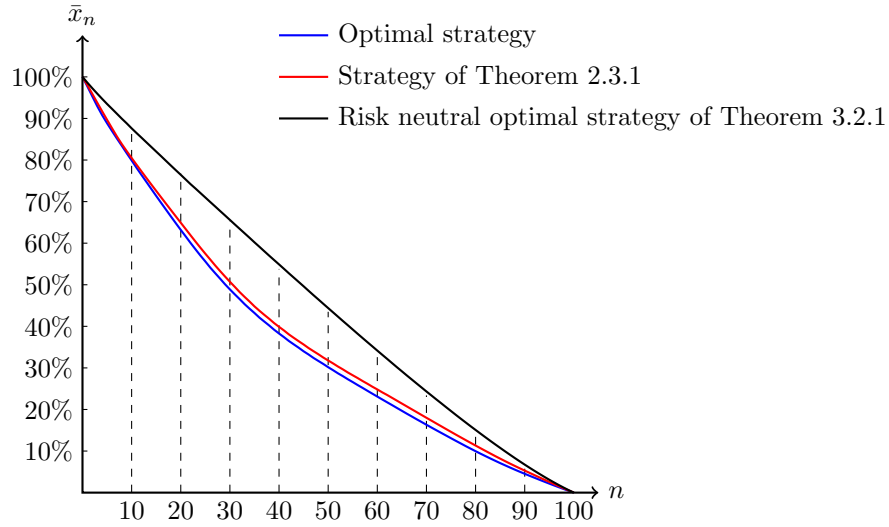
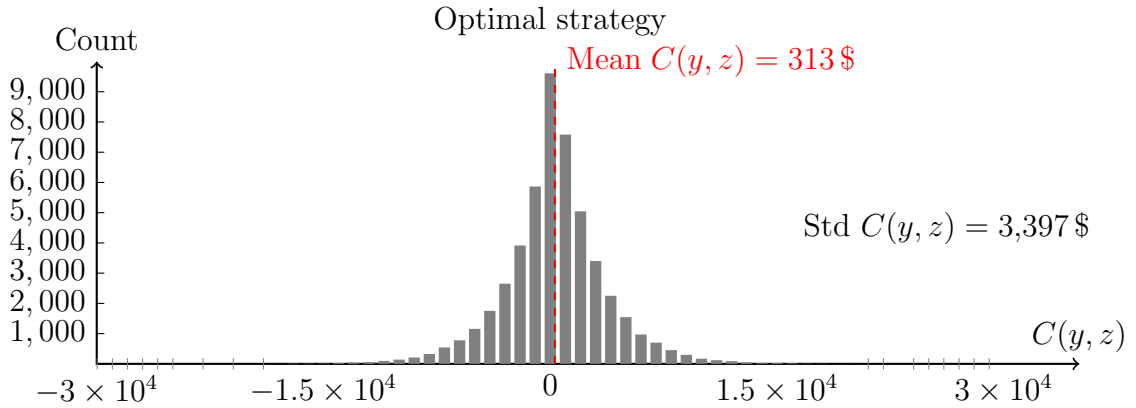
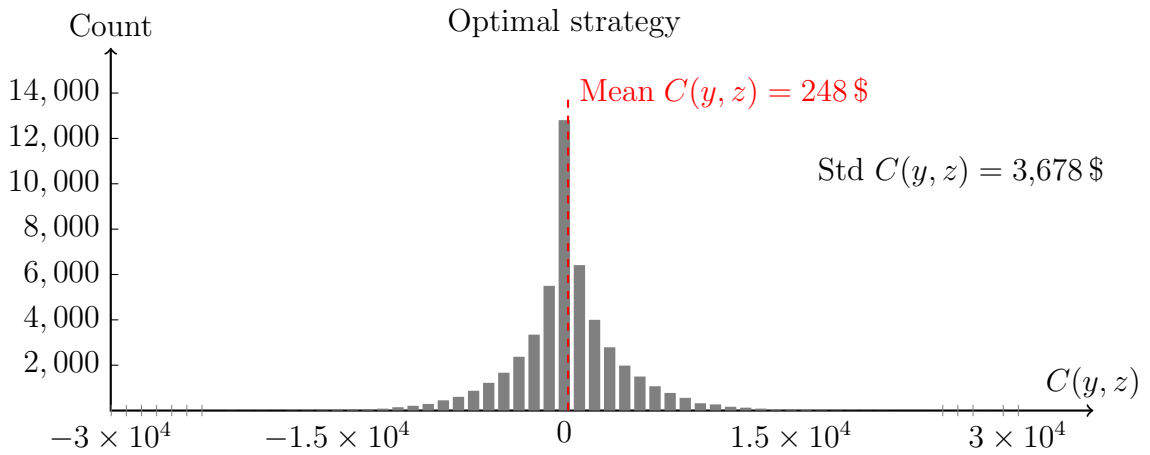


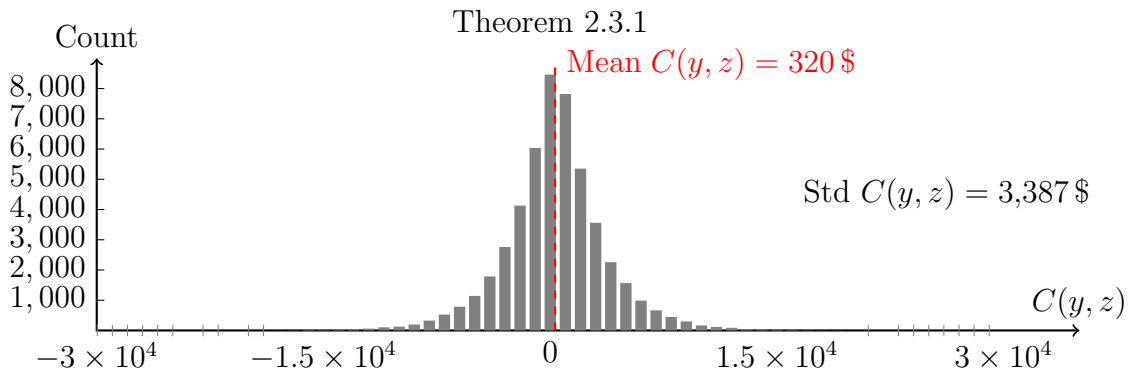
Figure 3.4: Figure on top shows average positions  $\bar{x}_n$  over different scenarios in the expected exponential case for the optimal strategies of Theorem 3.3.1, of Theorem 2.3.1 with  $\alpha = 10^{-5}$  and risk neutral optimal strategy of Theorem 3.2.1. Figure at the bottom shows average positions  $\bar{x}_n$  for the optimal strategy of Theorem 3.3.1 for three different parameter values of  $\alpha$ .



(a) Histogram of optimal strategy with  $q_1 = 1/1000$  and  $q_2 = 5/1000$



(b) Histogram of optimal strategy with  $q_1 = 4/1000$  and  $q_2 = 2/1000$



(c) Histogram of optimal strategy of Cheridito and Sepin (2014)

Figure 3.5: Histograms of  $C(y, z)$  for the optimal strategies in the expected exponential cost case under the restricted optimal strategy of Theorem 3.3.1 and Theorem 2.3.1 with  $\alpha = 10^{-5}$ .

# Chapter 4

## Multi–asset execution with a dark pool

In the case of executing orders for a portfolio of assets one also needs to consider market impact relations across assets to construct an optimal trading schedule. In this chapter we address a portfolio execution problem in the presence of stochastic market impact, stochastic liquidity and a dark pool.

In our setup we follow a multi–asset version of the impact model introduced in Chapter 2. However, the way we incorporate the dark pool is different than Chapter 3 and we closely follow the model of Cheridito and Sepin (2014c). First of all, we do not explicitly model adverse selection risk but account for it through an aversion term that penalizes the orders submitted to the dark pool. Moreover we assume that the available liquidity for each asset, hence the execution probability of each order submitted to the dark pool, can be determined indirectly by an observable Markov process. The volatility, market impact and dark pool liquidity are then allowed to be dependent on each other. Under these assumptions we find predictable strategies that minimize the expectation and the expected exponential of a cost functional that consists of the implementation shortfall of the execution program and an aversion

term for the orders submitted to the dark pool. The aversion term is a quadratic form that penalizes the orders crossed in the dark pool, which may reveal exploitable information about the position being executed. The strategy minimizing the expectation of the cost functional does not directly account for volatility risk of the portfolio positions and therefore trades slower on average. We compare the simulation mean and standard deviation of the multi-asset optimal strategy we find to ones obtained by using the risk neutral objective minimizing single-asset strategies found in Chapter 2 without a dark pool. It is observed that the multi-asset optimal strategy yields lower simulation mean and standard deviation for the cost of execution. In the expected exponential cost minimization case we provide the Bellman equation that the optimal strategy satisfies. However, the computation of optimal strategies is in most cases not possible by using the characterization of the optimal solution via Bellman equations, because it relies on using a numerical discretization procedure which may not be computationally feasible. Therefore, we first solve the problem by minimizing the expectation of the second order Taylor expansion of the exponential cost and then provide a general framework via duality which can be used to generate approximate strategies by specifying probabilities on the paths of outcomes. In all the cases we consider, the strategies we find are characterized as solutions of a set of stochastic forward-backward equations. Here we also observe that the multi-asset optimal strategy minimizing the expectation of the second order Taylor expansion of the cost yields lower simulation mean and standard deviation for the implementation cost, compared to the ones obtained from the single-asset strategies of Chapter 2 in the case where the expected exponential of the cost is minimized. Here we generalize Kratz and Schöneborn (2013a) who study the same problem in a discrete-time model. They first show the existence and uniqueness of an optimal solution in a framework with a general deterministic impact function and then find strategies that minimize the expectation of a cost function that consists of the implementation shortfall and a

quadratic portfolio risk term with constant volatility under a constant linear transient impact and dark pool execution probability.

In the next section we describe the market model. Section 4.2 describes the setup we use for the numerical studies. Section 4.3 deals with the risk neutral objective case. In Section 4.4 the expected exponential of the cost case is studied. Section 4.5 concludes and all proofs are given in Appendix A.3.

## 4.1 The model

We are interested in executing orders for a portfolio consisting of  $M$  different stocks with initial positions  $X_0 \in \mathbb{R}^M$ . Our goal is to attain a portfolio with final positions in the stocks given by  $X_N \in \mathbb{R}^M$ . Note that without loss of generality we can set  $X_N = 0$ ; therefore, we assume that initial position is specified for a portfolio such that  $X_N = 0$ . We can place an order to sell  $y_n^j \in \mathbb{R}$  shares for each asset  $j$  at every time  $t_{n-1} = (n-1)\Delta t$ . If  $y_n^j < 0$  it means that a buy order is submitted. We use  $y_n$  to denote the vector of orders submitted  $[y_n^1 \ y_n^2 \ \dots \ y_n^M]'$ . We assume that orders  $y_n^j$  execute in the interval  $[t_{n-1}, t_n)$  at a price of

$$\tilde{S}_n^j = S_{n-1}^j - \sum_{i=1}^M \eta_n^{ji} y_n^i. \quad (4.1.1)$$

$(S_n^j)$  denotes the mid-price process for asset  $j$ . By  $S_n$  we denote  $[S_n^1 \ S_n^2 \ \dots \ S_n^M]'$  and similarly by  $\tilde{S}_n$  the vector of execution prices  $[\tilde{S}_n^1 \ \tilde{S}_n^2 \ \dots \ \tilde{S}_n^M]'$ .  $(\eta_n^{ji})$  is a stochastic process modeling the temporary impact on the price of asset  $j$  which results from executing an order for asset  $i$  in the interval  $[t_{n-1}, t_n)$ . We denote the temporary impact process in matrix form by  $\Lambda_n = [\eta_n^{ji}]$ . Hence by using (4.1.1) we express  $\tilde{S}_n$  as

$$\tilde{S}_n = S_{n-1} - \Lambda_n y_n. \quad (4.1.2)$$

We assume that the dynamics of the mid-price process for each asset is given by

$$S_n^j = S_{n-1}^j + \sigma_n^j \sqrt{\Delta t} \xi_n^j. \quad (4.1.3)$$

$(\sigma_n^j)$  denotes a stochastic volatility process.  $\xi_n = [\xi_n^1 \ \xi_n^2 \ \dots \ \xi_n^M]'$  is a sequence of innovations with distribution  $\mathcal{N}(0, \Sigma)$  with  $\Sigma$  positive definite. We use  $\sigma_n$  to denote the vector  $[\sigma_n^1 \ \sigma_n^2 \ \dots \ \sigma_n^M]'$ . Furthermore,  $\text{diag}(\sigma_n)$  is used to denote the  $M \times M$  matrix with  $\sigma_n$  on its diagonal and with other entries zero. The vector mid-price process reads as,

$$S_n = S_{n-1} + \sqrt{\Delta t} \text{diag}(\sigma_n) \xi_n. \quad (4.1.4)$$

We assume that one can also submit orders to a dark pool to make use of hidden liquidity and reduce transaction costs. We use a stochastic process  $(\delta_n^j)$  to model the liquidity dynamics of the dark pool for asset  $j$ . It is assumed that  $\delta_n^j$  take distinct and finitely many values in a set  $D \subseteq \mathbb{R}$ . We assume that the orders submitted are of fill or kill type, that is when an order of  $z_n^j$  shares is submitted to the dark pool at time  $t_{n-1}$ , it results in the execution of 0 or  $z_n^j$  shares at a price of  $S_{n-1}^j$  and the order is cancelled immediately if it is not filled. The vector of orders submitted and filled are denoted by  $z_n$  and  $\text{diag}(b_{n-1}) z_n$ , where  $b_{n-1} = [b_{n-1}^1 \ b_{n-1}^2 \ \dots \ b_{n-1}^M]'$  denotes a vector of Bernoulli random variables. We define  $q^j : D \mapsto [0, 1]$  to denote the probability of success for the random variable  $b_{n-1}^j$ . Given  $\delta_{n-1}^j$ ,  $b_{n-1}^j$  is 1 with probability  $q^j(\delta_{n-1}^j)$ .  $\text{diag}(q(\delta_n))$  denotes the matrix with vector  $[q^1(\delta_n^1) \ q^2(\delta_n^2) \ \dots \ q^M(\delta_n^M)]$  on its diagonal and other entries zero.

We call the pairs  $(y_n, z_n) \in \mathbb{R}^M \times \mathbb{R}^M$  an execution strategy and denote by  $x_n$  the remaining orders to be executed at time  $t_n$  where  $x_n = X_0 - \sum_{i=1}^n (y_i + \text{diag}(b_{i-1}) z_i)$ . We suppose that  $\Lambda_n, \sigma_n, \delta_n, x_n$ , and  $S_n$  can be observed and define the filtration that they generate as

$$\mathcal{F}_n := \sigma(S_i, \Lambda_i, \sigma_i, \delta_i, x_i : -\infty < i \leq n).$$



An execution strategy is admissible if it is predictable with respect to  $(\mathcal{F}_n)$ ,  $x_0 = X_0$ ,  $x_N = 0$  and  $z_N = 0$ . This means that an admissible strategy must start with the specified initial stock positions and execute orders such that at the deadline the desired portfolio is attained. The condition  $z_N = 0$  ensures that  $x_N = 0$  is attained by forcing  $y_N = x_{N-1}$ . We denote the set of all admissible strategies by  $\mathcal{A}$ .

We assume that there is adverse selection risk which may result from information leakage when  $z_n^j$  is crossed in the dark pool. We model aversion to adverse selection by a risk term of the form  $\alpha z_n' \text{diag}(b_{n-1}) z_n$  for a parameter  $\alpha \geq 0$ . We define our cost functional as the sum of implementation shortfall and adverse selection risk. Implementation shortfall is defined as the difference of the initial portfolio and the execution value. The execution value is given by  $\sum_{n=1}^N (\tilde{S}'_n y_n + S'_{n-1} \text{diag}(b_{n-1}) z_n)$ , therefore for  $(y, z) \in \mathcal{A}$  the cost functional reads as,

$$C(x, y, z) = S'_0 X_0 - \sum_{n=1}^N \left( \tilde{S}'_n y_n + S'_{n-1} \text{diag}(b_{n-1}) z_n \right) + \alpha z_n' \text{diag}(b_{n-1}) z_n \quad (4.1.5)$$

One can express (4.1.5) also as

$$\begin{aligned} C(x, y, z) &= \sum_{n=1}^N y_n' \Lambda'_n y_n - \sqrt{\Delta t} \xi_n' \text{diag}(\sigma_n) (x_{n-1} - y_n - \text{diag}(b_{n-1}) z_n) \\ &\quad + \alpha z_n' \text{diag}(b_{n-1}) z_n. \end{aligned} \quad (4.1.6)$$

We assume that  $\xi_n$  is independent of  $\sigma(\mathcal{F}_{n-1}, \Lambda_n, \sigma_n, \delta_n, x_n)$ . Moreover, given  $\sigma(\mathcal{F}_{n-1}, \Lambda_n, \sigma_n, \delta_n, \xi_n)$ ,  $b_{n-1}^j$  only depends on  $\delta_{n-1}^j$ , and given  $\delta_{n-1}$ ,  $b_{n-1}^j$  is independent of  $b_{n-1}^i$ , hence  $\text{Cov}_{n-1}(b_{n-1}^j, b_{n-1}^i) = 0$  for  $i \neq j$ . We suppose that  $\Lambda_n$  takes finitely many different values in  $\mathbb{R}_+^{M \times M}$  and  $\sigma_n$  in  $\mathbb{R}_+^M$ . We define  $(\Lambda_n, \sigma_n, \delta_n)$  as a Markov chain with finite state space  $W \subseteq \mathbb{R}^{M \times M} \times \mathbb{R}_+^M \times \mathbb{R}_+^M$  and with

time-dependent transition probabilities

$$p_{n-1}^{vw} := \mathbb{P}[(\Lambda_n, \sigma_n, \delta_n) = w \mid (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1}) = v], \quad v, w \in W.$$

Notice that it is enough to observe  $(\Lambda_n, \sigma_n, \delta_n, x_n)$  for  $n \geq 0$  to decide on the optimal strategy. By  $\mathbb{E}_n^{v,s}$  we denote the conditional expectation  $\mathbb{E}[\cdot \mid (\Lambda_n, \sigma_n, \delta_n) = v, x_n = s]$  and by  $\mathbb{E}_n^v$ ,  $\mathbb{E}[\cdot \mid (\Lambda_n, \sigma_n, \delta_n) = v]$ . We use  $\mathbb{P}_n^v(b_n = u)$  to denote  $\mathbb{P}(b_n = u \mid (\Lambda_n, \sigma_n, \delta_n) = v)$  for  $v \in W$  and  $u \in \{0, 1\}^M$ . We assume that  $\mathbb{E}_n^v[\Lambda_{n+1}]$  is positive definite. In the rest of the chapter we frequently encounter the matrix form

$$\begin{bmatrix} U & B \\ B' & L \end{bmatrix}, \quad (4.1.7)$$

for matrices  $U, L, B \in \mathbb{R}^{M \times M}$ . The matrix forms given in (4.1.7) are positive definite. For a matrix  $A \in \mathbb{R}^{M \times M}$  we use  $\tilde{A}$  to denote the symmetric matrix  $A + A'$  and  $\text{Tr}(A)$  to denote the trace of  $A$ .

## 4.2 Parameter set for simulations

In this section we specify the parameters of the model for the numerical experiments we consider for risk neutral and expected exponential of the cost objectives. In our numerical studies we consider a portfolio of 2 assets. We consider two hypothetical companies; Company A and Company B and denote them by A and B.

We assume that  $\sigma_n^A = \sigma_n^B$ . This can be thought of having a common factor that derives the overall volatility level of market innovations. For instance if one is executing a portfolio of assets traded in S&P500 the volatility of S&P500 index can be used as the common factor. We denote the common factor volatility by  $\sigma_n^{CF}$ . We assume that the state space of  $\sigma_n^{CF}$  consists of three states and is given by

$\{\sigma_{low}^{CF}, \sigma_{med}^{CF}, \sigma_{high}^{CF}\}$ . Notice that in practice one may keep track of specific indices (i.e. financials, consumer goods) and can construct the volatility process  $\sigma_n$  by aggregating the volatility dynamics derived for the desired indices. Following such a procedure would allow to capture cross sectional dynamics more effectively. For applications  $\Sigma$ , the covariance matrix of market innovations, must be calculated by using statistical procedures. In practice,  $\Sigma$  is usually supplied by a third party and is integrated into the execution program. We use the following volatility parameter set, which is along the lines of Chapter 2.

$$\Sigma = \begin{bmatrix} 2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & 3/2 \end{bmatrix} \begin{bmatrix} \sigma_{low}^{CF} \\ \sigma_{med}^{CF} \\ \sigma_{high}^{CF} \end{bmatrix} = \begin{bmatrix} 3.51 \times 10^{-3} \\ 3.3 \times 10^{-2} \\ 1.172 \times 10^{-1} \end{bmatrix},$$

$$(p_{ij}^\sigma) = \begin{pmatrix} 0.93 & 0.04 & 0.03 \\ 0.72 & 0.22 & 0.06 \\ 0.44 & 0.48 & 0.08 \end{pmatrix}.$$

We assume that impact process follows the dynamics given by

$$\eta_n^{ji} = \eta_n^{jj} \frac{\text{Cov}_{n-1}(S_n^i, S_n^j)}{\sqrt{\text{Var}_{n-1}(S_n^i)} \sqrt{\text{Var}_{n-1}(S_n^j)}}. \quad (4.2.8)$$

Notice that  $\eta_n^{jj}$  can be estimated more easily as opposed to  $\eta_n^{ji}$  because its effect can be observed directly when one submits an order only for one asset. Then one can argue that cross impact will be proportional as in (4.2.8) with the coefficient of correlation of stock prices. In our case it computes as  $\eta_n^{ji} = \eta_n^{jj} \Sigma_{ij} / \sqrt{\Sigma_{jj}} \sqrt{\Sigma_{ii}}$  because we assume that  $\sigma_n = \sigma_n^{CF}$ . The impact matrix is then given by

$$\Lambda_n = \text{diag}(\eta_n) \text{diag}\left(\frac{1}{\sqrt{\Sigma_{jj}}}\right) \Sigma \text{diag}\left(\frac{1}{\sqrt{\Sigma_{jj}}}\right) \quad (4.2.9)$$

where  $\text{diag}(\eta_n)$  is the diagonal matrix with  $\eta_n = [\eta_n^A \ \eta_n^B]$  and  $\text{diag}\left(\frac{1}{\sqrt{\Sigma_{jj}}}\right)$  is the diagonal matrix with  $\frac{1}{\sqrt{\Sigma_{jj}}}$  terms on its diagonal. Notice that as long as  $\text{diag}(\eta_n)$  takes positive values in all states,  $\Lambda_n$  as defined in (4.2.9) will be positive definite because  $\Sigma$  was assumed to be positive definite hence all of  $\Lambda_n$ 's principle minors are positive for all states. We describe the state space of  $\eta_n^j$  by  $\{\eta_{low}^j, \eta_{med}^j, \eta_{high}^j\}$  for  $j \in \{A, B\}$  and specify the following state spaces,

$$\begin{bmatrix} \eta_{low}^A \\ \eta_{med}^A \\ \eta_{high}^A \end{bmatrix} = \begin{bmatrix} 2 \times 10^{-6} \\ 5.5 \times 10^{-6} \\ 32 \times 10^{-6} \end{bmatrix}, \quad \begin{bmatrix} \eta_{low}^B \\ \eta_{med}^B \\ \eta_{high}^B \end{bmatrix} = \begin{bmatrix} 1 \times 10^{-6} \\ 3 \times 10^{-6} \\ 35 \times 10^{-6} \end{bmatrix},$$

and transition probabilities,

$$(p_{ij}^A) = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.5 & 0.1 \\ 0.3 & 0.6 & 0.1 \end{pmatrix}, \quad (p_{ij}^B) = \begin{pmatrix} 0.55 & 0.35 & 0.10 \\ 0.15 & 0.70 & 0.15 \\ 0.1 & 0.1 & 0.8 \end{pmatrix},$$

We assume that  $(\eta_n^A)$ ,  $(\eta_n^B)$ , and  $(\sigma_n^{CF})$  are independent Markov chains with stationary transition probabilities. We define the dark pool process as follows,

$$\delta_n^j := \begin{cases} 1 & \text{if } \eta_n^j = \eta_{low}^j, \\ 0 & \text{otherwise.} \end{cases}$$

and specify the execution probabilities as  $q^j(1) = 1/500$  and  $q^j(0) = 1/1000$  for  $j \in \{A, B\}$ . Other parameter values we use are given in Table 4.1.

Parameter	Symbol	Value
Initial stock prices	$S_0^A, S_0^B$	175, 161 \$/share
Initial positions	$X_0^A, X_0^B$	30000, -25000 shares
Duration	$T$	50 minutes
Number of subintervals	$N$	50
Length of subintervals	$\Delta t$	1 minute

Table 4.1: Parameter values

### 4.3 Risk neutral objective

In this section we want to find the minimizer  $(y^*, z^*) \in \mathcal{A}$  of  $\mathbb{E}_0^{v, X_0} [C(x, y, z)]$  for  $v \in W$ . Let us define the set  $\mathcal{A}_n(s)$  as the set of admissible strategies such that  $x_n = s$ ,  $(y_i, z_i)_{i \geq n+1}$  is  $\mathcal{F}_i$  predictable and  $(x_N, z_N) = (0, 0)$ . Let us define

$$Q_n(x, y, z) := \sum_{i=n+1}^N y'_i \Lambda'_i y_i - \sqrt{\Delta t} \xi'_i \text{diag}(\sigma_n) (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i) + \alpha z'_i \text{diag}(b_{i-1}) z_i$$

and for  $v \in W$  define the value function

$$J_n^v(s) := \min_{(y, z) \in \mathcal{A}_n(s)} \mathbb{E}_n^{v, s} [Q_n(x, y, z)].$$

By a backward recursion we define the following matrices,

$$\begin{aligned} A_n^v &:= G_n^{v'} \mathbb{E}_n^v [\Lambda'_{n+1}] G_n^v + \alpha F_n^{v'} \text{diag}(q(v_3)) F_n^v \\ &+ \sum_w p_n^{vw} \left( F_n^{v'} \text{diag}((A_{n+1}^w)_{ii}) \text{diag}(q(v_3)(1 - q(v_3))) F_n^v \right. \\ &\left. + (I - G_n^v - \text{diag}(q(v_3)) F_n^v)' A_{n+1}^w (I - G_n^v - \text{diag}(q(v_3)) F_n^v) \right) \end{aligned}$$

for  $n \leq N - 2$  and  $A_{N-1}^v := \mathbb{E}_{N-1}^v [\Lambda'_N]$ . The matrices  $G_n^v$  and  $F_n^v$  are defined as

$$F_n^v := (S_n^v)^{-1} B_n^{v'} \left( I - (U_n^v)^{-1} \sum_w p_n^{vw} \tilde{A}_{n+1}^w \right), \quad G_n^v := (U_n^v)^{-1} \left( \sum_w p_n^{vw} \tilde{A}_{n+1}^w - B_n^v F_n^v \right)$$

where  $S_n^v$  is the Schur complement in  $U_n^v$  of the symmetric matrix

$$\begin{bmatrix} U_n^v & B_n^v \\ B_n^{v'} & L_n^v \end{bmatrix}$$

with  $U_n^v := \mathbb{E}_n^v[\tilde{\Lambda}_{n+1}] + \sum_w p_n^{vw} \tilde{A}_{n+1}^w$ ,  $B_n^v := \sum_w p_n^{vw} \tilde{A}_{n+1}^w \text{diag}(q(v_3))$ , and

$$\begin{aligned} L_n^v := & \sum_w p_n^{vw} \left( \text{diag}(q(v_3)) \tilde{A}_{n+1}^w \text{diag}(q(v_3)) + 2\alpha \text{diag}(q(v_3)) \right. \\ & \left. + 2\text{diag}((A_{n+1}^w)_{ii}) \text{diag}(q(v_3)(1 - q(v_3))) \right), \end{aligned}$$

for  $n \leq N - 2$ .

**Theorem 4.3.1.** *The optimal strategy minimizing the risk neutral objective is given by*

$$y_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v} = G_{n-1}^v x_{n-1}^*, \quad z_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v} = F_{n-1}^v x_{n-1}^*, \quad (4.3.1)$$

for  $n \leq N - 1$ .

## Simulation

Here we present the results obtained by simulating 10,000 scenarios in the setting we described in Section 4.2. All the scenarios start from the the initial state  $\sigma_0^{CF} = \sigma_{low}^{CF}$  and  $\eta_0^j = \eta_{low}^j$  for  $j \in \{A, B\}$ .  $\bar{x}_n^j$  denotes the average position in asset  $j$  at time  $t_n$  as percentage of the absolute of initial position in that asset, so negative values indicate buying and positive values indicate selling. Similarly  $\bar{y}_n^j$  and  $\bar{z}_n^j$  denote the average order sizes submitted into the regular exchange and the dark pool at times  $t_n$ . We denote by  $s_n$  the standard deviation of position  $x_n$  as observed in the simulations.

Figure 4.1 shows the simulation mean of the execution trajectory and one standard deviation around the mean. On the right side of Figure 4.1 we see one particular

realization of an execution trajectory. In this realization we see an order placed to the dark pool for Company A executes and a selling program is reversed into a buying program because the order submitted to the dark pool was larger than the number of shares that needed to be sold for Company A. The rationale behind this is better understood by examining Figure 4.2.

Figure 4.2 shows the simulation mean of orders submitted to the dark pool and to the regular exchange along with the simulation mean of execution trajectories as percentages of initial positions in the assets. Notice that the assets are negatively correlated. Therefore, one increases the stock price of Company B while selling the stock of Company A. By a similar observation we see that one will depress the stock price of Company A when buying the stock of Company B. Since the program needs to liquidate the shares for Company A and acquire the stock of Company B, every transaction made in one asset causes additional execution costs because of the correlation structure. To remedy this, the optimal strategy tries to reverse the positions in either of the assets by placing orders that are larger than the amount of shares that needs to be executed. For instance, as the program starts, an order to sell about 35% more than  $X_0^A$  is placed into the dark pool, likewise an order about 60% more than  $X_0^B$  to buy is put into the dark pool. One desirable outcome would be that only one of these orders gets crossed in the dark pool so that the negative correlation between the stocks can be taken advantage of to reduce the transaction costs in the rest of the execution. Another desirable outcome is that both dark pool orders are matched. For such an outcome, we see that the program will have reduced the amounts of shares to be executed to about 35% and 60% of initial positions at no cost. We also see that the program submits orders in the order of about 2% of the initial positions into the regular exchange to make sure that the amount of shares to be executed are decreased at each period in the case the orders in the dark pool do not get crossed.

Figure 4.3 shows the average size of sell orders submitted to the regular exchange and the dark pool for different values of adverse selection parameter  $\alpha$ . In practice, it may be risky to place large orders into the dark pool because of potential predatory trading. If it is discovered that a large position is being executed, the user program will be adversely selected by a predatory trader who will move the prices unfavorably for the user. We see that increasing the parameter of aversion for adverse selection reduces the size of the order put into the dark pool.

Figure 4.4 shows two histograms of the implementation cost. The multi-asset optimal strategy is obtained by Theorem 4.3.1 in the case when  $\alpha = 0$ . To construct the second histogram we have computed the risk neutral optimal single-asset strategy described in Theorem 2.2.1 for the two stocks we consider without the inclusion of a dark pool and by using the parameter set described in Section 4.2. Notice that since Theorem 2.2.1 considers a single-asset case the cross impact is also not accounted for in the strategies we obtain by using it. We see that if one assumes the parameter set for the cross impact and asset correlations is true and there is a dark pool where the orders may be crossed at the mid-price, then one may on average incur higher execution costs and standard deviation by not using these features. It is seen in Figure 4.4 that the multi-asset optimal strategy minimizing the expected implementation cost creates lower simulation mean and standard deviation.

## 4.4 Expected exponential cost

We fix a parameter of absolute risk aversion  $\lambda > 0$  and want to find the minimizer  $(y^*, z^*) \in \mathcal{A}$  of the expected exponential of the cost functional;  $\mathbb{E}_0^{v, X_0}[\exp(\lambda C(x, y, z))]$  for  $v \in W$ . We define the value function

$$J_n^v(s) := \min_{(y, z) \in \mathcal{A}_n(s)} \mathbb{E}_n^{v, s} [\exp(\lambda Q_n(x, y, z))]. \quad (4.4.1)$$



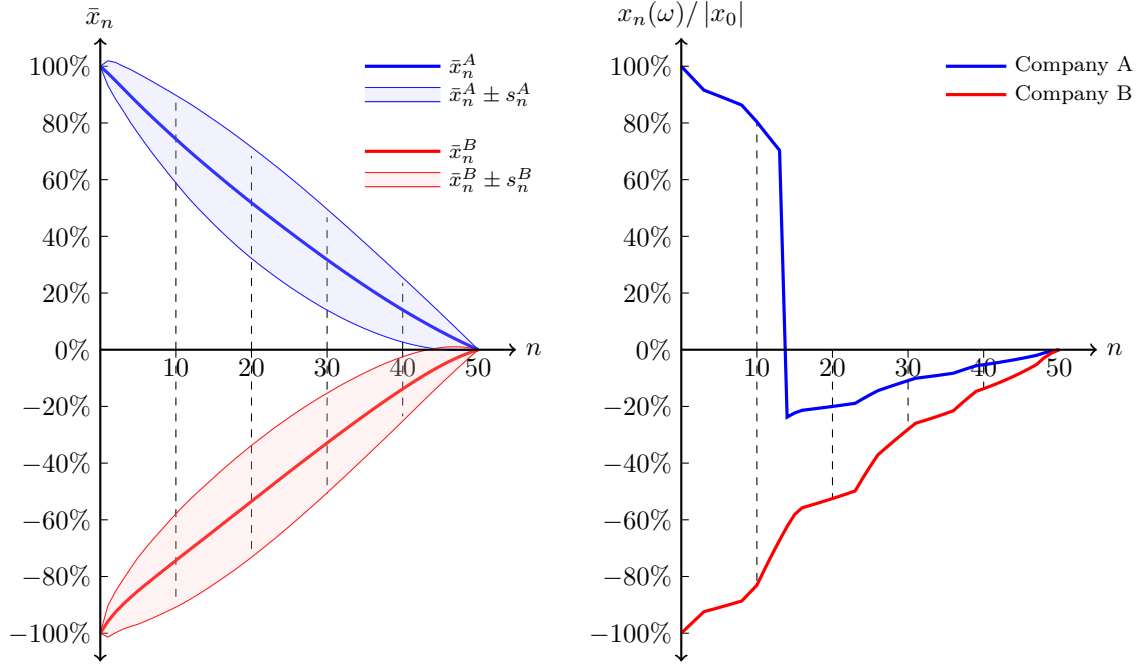


Figure 4.1: Average of net optimal positions for different scenarios and a realization of execution trajectories in the risk neutral case with  $\alpha = 0$ .

**Proposition 4.4.1.** *The value function  $J$  satisfies the Bellman equation*

$$\begin{aligned}
J_{N-1}^v(x_{N-1}) &= \sum_{w \in W} p_{N-1}^{vw} \exp\left(\lambda x'_{N-1} w'_1 x_{N-1}\right) \\
J_{n-1}^v(x_{n-1}) &= \min_{y_n, z_n} \sum_{w, u} p_{n-1}^{vw} \mathbb{P}_{n-1}^v(b_{n-1} = u) \exp\left(\lambda y'_n w'_1 y_n + \lambda \alpha z'_n \text{diag}(u) z_n\right. \\
&\quad \left. + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u) z_n)' \text{diag}(w_2) \Sigma \text{diag}(w_2)\right. \\
&\quad \left. (x_{n-1} - y_n - \text{diag}(u) z_n)\right) J_n^w(x_{n-1} - y_n - \text{diag}(u) z_n),
\end{aligned}$$

for  $n \leq N - 1$  and the minimizing  $(y_n^*, z_n^*)$  form the optimal strategy for the expected exponential of the cost.

The recursion equation of Proposition 4.4.1 must be solved by discretizing the control space. Since solving the problem on a grid is computationally very expensive, one will usually not be able to obtain the optimal strategy. We have not been able

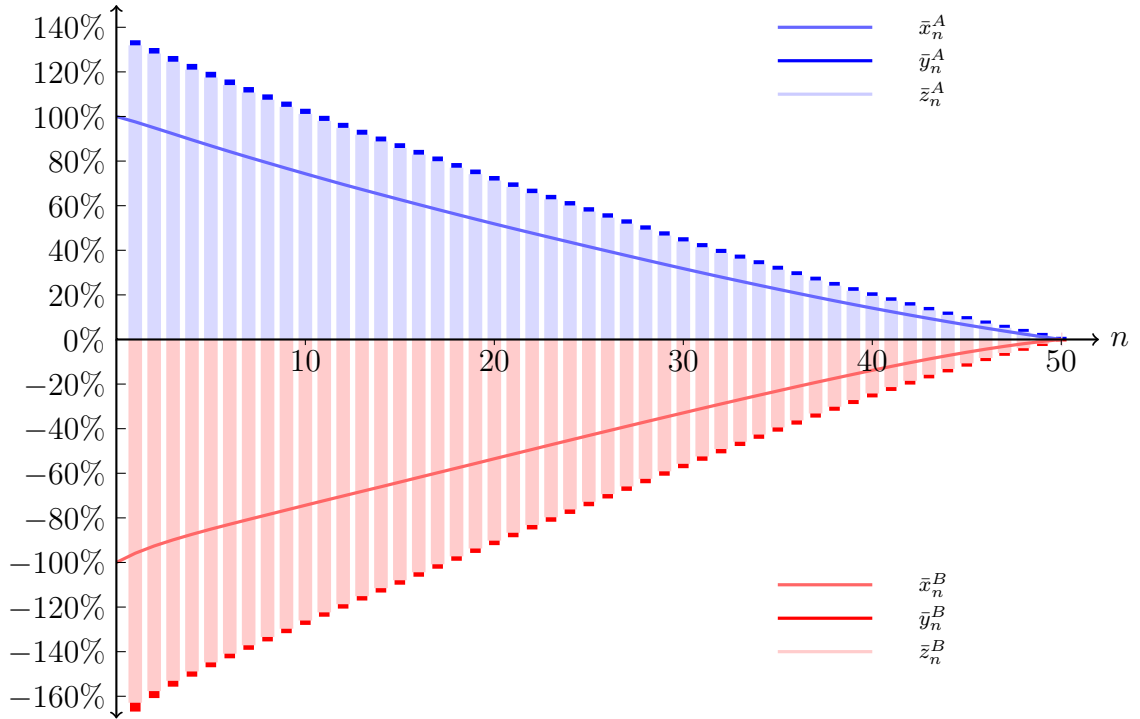


Figure 4.2: Detailed analysis of the negatively correlated pair Company A and B shows the net average stock positions, and average of order sizes submitted into the regular exchange and the dark pool as percentages of initial positions in the assets for the risk neutral case with  $\alpha = 0$ .

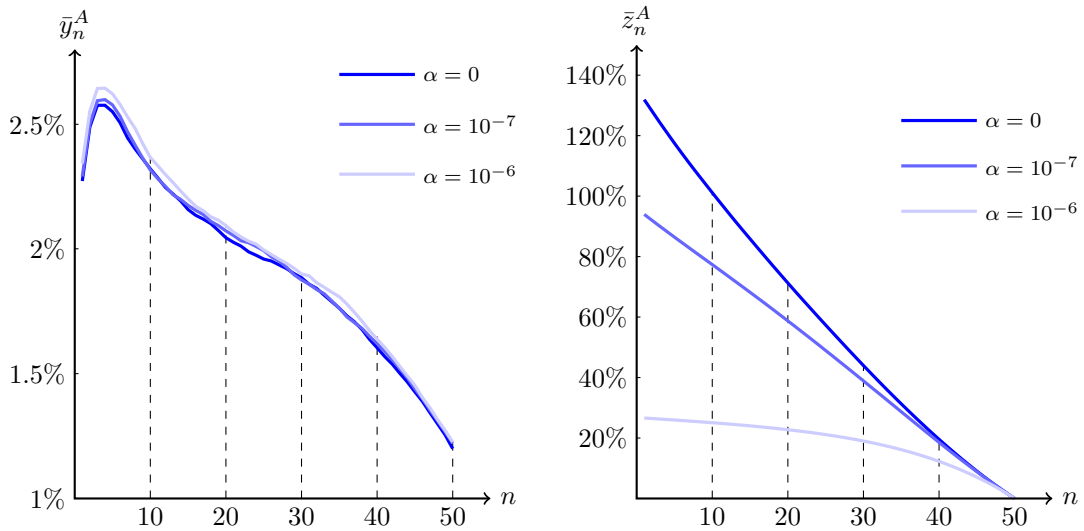


Figure 4.3: Average size of orders submitted to the regular exchange and to the dark pool for Company A as a function of the parameter  $\alpha$  in the risk neutral case.

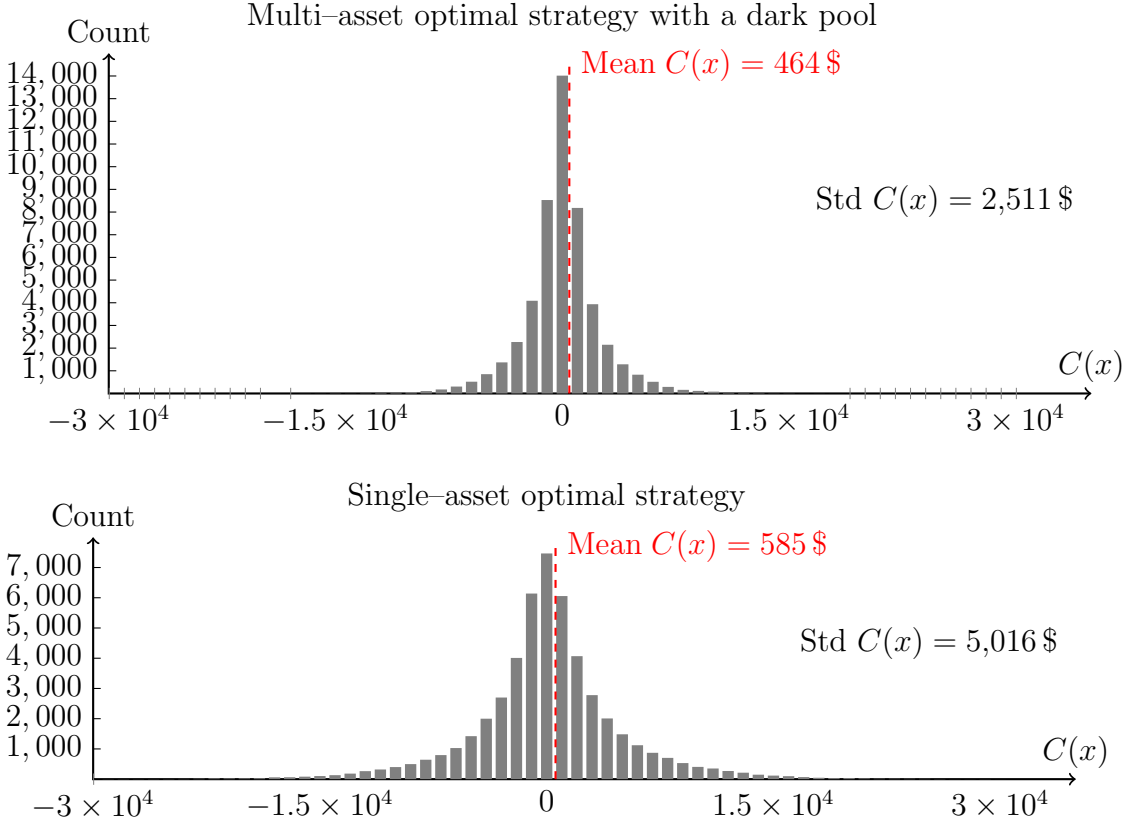


Figure 4.4: Histograms of  $C(x)$  computed for the risk neutral objective under the optimal strategy in the multi-asset case with a dark pool where  $\alpha = 0$  and the aggregation of two individual single-asset cases from from Theorem 2.2.1 without a dark pool.

to compute the optimal strategy numerically for reasonable choices of discretization parameters (i.e. restricting order submission to multiples of 2.5% of the initial positions in the assets). Therefore, we first propose to use an approximation for the exponential cost functional and then propose a dual method to generate approximate solutions to the dynamic programming equation of Proposition 4.4.1.

#### 4.4.1 Second order Taylor approximation

Let  $\hat{C}(x, y, z)$  be the second order Taylor expansion of  $\exp(\lambda C(x, y, z))$  around  $(x, y, z) = (0, 0, 0)$ . We want to find  $(y^*, z^*) \in \mathcal{A}$  that minimizes  $\mathbb{E}_0^{v, X_0}[\hat{C}(x, y, z)]$ .

Define the value function

$$J_n^v(s) := \min_{(y,z) \in \mathcal{A}_n(s)} \mathbb{E}_n^{v,s} \left[ \hat{Q}_n(x, y, z) \right],$$

where

$$\begin{aligned} \hat{Q}_n(x, y, z) := & \sum_{i=n+1}^N \frac{1}{2} y_i' \tilde{\Lambda}_i y_i + \alpha z_i' \text{diag}(b_{i-1}) z_i \\ & + \frac{\lambda}{2} \Delta t (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i)' \text{diag}(\sigma_i) \xi_i \xi_i' \text{diag}(\sigma_i) (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i). \end{aligned}$$

We recursively define the following matrices,

$$\begin{aligned} H_n^v := & \frac{1}{2} (\hat{G}_n^v)' \mathbb{E}_n^v [\tilde{\Lambda}_{n+1}] \hat{G}_n^v + (\hat{F}_n^v)' \left( \alpha \text{diag}(q(v_3)) + \left( \sum_w p_n^{vw} \text{diag}((H_{n+1}^w)_{ii}) \right. \right. \\ & \left. \left. + \frac{\lambda}{2} \Delta t \text{diag}(\mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})]_{ii}) \right) \text{diag}(q(v_3)(1 - q(v_3))) \right) \hat{F}_n^v \\ & + (I - \hat{G}_n^v - \text{diag}(q(v_3)) \hat{F}_n^v)' \left( \sum_w p_n^{vw} H_{n+1}^w + \frac{\lambda}{2} \Delta t \mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})] \right) \\ & (I - \hat{G}_n^v - \text{diag}(q(v_3)) \hat{F}_n^v) \end{aligned}$$

for  $n \leq N - 2$  and  $H_{N-1}^v := \frac{1}{2} \mathbb{E}_{N-1}^v [\tilde{\Lambda}_N]$ .  $\hat{G}_n^v$  and  $\hat{F}_n^v$  are defined as

$$\begin{aligned} \hat{F}_n^v := & (S_n^v)^{-1} (\hat{B}_n^v)' \left( I - (U_n^v)^{-1} \left( \sum_w p_n^{vw} \tilde{H}_{n+1}^w + \lambda \Delta t \mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})] \right) \right), \\ \hat{G}_n^v := & (U_n^v)^{-1} \left( \sum_w p_n^{vw} \tilde{H}_{n+1}^w + \lambda \Delta t \mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})] - \hat{B}_n^v \hat{F}_n^v \right), \end{aligned}$$

where  $\hat{S}_n^v$  is the Schur complement in  $\hat{U}_n^v$  of the matrix

$$\begin{bmatrix} \hat{U}_n^v & \hat{B}_n^v \\ (\hat{B}_n^v)' & \hat{L}_n^v \end{bmatrix}$$

with

$$\begin{aligned}\hat{U}_n^v &:= \mathbb{E}_n^v[\tilde{\Lambda}_{n+1}] + \sum_w p_n^{vw} \tilde{H}_{n+1}^w + \lambda \Delta t \mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})], \\ \hat{B}_n^v &:= \left( \sum_w p_n^{vw} \tilde{H}_{n+1}^w + \lambda \Delta t \mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})] \right) \text{diag}(q(v_3)), \quad \text{and} \\ \hat{L}_n^v &:= 2 \left( \alpha \text{diag}(q(v_3)) + \left( \sum_w p_n^{vw} \text{diag}((H_{n+1}^w)_{ii}) + \right. \right. \\ &\quad \left. \left. \frac{\lambda}{2} \Delta t \text{diag}(\mathbb{E}_n^v[\text{diag}(\sigma_{n+1}) \Sigma \text{diag}(\sigma_{n+1})]_{ii}) \right) \text{diag}(q(v_3)(1-q(v_3))) \right) + \text{diag}(q(v_3)) \hat{B}_n^v\end{aligned}$$

for  $n \leq N - 2$ .

**Theorem 4.4.2.** *The optimal strategy minimizing the expectation of second order Taylor approximation of the exponential cost is given by*

$$y_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v} = \hat{G}_{n-1}^v x_{n-1}^*, \quad z_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v} = \hat{F}_{n-1}^v x_{n-1}^*, \quad (4.4.2)$$

for  $n \leq N - 1$ .

## Simulation

Figure 4.5 shows the simulation average of net optimal stock positions for different values of parameter  $\lambda$  as percentage of the initial positions in the assets. It is observed that the average speed of execution increases as the parameter value  $\lambda$  is increased. On the right side of Figure 4.5, a realization of an execution trajectory is shown. Figure 4.6 shows a detailed analysis of the simulation average of orders submitted to the regular exchange and dark pool along with the average execution trajectory with parameter values  $\alpha = 0$ ,  $\lambda = 10^{-4}$ . As in the risk neutral case, we observe that the optimal strategy places orders larger than the remaining positions to be executed into the dark pool to reduce the transaction costs due to the correlation between Companies A and B. On the left side of Figure 4.7 the sample average of

execution trajectories and one standard deviation around this average is shown for parameter values of  $\lambda = 10^{-4}$  and  $\alpha = 0$ . On the right side of Figure 4.7 we see the effect of increasing the adverse selection parameter  $\alpha$  on the average size of sell orders submitted to the dark pool for Company A.

Figure 4.8 shows two histograms of the implementation cost. The multi-asset optimal strategy is obtained by Theorem 4.4.2 in the case when  $\lambda = 10^{-4}$  and  $\alpha = 0$ . The second histogram was created by using two individually optimal single-asset strategies found by using Theorem 2.3.1 for the expected exponential of the cost case. The single-asset strategies do not account for the dark pool, cross impact and market innovation correlations. Moreover they are obtained by a numerical procedure where one is allowed to submit orders in the multiples of 1% of the starting asset positions. It is assumed that the parameter of risk aversion is also  $10^{-4}$ . It is seen in Figure 4.8 that the multi-asset optimal strategy minimizing the expectation of the second order Taylor expansion of the exponential cost yields lower simulation mean and standard deviation.

#### 4.4.2 An approximation via duality

In this section our goal is to provide a framework so that one can generate approximate strategies by using the dual variables obtained from the dual representation of the optimization problems defined in (4.4.1). We introduce new notation, let us denote a path of the Markov chain and dark pool executions from time  $t_n$  to  $t_N$  by  $(v^n, u^n, \dots, v^{N-1}, u^{N-1}, v^N)$  where  $v^i \in W$  and  $u^i \in \{0, 1\}^M$ . We denote the probability of this outcome given  $(\Lambda_n, \sigma_n, \delta_n) = v^n$  by  $p(u^n, \dots, v^N \mid v^n)$  and  $\mu(u^n, \dots, v^N \mid v^n)$ .  $p(u^n, \dots, v^N \mid v^n)$  is given by

$$p_n^{v^n, v^{n+1}} \times \dots \times p_{N-1}^{v^{N-1}, v^N} \times \mathbb{P}_n^{v^n}(b_n = u^n) \times \dots \times \mathbb{P}_{N-1}^{v^{N-1}}(b_{N-1} = u^{N-1}),$$

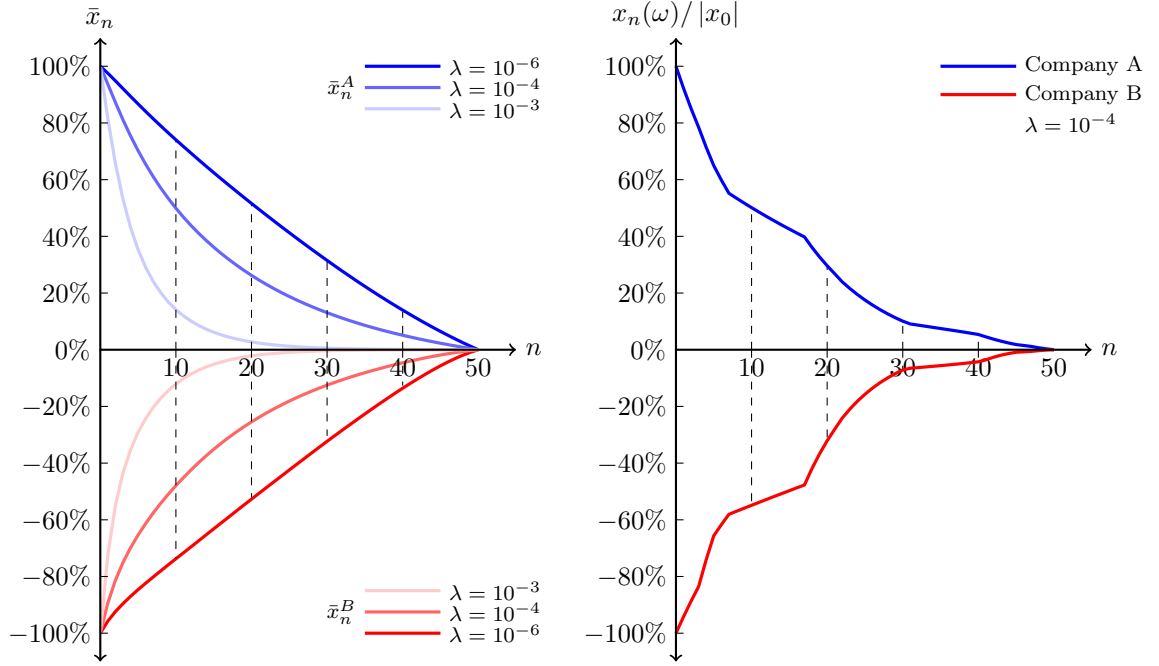


Figure 4.5: Average of net optimal positions for different scenarios as a function of risk aversion parameter  $\lambda$ , and a realization of execution trajectories in the second order Taylor approximation case with  $\alpha = 0$ .

and  $\mu(u^n, \dots, v^N | v^n)$  is such that

$$\sum_{u^n, \dots, v^N} \mu(u^n, \dots, v^N | v^n) = 1 \quad \text{and} \quad \mu(u^n, \dots, v^N | v^n) > 0, \quad (4.4.3)$$

where  $n \leq N - 2$ . For every path  $(v^n, u^n, \dots, v^{N-1}, u^{N-1}, v^N)$  we recursively define the following matrices,

$$\begin{aligned} A_n^\mu(u^n, \dots, v^N | v^n) &:= \lambda G_n^\mu(v^n)' v_1^{n+1} G_n^\mu(v^n) + \lambda \alpha F_n^\mu(v^n)' \text{diag}(u^n) F_n^\mu(v^n) \\ &\quad + (I - G_n^\mu(v^n) - \text{diag}(u^n) F_n^\mu(v^n))' \left( \frac{1}{2} \lambda^2 \Delta t \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \right. \\ &\quad \left. + A_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) \right) (I - G_n^\mu(v^n) - \text{diag}(u^n) F_n^\mu(v^n)) \end{aligned}$$

for  $n \leq N - 2$ . The terminal condition reads as  $A_{N-1}^\mu(u^{N-1}, v^N | v^{N-1}) := \lambda v_1^N$  for all  $u^{N-1} \in \{0, 1\}^M$  and  $v^{N-1}, v^N \in W$ . As a function of the probability distribution

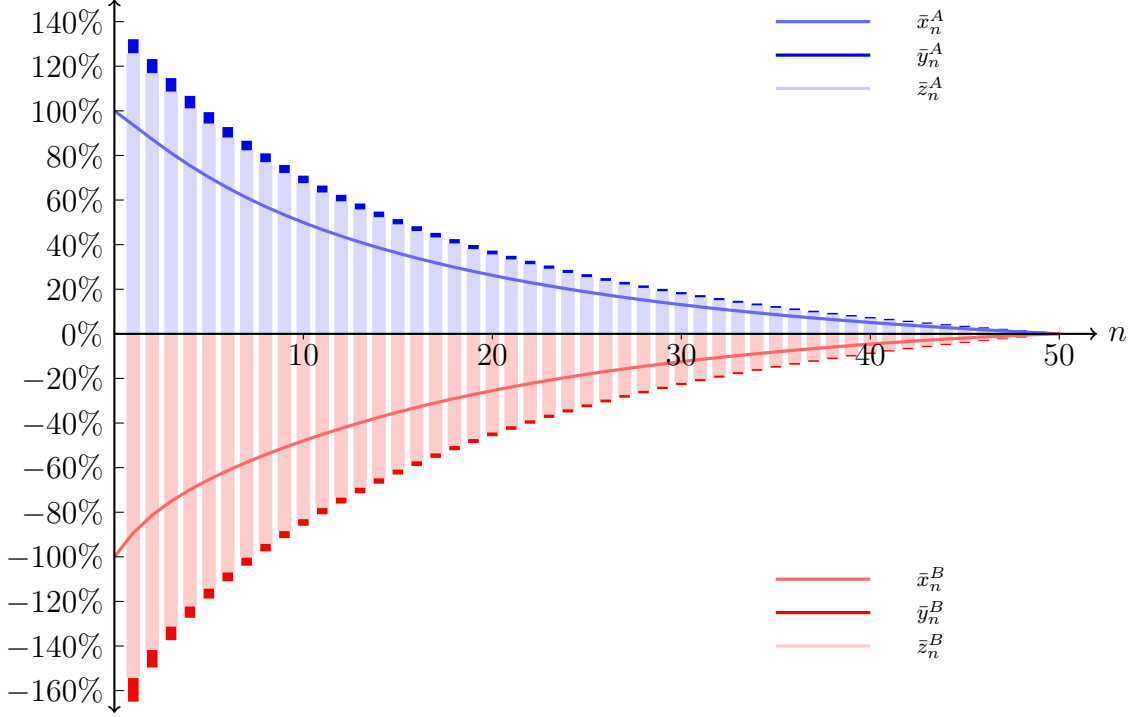


Figure 4.6: Detailed analysis of the negatively correlated pair Company A and B shows the net average stock positions, and average of order sizes submitted into the regular exchange and the dark pool as percentages of initial positions in the assets for the second order Taylor approximation of the expected exponential cost case with  $\alpha = 0$  and  $\lambda = 10^{-4}$ .

$\mu$  as defined in (4.4.3),  $G_n^\mu(v^n)$  and  $F_n^\mu(v^n)$  are defined as

$$\begin{aligned}
 F_n^\mu(v^n) &:= S_n^\mu(v^n)^{-1} B_n^\mu(v^n)' \left( I - U_n^\mu(v^n)^{-1} \left( \sum_{u^n \dots v^N} \mu(u^n, \dots, v^N | v^n) \right. \right. \\
 &\quad \left. \left. \left( \tilde{A}_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) + \lambda^2 \Delta t \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \right) \right) \right), \\
 G_n^\mu(v^n) &:= U_n^\mu(v^n)^{-1} \left( \sum_{u^n \dots v^N} \mu(u^n, \dots, v^N | v^n) \left( \tilde{A}_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) \right. \right. \\
 &\quad \left. \left. + \lambda^2 \Delta t \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \right) - B_n^\mu(v^n) F_n^\mu(v^n) \right),
 \end{aligned}$$



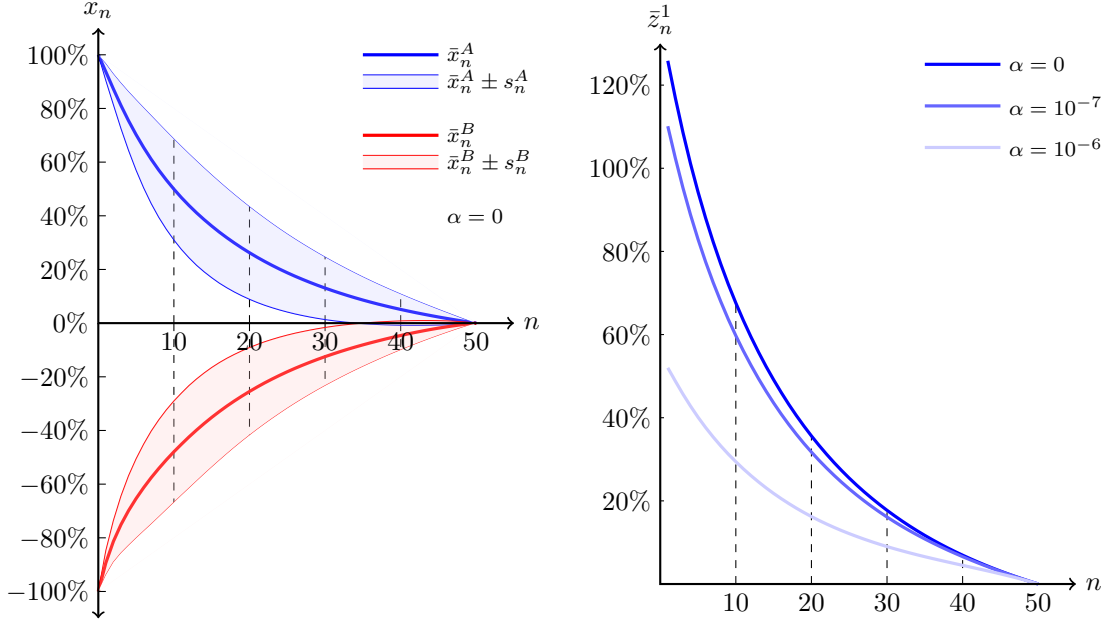


Figure 4.7: Standard deviation from the average execution trajectories and average size of orders submitted to the dark pool for Company A as a function of the parameter  $\alpha$  in the second order Taylor approximation case with  $\lambda = 10^{-4}$ .

where  $S_n^\mu(v^n)$  is the Schur complement in  $U_n^\mu(v^n)$  of the matrix

$$\begin{bmatrix} U_n^\mu(v^n) & B_n^\mu(v^n) \\ B_n^\mu(v^n)' & L_n^\mu(v^n) \end{bmatrix}$$

with

$$U_n^\mu(v^n) := \sum_{u^n \dots v^N} \mu(u^n, \dots, v^N | v^n) \left( \tilde{A}_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) + \lambda \tilde{v}_1^{n+1} \lambda^2 \Delta t \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \right),$$

$$B_n^\mu(v^n) := \sum_{u^n \dots v^N} \mu(u^n, \dots, v^N | v^n) \left( \tilde{A}_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) + \lambda^2 \Delta t \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \right) \text{diag}(u^n) \quad \text{and,}$$

$$L_n^\mu(v^n) := \sum_{u^n \dots v^N} \mu(u^n, \dots, v^N | v^n) \left( 2\lambda\alpha \text{diag}(u^n) + \text{diag}(u^n) \tilde{A}_{n+1}^\mu(u^{n+1}, \dots, v^N | v^{n+1}) \text{diag}(u^n) + \lambda^2 \Delta t \text{diag}(u^n) \text{diag}(v_2^{n+1}) \Sigma \text{diag}(v_2^{n+1}) \text{diag}(u^n) \right)$$

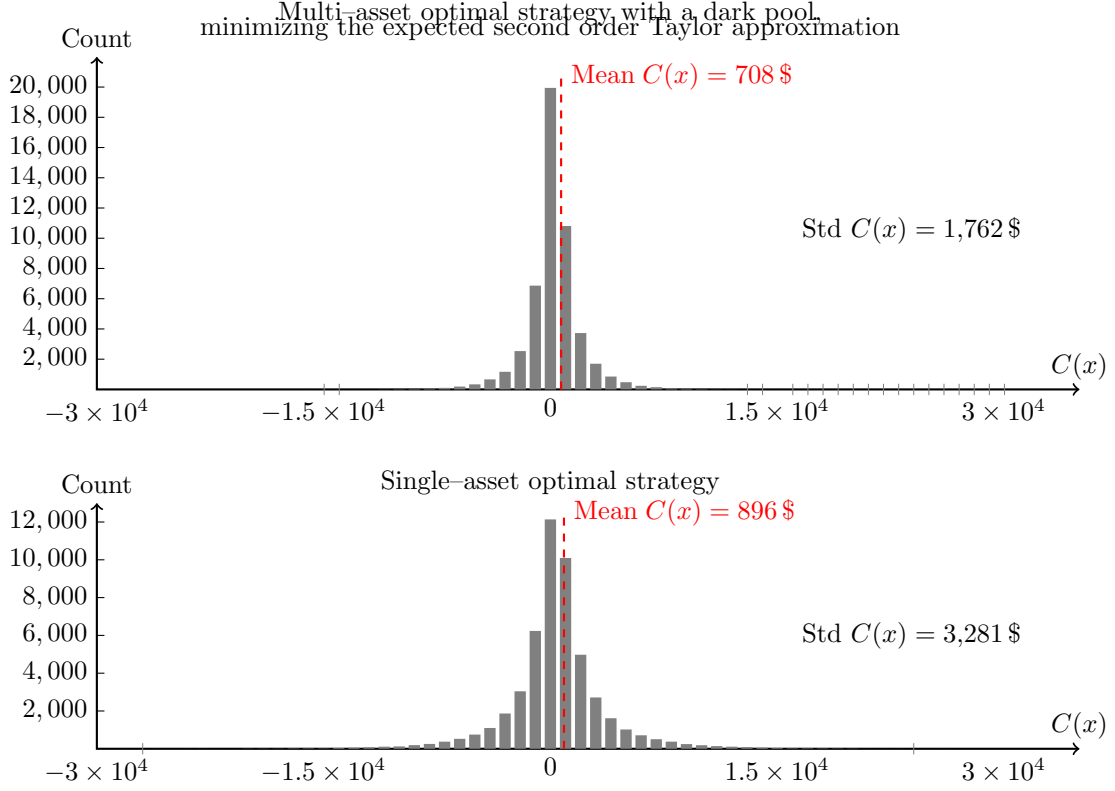


Figure 4.8: First histogram of  $C(x)$  is computed for the strategy minimizing the expectation of the second order exponential of the cost in the multi-asset case with a dark pool where  $\lambda = 10^{-4}$  and  $\alpha = 0$ . Second histogram of  $C(x)$  is computed by separately finding two individually optimal strategies minimizing the expected exponential of the cost in the single-asset case without a dark pool and where  $\lambda = 10^{-4}$ .

for  $n \leq N - 2$ . The following theorem characterizes the optimal strategy that solves the dynamic programming equation of Proposition 4.4.1.

**Theorem 4.4.3.** *For all  $n \leq N - 2$  and  $v^n \in W$ , there exist  $\mu^*(u^n, \dots, v^N | v^n)$  such that*

$$\sum_{u^n, \dots, v^N} \mu^*(u^n, \dots, v^N | v^n) = 1 \quad \text{and} \quad \mu^*(u^n, \dots, v^N | v^n) > 0,$$

and the strategy  $(y_n^*, z_n^*)$  given by

$$y_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1}) = v^{n-1}} = G_{n-1}^{\mu^*}(v^{n-1})x_{n-1}^* \quad \text{and}$$

$$z_n^* | x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1}) = v^{n-1} = F_{n-1}^{\mu^*}(v^{n-1})x_{n-1}^*,$$

solves the Bellman equations of Proposition 4.4.1.

Now one can generate approximate strategies for the expected exponential of the portfolio execution cost by using Theorem 4.4.3. One simply should specify probability distributions on the paths of the Markov chain and then use these distributions to recursively generate matrices  $G_{n-1}^\mu(v^{n-1})$  and  $F_{n-1}^\mu(v^{n-1})$ . We provide two examples.

**Example 4.4.4.** For all  $n \leq N - 1$  and  $v^n \in W$  choose

$$\mu(u^n, \dots, v^N | v^n) = \frac{1}{\#(W)^{N-n} \times 2^{M(N-n)}},$$

where  $\#(W)$  denotes the cardinality of the set  $W$  and  $2^M = \#\{0, 1\}^M$ . Define the following matrices,

$$\bar{\Lambda} = \frac{1}{\#(W)} \sum_{v \in W} v_1, \bar{\Sigma} = \frac{1}{\#(W)} \sum_{v \in W} \text{diag}(v_2) \Sigma \text{diag}(v_2), b\bar{\Sigma}b = \frac{1}{2^M} \sum_{u \in \{0,1\}^M} \text{diag}(u) \bar{\Sigma} \text{diag}(u),$$

set  $A_{N-1}^{\bar{\mu}} = \lambda \bar{\Lambda}$  and for  $n \leq N - 2$  recursively define

$$\begin{aligned} A_n^{\bar{\mu}} &= \lambda G_n^{\bar{\mu}'} \bar{\Lambda} G_n^{\bar{\mu}} + \frac{1}{2} \lambda \alpha F_n^{\bar{\mu}'} F_n^{\bar{\mu}} \\ &\quad + \frac{1}{2^M} \sum_{u \in \{0,1\}^M} (I - G_n^{\bar{\mu}} - \text{diag}(u) F_n^{\bar{\mu}})' \left( \frac{1}{2} \lambda^2 \Delta t \bar{\Sigma} + A_{n+1}^{\bar{\mu}} \right) (I - G_n^{\bar{\mu}} - \text{diag}(u) F_n^{\bar{\mu}}), \\ F_n^{\bar{\mu}} &:= (S_n^{\bar{\mu}})^{-1} B_n^{\bar{\mu}'} \left( I - (U_n^{\bar{\mu}})^{-1} (\tilde{A}_{n+1}^{\bar{\mu}} + \lambda^2 \Delta t \bar{\Sigma}) \right), G_n^{\bar{\mu}} := (U_n^{\bar{\mu}})^{-1} \left( \tilde{A}_{n+1}^{\bar{\mu}} + \lambda^2 \Delta t \bar{\Sigma} - B_n^{\bar{\mu}} F_n^{\bar{\mu}} \right), \\ U_n^{\bar{\mu}} &:= \tilde{A}_{n+1}^{\bar{\mu}} + \lambda \tilde{\Lambda} + \lambda^2 \Delta t \bar{\Sigma}, B_n^{\bar{\mu}} := \tilde{A}_{n+1}^{\bar{\mu}} + \frac{1}{2} \lambda^2 \Delta t \bar{\Sigma}, L_n^{\bar{\mu}} := \lambda \alpha I + b \tilde{A}_{n+1}^{\bar{\mu}} b + \lambda^2 \Delta t b \bar{\Sigma} b, \end{aligned}$$

and  $b \tilde{A}_{n+1}^{\bar{\mu}} b := \frac{1}{2^M} \sum_{u \in \{0,1\}^M} \text{diag}(u) \tilde{A}_{n+1}^{\bar{\mu}} \text{diag}(u)$ , for a uniform distribution  $\bar{\mu}$  on execution paths. It follows that for such a choice of  $\mu$ , the approximate (and deter-

ministic) strategy is given by

$$y_n |_{x_{n-1}, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v^{n-1}} = G_{n-1}^{\bar{\mu}} x_{n-1}, \quad z_n |_{x_{n-1}, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v^{n-1}} = F_{n-1}^{\bar{\mu}} x_{n-1}$$

where  $n \leq N - 1$ .

**Example 4.4.5.** Pick  $\mu(u^n, \dots, v^N | v^n) = p(u^n, \dots, v^N | v^n)$ , for all  $n \leq N - 1$  and  $v^n \in W$ . For this particular choice of  $\mu$  the strategy given in Theorem 4.4.2 is optimal.

Note that one should make reasonable choices for  $\mu$ . For instance, in the case where the actual probabilities on the paths of the chain and the uniform distribution are far from each other (such as in our case), Example 4.4.4 will not produce good results. The dual of the problems given in (4.4.1) is a type of entropy maximization problem where the objective function consists of the sum of relative entropy between  $\mu$  and  $p$  and a term that represents  $y$  and  $z$  in terms of  $\mu$  and  $x$ . The choice of  $\mu$  as in Example 4.4.5 maximizes the entropy term but will not be optimal for the other term in the objective function. The user of the program may develop heuristic approaches to generate approximate strategies that perform better by making these observations. One may also use a duality gap argument to estimate the error of an approximation.

## 4.5 Conclusion

We have addressed an optimal portfolio execution problem with a dark pool by using a framework that assumes stochastic volatility, price impact and dark pool liquidity. We found a predictable strategy that minimizes the risk neutral objective and provided a solution minimizing the expectation of the second order Taylor expansion of the exponential cost. Moreover, we have developed a framework via duality that can be used to treat the exponential case by generating approximate strategies. Although we

have made independence assumptions for the Markov processes we have in our setup, our formulation can be used to account for codependence between volatility, market impact, and dark pool liquidity by an appropriate choice of transition probabilities. Moreover, one can capture the time varying dynamics of the markets by specifying non-stationary transition probabilities. In practice, the adverse selection parameter  $\alpha$  can be used to adjust the size of orders submitted to the dark pool to levels that are deemed appropriate for the execution program to be used. In practice one executes orders for a portfolio of S&P500 stocks, and dimensionality can be a big problem for the computation of optimal strategy. Therefore, the setup we proposed in Section 4.2 for the numerical studies, which makes use of common factors to describe the volatility and market impact processes can be used to reduce the dimensionality of the problem.

# Chapter 5

## Robust execution of a single–asset

In practice the user of any model must find the corresponding parameter set which fits the model that has been specified. In the case of stochastic dynamics one also assumes a distribution for these parameters and believes that this distribution reflects the reality for the probability of outcomes. One issue with this approach is that it does not account for the possibility of model misspecification. When the distribution is known, risk can be calculated. On the other hand, when there exists ambiguity regarding the distribution, one will not be able to calculate the risks associated with the decision making process with certainty. This concept was first introduced by Knight (1921) and then has been formalized by Gilboa and Schmeidler (1989). Hansen and Sargent (2001) studied a control problem under model uncertainty by using multiplier preferences. Klibanoff et al. (2005) proposed a model for decision making by introducing second order certainty equivalence. Subsequently, variational preferences have been introduced by Maccheroni et al. (2006a) and have been extended to time-consistent dynamic preferences by Maccheroni et al. (2006b). Representation of preferences is also closely related to risk measures. Coherent risk measures have been extended to a dynamic framework by Riedel (2004). Cheridito and Kupper (2011) have shown how one can see time-consistent risk measures as compositions of one-step risk measures.

The purpose of this chapter is to compute robust execution strategies in an Almgren–Chriss framework by making use of time-consistent risk measures, first order multi-prior certainty equivalence and second order certainty equivalence frameworks. We use the program developed in Chapter 2. Here we assume that there exists ambiguity for the transition probabilities of this Markov chain and we consider three setups that allow to account for both ambiguity and risk aversion. We see the total cost of execution as sum of costs that arise at each period and call these costs periodic costs. The first setup we consider aggregates worst case expectations of periodic costs by penalizing the transition probabilities with dynamic ambiguity indices. The second setup we consider formulates the problem as the first order certainty equivalent of expected execution cost with multiple priors. The last setup uses a second order certainty equivalence formulation. We find predictable strategies that minimize the cost functionals of these setups. Our work is close to Schied (2013) who find a robust strategy minimizing the execution cost by assuming uncertainty only for the distribution of the unaffected stock price process in an Almgren–Chriss setup. However, our approach differs because we assume ambiguity for the volatility and liquidity and think that the law of market innovations are known with certainty.

The paper is organized such that in Section 5.1 we introduce the dynamics of the market model with linear transient impact. In Section 5.2 we study the case when the objective functional is assembled by aggregating periodic risks with dynamic ambiguity indices of indicator, relative entropy and Gini type. Section 5.3 studies the first order certainty equivalence formulation with multi-priors and Section 5.4 the second order certainty equivalence case. All proofs are given in the Appendix A.4.

## 5.1 The model

We consider the same problem of liquidating  $X \in \mathbb{R}_+$  shares as in Chapter 2. After dropping the constant  $cX^2/2$  from the cost functional  $C(x)$  we express it as  $V(x) = \sum_{n=1}^N V_n(x)$ , where  $V_n(x)$  are the costs incurred in the time interval  $(t_{n-1}, t_n]$  that read as

$$V_n(x) := (x_{n-1} - x_n)^2 \tilde{\eta}_n - x_n \sigma_n \sqrt{\Delta t} \xi_n. \quad (5.1.1)$$

The filtration  $(\mathcal{F}_n)$  and the definition of the set of admissible strategies  $\mathcal{A}$  is the same as in Chapter 2. We again assume that  $\xi_n$  is independent of  $\sigma(\mathcal{F}_{n-1}, \sigma_n, \eta_n)$ .  $(\sigma_n, \eta_n)$  is assumed to take finitely many and different values in  $\mathbb{R}_+^2$ , moreover, we specify  $(\sigma_n, \eta_n)$  as a Markov chain with finite state space  $W \subseteq \mathbb{R}_+^2$ . Therefore, when conditioned on  $\mathcal{F}_{n-1}$ , its distribution only depends on  $(\sigma_{n-1}, \eta_{n-1})$ . However, we assume there exists ambiguity for the transition probabilities of  $(\sigma_n, \eta_n)$ . We define time-dependent reference transition probabilities as

$$p_{n-1}^{vw} := \mathbb{P}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v], \quad v, w \in W. \quad (5.1.2)$$

We assume  $p_{n-1}^{vw} > 0$  for all  $n \leq N-1$  and  $v, w \in W$ . Reference transition probability for  $(\sigma_n, \eta_n)$  given  $(\sigma_{n-1}, \eta_{n-1}) = v$  is denoted by  $\mathbb{P}_{n-1}^v := (p_{n-1}^{vw})_{w \in W}$ . We denote by  $\mathbb{E}_{\mathbb{P}}$  the expectation taken by using the time-dependent reference transition matrix defined in (5.1.2). Let  $\mathbb{Q}$  denote the set of all discrete probability distributions on the state space  $W$ .  $Q \in \mathbb{Q}$  is then given by  $(q_w)_{w \in W}$  such that  $\sum_{w \in W} q_w = 1$  and  $q_w \geq 0$  for all  $w \in W$ . For a  $Q \in \mathbb{Q}$ ,  $\mathbb{E}_Q$  and  $\mathbb{E}_Q^{n,v}$  respectively denote the expectation under  $Q$  and the conditional expectation  $\mathbb{E}_Q[\cdot \mid (\sigma_n, \eta_n) = v]$ . Notice that for  $Q \in \mathbb{Q}$  we only specify a distribution on the state space of Markov chain  $(\sigma_n, \eta_n)$ ; therefore, regardless of the  $Q$  specified,  $(\xi_n)$  has a standard normal distribution. It is enough to consider the process  $(\sigma_n, \eta_n)$ ,  $n \geq 0$  to derive the optimal strategies. Moreover, we define  $\tilde{\eta}_n := \eta_n - c/2$  and assume that  $w_2 - c/2 > 0$  for all  $w \in W$ .



## 5.2 Aggregation with dynamic ambiguity indices

For  $v \in W$  and  $n \leq N - 1$  we define a dynamic ambiguity index  $\alpha_n^v : \mathbb{Q} \mapsto [0, \infty]$  to penalize the distribution one uses on  $W$  when aggregating expected costs associated with  $V(x)$ . We define a risk map as follows

$$\rho_n(Z) := \sup_{Q \in \mathbb{Q}} \mathbb{E}_Q [Z - \alpha_n^{(\sigma_n, \eta_n)}(Q) \mid \mathcal{F}_n], \quad (5.2.1)$$

and

$$\rho_n^v(Z) := \sup_{Q \in \mathbb{Q}} \mathbb{E}_Q^{n,v} [Z] - \alpha_n^v(Q), \quad (5.2.2)$$

for an  $\mathcal{F}_{n+1}$ -measurable random variable  $Z$  and  $v \in W$ . Let  $u : \mathbb{R} \mapsto \mathbb{R}$  be a convex function and define the aggregation,

$$\rho_0^v \circ \rho_1 \cdots \rho_{N-1} \left( \sum_{n=1}^N \delta^{n-1} u(V_n(x)) \right) \quad (5.2.3)$$

for  $v \in W$ , and a discounting parameter  $\delta \in (0, 1]$ . Let  $\int u(V_n(x))$  denote  $\mathbb{E}[u(V_n(x)) \mid \mathcal{F}_{n-1}, \sigma_n, \eta_n]$ , that is we integrate only with respect to innovation at period  $n$ . Also define

$$V_n^w(x) := (x_{n-1} - x_n)^2(w_2 - c/2) - x_n w_1 \sqrt{\Delta t} \xi_n,$$

for  $w \in W$ . In this section our goal is to find an  $x \in \mathcal{A}$  that minimizes the aggregation of equation (5.2.3) by assuming different ambiguity indices.

### 5.2.1 Indicator index

For all  $n \leq N - 1$ , and  $v \in W$  we fix a set of distributions  $\{P_{n,1}^v, \dots, P_{n,M_n}^v\}$  on  $W$ , where  $M_n \in \mathbb{N}$ . By  $\mathcal{P}_n^v$  we denote the convex hull of  $\{P_{n,1}^v, \dots, P_{n,M_n}^v\}$ . First we consider the problem obtained by assuming an ambiguity index of indicator function

type:

$$\alpha_n^v(Q) = \begin{cases} \infty, & \text{if } Q \notin \mathcal{P}_n^v \\ 0, & \text{if } Q \in \mathcal{P}_n^v. \end{cases} \quad (5.2.4)$$

Define the value function

$$J_n^v(z) := \min_{x \in A_n(z)} \rho_n^v \circ \rho_{n+1} \cdots \rho_{N-1} \left( \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right). \quad (5.2.5)$$

The following theorem characterizes the optimal strategy  $x^* \in \mathcal{A}$  that minimizes (5.2.3) with an ambiguity index of indicator function type.

**Theorem 5.2.1.** *The value function  $J$  satisfies the Bellman equation,*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \sup_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1, M_{N-1}}^v\}} \sum_{w \in W} q_w \delta^{N-1} \int u(V_N^w(x)) \Big|_{x_N=0} \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \sum_{w \in W} q_w \left( \delta^{n-1} \int u(V_n^w(x)) + J_n^v(x_n) \right) \end{aligned}$$

for  $n \leq N-1$ , and the minimizing  $(x_n^*)$  form the unique optimal strategy.

In the case when  $u$  is not linear, the risks associated with the volatility of the asset price will be accounted for in the optimal strategy found by using Theorem 5.2.1. The optimal strategy can be found by discretizing the space of controls if necessary. One special case is when  $u$  is assumed to be linear. This case would be of interest to construct a robust strategy when there is no aversion for risks arising from volatility. Next corollary gives the optimal strategy for a linear function  $u$ :

**Corollary 5.2.2.** *When one assumes  $u$  is linear, the unique optimal strategy of Theorem 5.2.1 is given by*

$$x_n^* |_{x_{n-1}^*, (\sigma_{n-1}, \eta_{n-1})=v} = x_{n-1}^* \frac{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1, v}[\tilde{\eta}_n]}{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1, v}[\tilde{\eta}_n] + \sum_{w \in W} q_w^* a_n^w}, \quad n = 1, \dots, N-1, \quad (5.2.6)$$

where the coefficients  $a_n^w$  satisfy the backwards recursion:

$$a_{N-1}^v = \delta^{N-1} \mathbb{E}_{Q^*}^{N-1,v}[\tilde{\eta}_N], \quad a_{n-1}^v = \frac{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n] \sum_{w \in W} q_w^* a_n^w}{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n] + \sum_{w \in W} q_w^* a_n^w}, \quad n \leq N-1. \quad (5.2.7)$$

and the optimal choice of transition probabilities  $Q^* = (q_w^*)_{w \in W}$  satisfy

$$Q^* |_{(\sigma_{N-1}, \eta_{N-1})=v} = \arg \max_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1, M_{N-1}}^v\}} \delta^{N-1} \mathbb{E}_Q^{N-1,v}[\tilde{\eta}_N]. \quad (5.2.8)$$

and for  $n \leq N-1$ ,

$$Q^* |_{(\sigma_{n-1}, \eta_{n-1})=v} = \arg \max_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \frac{\delta^{n-1} \mathbb{E}_Q^{n-1,v}[\tilde{\eta}_n] \sum_{w \in W} q_w a_n^w}{\delta^{n-1} \mathbb{E}_Q^{n-1,v}[\tilde{\eta}_n] + \sum_{w \in W} q_w a_n^w}. \quad (5.2.9)$$

(5.2.6)–(5.2.9) define a system of forward–backward stochastic equations. The optimal strategy is to reduce the position in the asset to a fraction of the previous position. This fraction depends on the ratio of the next periods discounted expected liquidity under the worst case distribution to a weighted path of discounted worst case expected liquidity terms. Note that the strategy of Corollary 5.2.2 corresponds to the predictable risk neutral strategy given by Theorem 2.2.1 when the set  $\mathcal{P}_n^v$  is a singleton that consists of  $\mathbb{P}_n^v$  and  $\delta = 1$ . In the special case of having constant liquidity along with  $\delta = 1$ , the optimal solution is the constant speed sell strategy found in Bertsimas and Lo (1998).

## 5.2.2 Relative entropy index

For a parameter  $\theta \in \mathbb{R}_+$ , we define the next dynamic ambiguity index we use as the relative entropy:

$$\alpha_n^v(Q) := \mathbb{E}_Q \left[ \frac{1}{\theta} \log \left( \frac{dQ}{d\mathbb{P}_n^v} \right) \right] = \sum_{w \in W} q_w \frac{1}{\theta} \log \left( \frac{q_w}{p_n^{vw}} \right), \quad (5.2.10)$$

where  $Q \in \mathbb{Q}$ . Our goal is to find a strategy that minimizes (5.2.3), where this time the ambiguity index is the relative entropy. Define the value function

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \exp \left( \theta \rho_n^v \circ \rho_{n+1} \cdots \rho_{N-1} \left( \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right) \right).$$

It is well known that the relative entropy ambiguity index assembles the exponential utility. The following theorem gives the Bellman equation which characterizes the optimal strategy for this case:

**Theorem 5.2.3.** *The value function  $J$  satisfies the Bellman equation,*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \sum_{w \in W} p_{N-1}^{vw} \exp \left( \delta^{N-1} \theta \int u(V_N^w(x)) \right) \Big|_{x_N=0} \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in W} p_{n-1}^{vw} \exp \left( \delta^{n-1} \theta \int u(V_n^w(x)) \right) J_n^w(x_n) \end{aligned}$$

for  $n \leq N - 1$ , and the minimizing  $(x_n^*)$  form the unique optimal strategy.

The Bellman equation of Theorem 5.2.3 is solved by discretizing the space of controls. When one assumes that  $u$  is linear, the optimal solution will not account for volatility risk. In such a case one obtains a robust strategy minimizing only the expected exponential of liquidity costs. Following corollary describes such a strategy:

**Corollary 5.2.4.** *When one assumes  $u$  is linear the value function  $J$  in Theorem 5.2.3 satisfies the Bellman equation*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \sum_{w \in W} p_{N-1}^{vw} \exp \left( \theta \delta^{N-1} x_{N-1}^2 (w_2 - c/2) \right) \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in W} p_{n-1}^{vw} \exp \left( \theta \delta^{n-1} (x_{n-1} - x_n)^2 (w_2 - c/2) \right) J_n^w(x_n), \end{aligned}$$

for  $n \leq N - 1$  and the minimizing  $(x_n^*)$  form the unique optimal strategy.

### 5.2.3 Gini index

The next ambiguity index we consider belongs to the family of indices given by:

$$\alpha_n^v(Q) := \mathbb{E}_{\mathbb{P}_n^v} \left[ \phi \left( \frac{dQ}{d\mathbb{P}_n^v} \right) \right] = \sum_{w \in W} p_n^{vw} \phi \left( \frac{q_w}{p_n^{vw}} \right), \quad (5.2.11)$$

where  $Q \in \mathbb{Q}$  and  $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ , is a strictly convex differentiable function that grows super-linearly with  $\phi'(1) = \phi(1) = 0$ . This index is known as the divergence index. Proposition A.4.3 in the appendix shows the necessary and sufficient conditions for finding the optimal choice of transition probabilities which can be used to solve (5.2.3) when  $\phi$  is used to describe the ambiguity index. We only consider the Gini index here. It is given by  $\phi(x) = \lambda(x-1)^2/2$  for a parameter  $\lambda \in \mathbb{R}_+$ . Define the value function

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \rho_n^v \circ \dots \circ \rho_{N-1} \left( \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right).$$

**Theorem 5.2.5.** *For a sufficiently large parameter  $\lambda \in \mathbb{R}_+$ , the value function  $J$  satisfies the Bellman equation,*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \delta^{N-1} \mathbb{E}_{\mathbb{P}}^{N-1,v} \left[ \int u(V_N(x)) \right] + \frac{1}{2\lambda} \delta^{2(N-1)} \text{Var}_{\mathbb{P}}^{N-1,v} \left( \int u(V_N(x)) \right) \Big|_{x_N=0} \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \delta^{n-1} \mathbb{E}_{\mathbb{P}}^{n-1,v} \left[ \int u(V_n(x)) \right] + \frac{1}{2\lambda} \delta^{2(n-1)} \text{Var}_{\mathbb{P}}^{n-1,v} \left( \int u(V_n(x)) \right) \\ &\quad + \mathbb{E}_{\mathbb{P}}^{n-1,v} [J_n(x_n)] + \frac{1}{2\lambda} \text{Var}_{\mathbb{P}}^{n-1,v} (J_n(x_n)) + \frac{1}{\lambda} \delta^{n-1} \text{Cov}_{\mathbb{P}}^{n-1,v} \left( J_n(x_n), \int u(V_n(x)) \right) \end{aligned}$$

for  $n \leq N-1$ , and the minimizing  $(x_n^*)$  form the optimal strategy.

We solve the Bellman equation of Theorem 5.2.5 by discretizing the space of controls when the value function can not be obtained in closed form. In the special case of linear  $u$  the best robust strategy can be obtained by using the following corollary, where given  $(\sigma_n, \eta_n) = v$ ,  $\text{Var}_{\mathbb{P}}^{n,v}$ ,  $\text{Cov}_{\mathbb{P}}^{n,v}$  denote the conditional variance and covariance computed using  $\mathbb{P}$ .

**Corollary 5.2.6.** *For a sufficiently large parameter  $\lambda \in \mathbb{R}_+$  and linear  $u$ , the value function  $J$  satisfies the Bellman equation,*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= x_{N-1}^2 \mathbb{E}_{\mathbb{P}}^{N-1,v}[\tilde{\eta}_N] + \frac{1}{2\lambda} x_{N-1}^4 \text{Var}_{\mathbb{P}}^{N-1,v}(\tilde{\eta}_N) \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} (x_{n-1} - x_n)^2 \mathbb{E}_{\mathbb{P}}^{n-1,v}[\tilde{\eta}_n] + \frac{1}{2\lambda} (x_{n-1} - x_n)^4 \text{Var}_{\mathbb{P}}^{n-1,v}(\tilde{\eta}_n) \\ &\quad + \mathbb{E}_{\mathbb{P}}^{n-1,v}[J_n(x_n)] + \frac{1}{2\lambda} \text{Var}_{\mathbb{P}}^{n-1,v}(J_n(x_n)) + \frac{1}{\lambda} (x_{n-1} - x_n)^2 \text{Cov}_{\mathbb{P}}^{n-1,v}(J_n(x_n), \tilde{\eta}_n) \end{aligned}$$

for  $n \leq N - 1$  and the minimizing  $(x_n^*)$  form the optimal strategy.

### 5.3 First order certainty equivalent

In this section our goal is to find the strategy minimizing,

$$\mu_0^v \circ \mu_1 \circ \dots \circ \mu_{N-1}(V(x)) \tag{5.3.1}$$

over  $x \in \mathcal{A}$  for  $v \in W$  where,

$$\mu_n(Z) = \frac{1}{\theta} \log \left( \sup_{Q \in \mathcal{P}_n^{(\sigma_n, \eta_n, \varphi_n)}} \mathbb{E}_Q[\exp(\theta Z) \mid \mathcal{F}_n] \right) \tag{5.3.2}$$

and

$$\mu_n^v(Z) := \frac{1}{\theta} \log \left( \sup_{Q \in \mathcal{P}_n^v} \mathbb{E}_Q^{n,v}[\exp(\theta Z)] \right)$$

for a parameter  $\theta \in \mathbb{R}_+$  and an  $\mathcal{F}_{n+1}$ -measurable random variable  $Z$ . The sets  $\{P_{n,1}^v, \dots, P_{n,M_n}^v\}$  and  $\mathcal{P}_n^v$  are defined as in Section 5.2.1. The reason we only consider the exponential function to specify the composition in (5.3.1) is because only exponential and linear functions constitute a time-consistent formulation for a first order certainty equivalence setup (i.e. see Cheridito and Kupper (2006)). Define the

value function

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \exp \left( \theta \mu_n^v \circ \mu_{n+1} \dots \mu_{N-1} \left( \sum_{i=n+1}^N V_i(x) \right) \right).$$

The following theorem characterizes the optimal solution  $x^* \in \mathcal{A}$  that minimizes (5.3.1).

**Theorem 5.3.1.** *The value function  $J$  satisfies the Bellman equation,*

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \sup_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1,M_{N-1}}^v\}} \sum_{w \in W} q_w \exp(\theta x_{N-1}^2 (w_2 - c/2)) \\ J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \{P_{N-1}^1, \dots, P_{N-1}^{M_{N-1}}\}} \\ &\quad \sum_{w \in W} q_w \exp \left( \theta (x_{n-1} - x_n)^2 (w_2 - c/2) + \frac{1}{2} \theta^2 x_n^2 w_1^2 \Delta t \right) J_n^w(x_n) \end{aligned}$$

for  $n \leq N - 1$ , and the minimizing  $(x_n^*)$  form the unique optimal strategy.

Notice that when one assumes that  $\mathcal{P}_n^v$  is a singleton given by  $(p_n^{vw})_{w \in W}$  one will obtain the optimal strategy minimizing the expected exponential cost given by Theorem 2.3.1. Hence the minimizing strategy is also expected to be closely related to the predictable strategy minimizing a mean–variance criterion with parameter  $\theta/2$  and in the case of constant liquidity and volatility one obtains the mean–variance minimizing deterministic strategy of Almgren and Chriss (2001).

## 5.4 Second order certainty equivalent

In this section we assume  $u : \mathbb{R} \mapsto \mathbb{R}$  is an increasing convex function, and introduce a continuous and increasing function  $\nu : \mathbb{R} \mapsto \mathbb{R}$ . For each  $n \leq N - 1$  and  $v \in W$  we fix a probability distribution  $(\pi_{n,i}^v)_{i=1}^{M_n}$  on the set of distributions  $\{P_{n,1}^v, \dots, P_{n,M_n}^v\}$  on

$W$ . Our goal is to find  $x^* \in \mathcal{A}$  that gives  $J_0^v(X)$  for  $v \in W$  where

$$J_{N-1}^v(x_{N-1}) = \min_{x \in \mathcal{A}_{N-1}(x_{N-1})} \nu^{-1} \left( \sum_{i=1}^{M_{N-1}} \nu \circ u^{-1} \left( \mathbb{E}_{P_{N-1,i}^v}^{N-1,v} [u(V_N(x))] \right) \pi_{N-1,i}^v \right)$$

$$J_{n-1}^v(x_{n-1}) = \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \nu^{-1} \left( \sum_{i=1}^{M_{n-1}} \nu \circ u^{-1} \left( \mathbb{E}_{P_{n-1,i}^v}^{n-1,v} [u(V_n(x)) + \delta u(J_n(x_n))] \right) \pi_{n-1,i}^v \right)$$

for  $n \leq N-1$  and a parameter  $\delta \in (0, 1]$ . The following proposition gives the optimal solution to this problem.

**Proposition 5.4.1.** *The optimal strategy  $(x_n^*)$  satisfying the following Bellman equations*

$$\tilde{J}_{N-1}^v(x_{N-1}) = \sum_{i=1}^{M_{N-1}} \nu \circ u^{-1} \left( \mathbb{E}_{P_{N-1,i}^v}^{N-1,v} [u(V_N(x))] \right) \pi_{N-1,i}^v \Big|_{x_N=0}$$

$$\tilde{J}_{n-1}^v(x_{n-1}) = \min_{0 \leq x_n \leq x_{n-1}} \sum_{i=1}^{M_{n-1}} \nu \circ u^{-1} \left( \mathbb{E}_{P_{n-1,i}^v}^{n-1,v} [u(V_n(x)) + \delta u \circ \nu^{-1}(\tilde{J}_n(x_n))] \right) \pi_{n-1,i}^v$$

for  $n \leq N-1$ , is the optimal solution for the second order certainty equivalence objective.

As a special case assume  $\nu = u$ . Then the optimal strategy obtained from Proposition 5.4.1 also minimizes (5.2.3) with indicator indices given by,

$$\alpha_n^v(Q) = \begin{cases} \infty, & \text{if } Q \neq \sum_{i=1}^{M_n} P_{n,i}^v \pi_{n,i}^v \\ 0, & \text{if } Q = \sum_{i=1}^{M_n} P_{n,i}^v \pi_{n,i}^v. \end{cases}$$

Note that in the case  $\nu = u$ , aversion towards uncertainty is not penalized. Moreover if  $u$  is linear,  $\delta = 1$  and  $\sum_{i=1}^{M_n} P_{n,i}^v \pi_{n,i}^v = \mathbb{P}_n^v$  for all  $n \leq N-1$  then the optimal strategy we obtain is that of the risk neutral objective given by Theorem 2.2.1.



# Appendix A

## A.1 Proofs for Chapter 2

### Proof of Theorem 2.2.1

We prove the theorem by backwards induction. Since  $\mathbb{E}_0^v[Q(x)] = \mathbb{E}_0^v[R(x)]$  for

$$R_n(x) = \sum_{i=n+1}^N (x_{i-1} - x_i)^2 \tilde{\eta}_i,$$

we denote by  $\mathcal{A}_n(z)$  the set of predictable strategies  $(x_i)_{i=n}^N$  such that  $x_n = z$ ,  $x_{i-1} \geq x_i$ ,  $x_N = 0$  and define

$$J_n^v(z) := \min_{x \in \mathcal{A}_n(z)} \mathbb{E}_n^v[R_n(x)].$$

Then

$$J_{N-1}^v(x_{N-1}) = x_{N-1}^2 \mathbb{E}_{N-1}^v[\tilde{\eta}_N] = x_{N-1}^2 a_{N-1}^v,$$

and inductively,

$$\begin{aligned} J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v[R_{n-1}(x)] \\ &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} (x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \mathbb{E}_{n-1}^v[R_n(x_n)] \\ &= \min_{x_n} (x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + x_n^2 \sum_{w \in V^k} p_{n-1}^{vw} a_n^w, \quad n \leq N-1. \end{aligned}$$

It follows that the unique optimal strategy is given by

$$x_n^* = x_{n-1}^* \frac{\mathbb{E}_{n-1}^v[\tilde{\eta}_n]}{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \sum_{w \in V^k} p_{n-1}^{vw} a_n^w},$$

and  $J_{n-1}^v(x_{n-1}^*)$  becomes

$$(x_{n-1}^*)^2 \frac{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] \sum_{w \in V^k} p_{n-1}^{vw} a_n^w}{\mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \sum_{w \in V^k} p_{n-1}^{vw} a_n^w} = (x_{n-1}^*)^2 a_{n-1}^w.$$

In particular,  $J_0^v(X) = X^2 a_0^v$ . □

### Proof of Theorem 2.3.1

Define

$$R_n(x) := \exp \left( \sum_{i=n+1}^N \alpha (x_{i-1} - x_i)^2 \tilde{\eta}_i + \frac{1}{2} \alpha^2 x_i^2 \sigma_i^2 \Delta t \right)$$

and note that  $\mathbb{E}_n^v[\exp(\alpha Q_n(x))] = \mathbb{E}_n^v[R_n(x)]$ . So

$$J_{N-1}^v(x_{N-1}) = \sum_{w \in V} p_{N-1}^{vw} \exp(\alpha x_{N-1}^2 (w_2 - c/2))$$

and

$$\begin{aligned} J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v \left[ \exp \left( \alpha (x_{n-1} - x_n)^2 \tilde{\eta}_n + \frac{1}{2} \alpha^2 x_n^2 \sigma_n^2 \Delta t \right) R_n(x_n) \right] \\ &= \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in V} p_{n-1}^{vw} \exp \left( \alpha (x_{n-1} - x_n)^2 (w_2 - c/2) + \frac{1}{2} \alpha^2 x_n^2 w_1^2 \Delta t \right) J_n^w(x_n), \end{aligned}$$

$n \leq N - 1$ . Since in every step,  $x_n^*$  minimizes a strictly convex function, the optimal strategy  $x^*$  is unique.

**Proposition A.1.1.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $E$  a non-empty subset of  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . Denote*

$$V^\lambda(X) := \mathbb{E}X + \lambda \text{Var}(X), \quad V^{\lambda, \mu}(X) := \mu \mathbb{E}X + \lambda \mathbb{E}X^2, \quad X \in L^2,$$

and assume  $V^\lambda$  attains a minimum over  $E$  at  $X^* \in E$ . Then  $X^*$  minimizes  $V^{\lambda,\mu}$  over  $E$  for  $\mu = 1 - 2\lambda\mathbb{E}X^*$ .

*Proof.* One can write  $V^\lambda(X) = f(\mathbb{E}X, \mathbb{E}X^2)$  for  $f(x) = x_1 - \lambda x_1^2 + \lambda x_2$ . Since  $f$  is concave, one has

$$f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*) = f(x^*) + (1 - 2\lambda x_1^*)(x_1 - x_1^*) + \lambda(x_2 - x_2^*).$$

So if  $X^* \in E$  is a minimizer of  $V^\lambda$  and  $\mu = 1 - 2\lambda\mathbb{E}X^*$ , then

$$V^\lambda(X^*) \leq V^\lambda(X) \leq V^\lambda(X^*) + \mu(\mathbb{E}X - \mathbb{E}X^*) + \lambda(\mathbb{E}X^2 - \mathbb{E}(X^*)^2)$$

for all  $X \in E$ , and therefore,  $V^{\lambda,\mu}(X^*) \leq V^{\lambda,\mu}(X)$ . □

### Proof of Theorem 2.4.1

Define

$$R_n(x) := \sum_{i=n+1}^N \mu(x_{i-1} - x_i)^2 \tilde{\eta}_i + \lambda x_i^2 \sigma_i^2 \Delta t + \lambda(x_{i-1} - x_i)^4 \tilde{\eta}_i^2 + 2\lambda(x_{i-1} - x_i)^2 \tilde{\eta}_i^2 \sum_{j>i} (x_{j-1} - x_j)^2 \tilde{\eta}_j$$

and note that

$$\mathbb{E}_n^v[R_n(x)] = \mathbb{E}_n^v[\mu Q_n(x) + \lambda Q_n^2(x)].$$

So one has

$$J_{N-1}^v(x_{N-1}) = \mu x_{N-1}^2 \mathbb{E}_{N-1}^v[\tilde{\eta}_N] + \lambda x_{N-1}^4 \mathbb{E}_{N-1}^v[\tilde{\eta}_N^2]$$

and

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{R}_{n-1}(x_{n-1})} \mu(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v[\sigma_n^2] \\
&\quad + \lambda(x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v[\tilde{\eta}_n^2] + 2\lambda(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v \left[ \tilde{\eta}_n \sum_{j=n+1}^N (x_{j-1} - x_j)^2 \tilde{\eta}_j \right] \\
&\quad + \mathbb{E}_{n-1}^v[R_n(x)] \\
&= \min_{0 \leq x_n \leq x_{n-1}} \mu(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v[\sigma_n^2] \\
&\quad + \lambda(x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v[\tilde{\eta}_n^2] + 2\lambda(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v \left[ \tilde{\eta}_n \sum_{j=n+1}^N (x_{j-1} - x_j)^2 \tilde{\eta}_j \right] \\
&\quad + \sum_{w \in V} p_{n-1}^{vw} J_n^w(x_n), \quad n \leq N-1.
\end{aligned}$$

So a  $x^* \in \mathcal{R}$  is optimal if it realizes the minimum for every  $n = 1, \dots, N-1$ . By assumption, one has  $\mathbb{E}_{n-1}^v[\tilde{\eta}_n] > 0$ . Therefore if  $\mu > 0$ , then in every step, the function to be minimized is strictly convex. It follows that the optimal strategy  $x^* \in \mathcal{R}$  is unique.  $\square$

### Proof of Theorem 2.4.3

It is clear that

$$J_{N-1}^v(h, x_{N-1}) = (\mu + 2\lambda h)x_{N-1}^2 \mathbb{E}_{N-1}^v[\tilde{\eta}_N] + \lambda x_{N-1}^4 \mathbb{E}_{N-1}^v[\tilde{\eta}_N^2].$$

Moreover, for  $n \leq N-1$ ,

$$\begin{aligned}
J_{n-1}^v(h, x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v [(\mu + 2\lambda h)Q_{n-1}(x) + \lambda Q_{n-1}^2(x)] \\
&= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v [(\mu + 2\lambda h)(u_n + Q_n(x)) + \lambda u_n^2 + 2\lambda u_n Q_n(x) + \lambda Q_n^2(x)],
\end{aligned}$$

where  $u_n = (x_{n-1} - x_n)^2 \tilde{\eta}_n - x_n \sigma_n \xi_n$ . So

$$\begin{aligned}
J_{n-1}^v(h, x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} (\mu + 2\lambda h)(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \lambda(x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v[\tilde{\eta}_n^2] \\
&\quad + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v[\sigma_n^2] + \mathbb{E}_{n-1}^v [(\mu + 2\lambda h + 2\lambda u_n)Q_n(x) + \lambda Q_n^2(x)] \\
&= \min_{0 \leq x_n \leq x_{n-1}} (\mu + 2\lambda h)(x_{n-1} - x_n)^2 \mathbb{E}_{n-1}^v[\tilde{\eta}_n] + \lambda(x_{n-1} - x_n)^4 \mathbb{E}_{n-1}^v[\tilde{\eta}_n^2] \\
&\quad + \lambda x_n^2 \Delta t \mathbb{E}_{n-1}^v[\sigma_n^2] \\
&\quad + \sum_{w \in V} p_{n-1}^{vw} \int_{\mathbb{R}} J_n^w \left( h + (x_{n-1} - x_n)^2 (w_2 - c/2) - x_n w_1 \sqrt{\Delta t} \xi, x_n \right) \rho(\xi) d\xi,
\end{aligned}$$

and  $x^* \in \mathcal{A}$  is optimal if in every step,  $x_n^*$  realizes the minimum for  $h = h_{n-1}(x^*)$  and  $x_{n-1} = x_{n-1}^*$ . Finally, for  $\mu \geq 0$ ,  $\mathbb{E}_0^v[\mu Q(x) + \lambda Q(x)^2]$  is strictly convex in  $x$ , and the optimal strategy  $x^* \in \mathcal{A}$  is unique.  $\square$

## A.2 Proofs for Chapter 3

### Proof of Theorem 3.2.1

Note that  $\mathbb{E}_0^v C(y, z) = \mathbb{E}_0^v R_0(y, z)$ , where

$$R_n(y, z) := \sum_{i=n+1}^N c x_i y_i + y_i^2 \eta_i.$$

For  $v \in V$  and  $x \in \mathbb{R}$ , we denote

$$J_n^v(x) := \min_{(y,z) \in \mathcal{A}_n(x)} \mathbb{E}_n^v R_n(y, z),$$

where  $\mathcal{A}_n(x)$  is the set of admissible strategies  $(y, z)$  with  $x_n = x$ . Then

$$J_{N-1}^v(x_{N-1}) = x_{N-1}^2 \mathbb{E}_{N-1}^v \eta_N = a_{N-1}^v x_{N-1}^2.$$

Furthermore, assuming  $J_n^w(x) = a_n^w x^2$  and  $a_n^w \geq 0$ , one obtains for  $\bar{q} = (q_1 + q_2)/2$ ,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v R_{n-1}(x_{n-1}) \\
&= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} cy_n \mathbb{E}_{n-1}^v x_n + y_n^2 \mathbb{E}_{n-1}^v \eta_n + \mathbb{E}_{n-1}^v R_n(x_n) \\
&= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} cx_{n-1}y_n - cy_n^2 - cy_n z_n \mathbb{E}_{n-1}^v b_n + y_n^2 \mathbb{E}_{n-1}^v \eta_n + \mathbb{E}_{n-1}^v R_n(x_n) \\
&= \min_{y_n, z_n} cx_{n-1}y_n - c\bar{q}y_n z_n + y_n^2 \mathbb{E}_{n-1}^v \tilde{\eta}_n \\
&\quad + \sum_w p_{n-1}^{vw} (\bar{q}J_n^w(x_{n-1} - (y_n + z_n)) + (1 - \bar{q})J_n^w(x_{n-1} - y_n)) \\
&= \min_{y_n, z_n} cx_{n-1}y_n - c\bar{q}y_n z_n + y_n^2 \mathbb{E}_{n-1}^v \tilde{\eta}_n \tag{A.2.1} \\
&\quad + \sum_w p_{n-1}^{vw} a_n^w (\bar{q}(x_{n-1} - y_n - z_n)^2 + (1 - \bar{q})(x_{n-1} - y_n)^2).
\end{aligned}$$

Provided that  $c \geq 0$  is small enough, the pair  $(y_n, z_n)$  must be chosen according to the first order condition

$$\begin{aligned}
y_n &= \frac{(x_{n-1} - \bar{q}z_n) (\sum_w p_{n-1}^{vw} a_n^w - c/2)}{\mathbb{E}_{n-1}^v \tilde{\eta}_n + \sum_w p_{n-1}^{vw} a_n^w} \\
z_n &= x_{n-1} - \frac{\sum_w p_{n-1}^{vw} a_n^w - c/2}{\sum_w p_{n-1}^{vw} a_n^w} y_n.
\end{aligned}$$

Therefore,

$$y_n = d_{n-1}^v x_{n-1}, \quad z_n = e_{n-1}^v x_{n-1},$$

and the value function becomes  $J_{n-1}^v(x_{n-1}) = a_{n-1}^v x_{n-1}^2$ . Moreover, for  $c \geq 0$  sufficiently small, one has  $a_{n-1}^v \geq 0$ .  $\square$

### Proof of Theorem 3.3.1

One clearly has

$$J_{N-1}^v(x_{N-1}) = \sum_{w \in V} p_{N-1}^{vw} \exp(\alpha x_{N-1}^2 w_2).$$

If  $\rho$  denotes the standard normal density, then for  $n \leq N - 1$ ,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^v \exp \left( \alpha (cx_n y_n + y_n^2 \eta_n - x_n \sigma_n \sqrt{\Delta t} \xi_n) + C_n(y, z) \right) \\
&= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \sum_{w \in V, b \in \{0,1\}} \int_{\mathbb{R}} d\xi \rho(\xi) p_{n-1}^{vw} \mathbb{P}[b_n = b | \xi, z_n] \times \\
&\quad \exp \left( \alpha (c(x_{n-1} - y_n - bz_n)y_n + y_n^2 w_2 - (x_{n-1} - y_n - bz_n)w_1 \sqrt{\Delta t} \xi) \right) \\
&\quad \times \mathbb{E}_n^w C_n(y, z) \\
&= \min_{y_n, z_n} \sum_{w \in V, b \in \{0,1\}} p_{n-1}^{vw} \psi(b, z_n, \alpha w_1 \sqrt{\Delta t} (x_{n-1} - y_n - bz_n)) \times \tag{A.2.2} \\
&\quad \exp \left( \alpha c(x_{n-1} - y_n - bz_n)y_n + \alpha y_n^2 w_2 + \frac{\alpha^2}{2} w_1^2 \Delta t (x_{n-1} - y_n - bz_n)^2 \right) \times \\
&\quad J_n^w(x_{n-1} - y_n - bz_n).
\end{aligned}$$

and since  $c$  was assumed to be sufficiently small the objective function in (A.2.2) is positive semidefinite in  $y_n$  and  $z_n$  hence the minimum is attained.  $\square$

## A.3 Proofs for Chapter 4

### Proof of Theorem 4.3.1

First notice that  $\mathbb{E}_n^{v,s}[Q_n(x, y, z)] = \mathbb{E}_n^{v,s}[R_n(y, z)]$  for

$$R_n(y, z) = \sum_{i=n+1}^N y_i' \Lambda_i' y_i + \alpha z_i' \text{diag}(b_{i-1}) z_i$$

It follows easily that

$$J_{N-1}^v(x_{N-1}) = x_{N-1}' \mathbb{E}_{N-1}^v [\Lambda_N'] x_{N-1} = x_{N-1}' A_{N-1}^v x_{N-1},$$

and inductively,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^{v,x_{n-1}}[R_{n-1}(y,z)] \\
&= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} y_n' \mathbb{E}_{n-1}^v[\Lambda_n'] y_n + \alpha z_n' \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n \\
&\quad + \mathbb{E}_{n-1}^{v,x_{n-1}} [\mathbb{E}[R_n(y,z) \mid (\Lambda_n, \sigma_n, \varphi_n, \delta_n), x_n]] \\
&= \min_{y_n, z_n} y_n' \mathbb{E}_{n-1}^v[\Lambda_n'] y_n + \alpha z_n' \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n \\
&\quad + \sum_w p_{n-1}^{vw} \left( \text{Tr}(A_n^w \text{Cov}_{n-1}^v(\text{diag}(b_{n-1}) z_n, \text{diag}(b_{n-1}) z_n)) \right. \\
&\quad \left. + (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n)' A_n^w \right. \\
&\quad \left. (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n) \right) \\
&= \min_{y_n, z_n} y_n' \mathbb{E}_{n-1}^v[\Lambda_n'] y_n + \alpha z_n' \text{diag}(q(v_3)) z_n \\
&\quad + \sum_w p_{n-1}^{vw} \left( z_n' \text{diag}((A_n^w)_{ii}) \text{diag}(q(v_3)(1 - q(v_3))) z_n \right. \\
&\quad \left. + (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n)' A_n^w (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n) \right)
\end{aligned}$$

for  $n \leq N - 1$ . The optimal  $(y_n^*, z_n^*)$  pair must satisfy,

$$\begin{bmatrix} U_{n-1}^v & B_{n-1}^v \\ (B_{n-1}^v)' & L_{n-1}^v \end{bmatrix} \begin{bmatrix} y_n^* \\ z_n^* \end{bmatrix} = \begin{bmatrix} \sum_w p_{n-1}^{vw} \tilde{A}_n^w x_{n-1}^* \\ (B_{n-1}^v)' x_{n-1}^* \end{bmatrix}$$

It follows that the unique optimal strategy is given by

$$y_n^* = G_{n-1}^v x_{n-1}^*, \quad \text{and} \quad z_n^* = F_{n-1}^v x_{n-1}^*.$$



and hence

$$\begin{aligned}
J_{n-1}^v(x_{n-1}^*) &= x_{n-1}^{*'} \left( (G_{n-1}^v)' \mathbb{E}_{n-1}^v [\Lambda_n'] G_{n-1}^v + \alpha (F_{n-1}^v)' \text{diag}(q(v_3)) (F_{n-1}^v) \right. \\
&\quad + \sum_w p_{n-1}^{vw} \left( (F_{n-1}^v)' \text{diag}((A_n^w)_{ii}) \text{diag}(q(v_3)(1-q(v_3))) (F_{n-1}^v) \right. \\
&\quad \left. \left. + (I - G_{n-1}^v - \text{diag}(q(v_3)) F_{n-1}^v)' A_n^w (I - G_{n-1}^v - \text{diag}(q(v_3)) F_{n-1}^v) \right) \right) x_{n-1}^* \\
&= x_{n-1}^{*'} A_{n-1}^v x_{n-1}^*
\end{aligned}$$

for  $n \leq N - 1$ . □

### Proof of Proposition 4.4.1

Define

$$\begin{aligned}
R_n(x, y, z) &= \sum_{i=n+1}^N \exp \left( \lambda y_i' \Lambda_i y_i + \lambda \alpha z_i' \text{diag}(b_{i-1}) z_i \right. \\
&\quad \left. + \frac{1}{2} \lambda^2 \Delta t (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i)' \text{diag}(\sigma_i) \Sigma \text{diag}(\sigma_i) \right. \\
&\quad \left. (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i) \right).
\end{aligned}$$

It follows that  $\mathbb{E}_n^v [\exp(\alpha Q_n(x, y, z))] = \mathbb{E}_n^v [R_n(x, y, z)]$  and hence

$$J_{N-1}^v(x_{N-1}) = \sum_{w \in W} p_{N-1}^{vw} \exp \left( \lambda x_{N-1}' w_1 x_{N-1} \right).$$

By induction one obtains

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_n(x_{n-1})} \sum_u \mathbb{P}_{n-1}^v(b_{n-1} = u) \mathbb{E}_{n-1}^v \left[ \exp \left( \lambda y'_n \Lambda'_n y_n + \lambda \alpha z'_n \text{diag}(u) z_n \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u) z_n)' \text{diag}(\sigma_n) \Sigma \right. \right. \\
&\quad \left. \left. \text{diag}(\sigma_n) (x_{n-1} - y_n - \text{diag}(u) z_n) \right) R_n(x, y, z) \mid b_{n-1} = u \right] \\
&= \min_{y_n, z_n} \sum_{w, u} p_{n-1}^{vw} \mathbb{P}_{n-1}^v(b_{n-1} = u) \exp \left( \lambda y'_n w'_1 y_n + \lambda \alpha z'_n \text{diag}(u) z_n \right. \\
&\quad \left. + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u) z_n)' \text{diag}(w_2) \Sigma \text{diag}(w_2) \right. \\
&\quad \left. (x_{n-1} - y_n - \text{diag}(u) z_n) \right) J_n^w(x_{n-1} - y_n - \text{diag}(u) z_n),
\end{aligned}$$

for  $n \leq N - 1$ . Therefore the minimizing  $(y_n^*, z_n^*)$  form the unique optimal strategy for the expected exponential cost.  $\square$

#### Proof of Theorem 4.4.2

We prove the Theorem by using backward induction. We first observe that

$$\mathbb{E}_n^{v,s} \left[ \hat{Q}_n(x, y, z) \right] = \mathbb{E}_n^{v,s} \left[ \hat{R}_n(x, y, z) \right]$$

where

$$\begin{aligned}
\hat{R}_n(x, y, z) &:= \sum_{i=n+1}^N \frac{1}{2} y'_i \tilde{\Lambda}_i y_i + \alpha z'_i \text{diag}(b_{i-1}) z_i \\
&\quad + \frac{\lambda}{2} \Delta t (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i)' \text{diag}(\sigma_i) \Sigma \text{diag}(\sigma_i) \\
&\quad (x_{i-1} - y_i - \text{diag}(b_{i-1}) z_i).
\end{aligned}$$

Then the terminal conditional computes as,

$$J_{N-1}^v(x_{N-1}) = \frac{1}{2} x'_{N-1} \mathbb{E}_{N-1}^v [\tilde{\Lambda}_N] x_{N-1}.$$

It follows,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{n-1}^{v,x_{n-1}}[\hat{R}_{n-1}(x,y,z)] \\
&= \min_{(y,z) \in \mathcal{A}_{n-1}(x_{n-1})} \frac{1}{2} y_n' \mathbb{E}_{n-1}^v[\tilde{\Lambda}_n] y_n + \alpha z_n' \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n \\
&\quad + \frac{\lambda}{2} \Delta t (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n)' \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \\
&\quad \quad (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n) \\
&\quad + \frac{\lambda}{2} \Delta t \text{Tr} \left( \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \text{Cov}_{n-1}^v(\text{diag}(b_{n-1}) z_n, \text{diag}(b_{n-1}) z_n) \right) \\
&\quad + \mathbb{E}_{n-1}^{v,x_{n-1}} \left[ \mathbb{E} \left[ \hat{R}_n(x,y,z) \mid (\Lambda_n, \sigma_n, \varphi_n, \delta_n), x_n \right] \right] \\
&= \min_{y_n, z_n} \frac{1}{2} y_n' \mathbb{E}_{n-1}^v[\tilde{\Lambda}_n] y_n + \alpha z_n' \text{diag}(q(v_3)) z_n \\
&\quad + \frac{\lambda}{2} \Delta t (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n)' \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \\
&\quad \quad (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n) \\
&\quad + \frac{\lambda}{2} \Delta t z_n' \text{diag} \left( \left( \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \right)_{ii} \right) \text{diag}(q(v_3)(1 - q(v_3))) z_n \\
&\quad + \sum_w p_{n-1}^{vw} \left( \text{Tr} \left( H_n^w \text{Cov}_{n-1}^v(\text{diag}(b_{n-1}) z_n, \text{diag}(b_{n-1}) z_n) \right) \right. \\
&\quad \quad \left. + (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n)' H_n^w \right. \\
&\quad \quad \quad \left. (x_{n-1} - y_n - \mathbb{E}_{n-1}^v[\text{diag}(b_{n-1})] z_n) \right) \\
&= \min_{y_n, z_n} \frac{1}{2} y_n' \mathbb{E}_{n-1}^v[\tilde{\Lambda}_n] y_n + z_n' \left( \alpha \text{diag}(q(v_3)) + \left( \sum_w p_{n-1}^{vw} \text{diag}((H_n^w)_{ii}) \right) \right. \\
&\quad \left. + \frac{\lambda}{2} \Delta t \text{diag} \left( \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)]_{ii} \right) \text{diag}(q(v_3)(1 - q(v_3))) \right) z_n \\
&\quad + (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n)' \\
&\quad \quad \left( \sum_w p_{n-1}^{vw} H_n^w + \frac{\lambda}{2} \Delta t \mathbb{E}_{n-1}^v[\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \right) \\
&\quad \quad (x_{n-1} - y_n - \text{diag}(q(v_3)) z_n)
\end{aligned}$$

for  $n \leq N - 1$ . The optimal  $(y_n^*, z_n^*)$  pair must satisfy,

$$\begin{bmatrix} \hat{U}_{n-1}^v & \hat{B}_{n-1}^v \\ (\hat{B}_{n-1}^v)' & \hat{L}_{n-1}^v \end{bmatrix} \begin{bmatrix} y_n^* \\ z_n^* \end{bmatrix} = \begin{bmatrix} \left( \sum_w p_{n-1}^{vw} \tilde{H}_n^w + \lambda \Delta t \mathbb{E}_{n-1}^v [\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \right) x_{n-1}^* \\ (\hat{B}_{n-1}^v)' x_{n-1}^* \end{bmatrix}$$

Therefore the unique optimal strategy is given by

$$y_n^* = \hat{G}_{n-1}^v x_{n-1}^*, \quad \text{and} \quad z_n^* = \hat{F}_{n-1}^v x_{n-1}^*.$$

and hence

$$\begin{aligned} J_{n-1}^v(x_{n-1}^*) &= x_{n-1}^{*'} \left( \frac{1}{2} (\hat{G}_{n-1}^v)' \mathbb{E}_{n-1}^v [\tilde{\Lambda}_n] \hat{G}_{n-1}^v + (\hat{F}_{n-1}^v)' \left( \alpha \text{diag}(q(v_3)) \right. \right. \\ &\quad \left. \left. + \left( \sum_w p_{n-1}^{vw} \text{diag}((H_n^w)_{ii}) + \frac{\lambda}{2} \Delta t \text{diag}(\mathbb{E}_{n-1}^v [\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)]_{ii}) \right) \right. \right. \\ &\quad \left. \left. \text{diag}(q(v_3)(1 - q(v_3))) \right) \hat{F}_{n-1}^v + (I - \hat{G}_{n-1}^v - \text{diag}(q(v_3)) \hat{F}_{n-1}^v)' \right. \\ &\quad \left. \left( \sum_w p_{n-1}^{vw} H_n^w + \frac{\lambda}{2} \Delta t \mathbb{E}_{n-1}^v [\text{diag}(\sigma_n) \Sigma \text{diag}(\sigma_n)] \right) \right. \\ &\quad \left. (I - \hat{G}_{n-1}^v - \text{diag}(q(v_3)) \hat{F}_{n-1}^v) \right) x_{n-1}^* \\ &= x_{n-1}^{*'} H_{n-1}^v x_{n-1}^* \end{aligned}$$

for  $n \leq N - 1$ . □

### Proof of Theorem 4.4.3

We prove the theorem by finding the dual of the optimization problems defined as in (4.4.1) and carrying out the backward recursion for the optimal value function. The terminal condition directly reads as

$$J_{N-1}^{v^{N-1}}(x_{N-1}^*) = \sum_{v^N \in W} p_{N-1}^{v^{N-1} v^N} \exp \left( \lambda x_{N-1}^{*'} v_1^N x_{N-1}^* \right),$$

hence  $A_{N-1}^{\mu^*}(u^{N-1}, v^N | v^{N-1}) := \lambda v_1^N$ . Assume that

$$J_n^{v^n}(x_n) = \sum_{u^n \dots v^N} p(u^n \dots v^N | v^n) \exp\left(x_n' A_n^\mu(u^n \dots v^N | v^n) x_n\right)$$

Now we find the dual of the following problem

$$\begin{aligned} J_{n-1}^{v^{n-1}}(x_{n-1}) &= \min_{(y,z) \in \mathcal{A}_n(x_{n-1})} \mathbb{E}_n^{v, x_{n-1}} [\exp(\lambda Q_n(x, y, z))] \\ &= \min_{y_n, z_n} \sum_{u^{n-1}, v^n} p_{n-1}^{v^{n-1} v^n} \mathbb{P}_{n-1}^{v^{n-1}}(b_{n-1} = u^{n-1}) \exp\left(\lambda y_n' v_1^{n'} y_n \right. \\ &\quad \left. + \lambda \alpha z_n' \text{diag}(u^{n-1}) z_n + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)' \text{diag}(v_2^n) \Sigma \right. \\ &\quad \left. \text{diag}(v_2^n) (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)\right) J_n^{v^n}(x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n), \\ &= \min_{y_n, z_n} \sum_{u^{n-1} \dots v^N} p(u^{n-1} \dots v^N | v^{n-1}) \exp\left(\lambda y_n' v_1^{n'} y_n + \lambda \alpha z_n' \text{diag}(u^{n-1}) z_n \right. \\ &\quad \left. + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)' \left(\text{diag}(v_2^n) \Sigma \text{diag}(v_2^n) \right. \right. \\ &\quad \left. \left. + A_n^\mu(u^n \dots v^N | v^n)\right) (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)\right) \end{aligned} \quad (\text{A.3.3})$$

One can equivalently solve the following problem instead of (A.3.3),

$$\begin{aligned} &\min_{y_n, z_n, \gamma} \log \left( \sum_{u^{n-1} \dots v^N} p(u^{n-1} \dots v^N | v^{n-1}) \exp(\gamma(u^{n-1} \dots v^N | v^{n-1})) \right) \quad (\text{A.3.4}) \\ \text{s.t.} \quad &\exp\left(\lambda y_n' v_1^{n'} y_n + \lambda \alpha z_n' \text{diag}(u^{n-1}) z_n + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)' \right. \\ &\quad \left. \left(\text{diag}(v_2^n) \Sigma \text{diag}(v_2^n) + A_n^\mu(u^n \dots v^N | v^n)\right) (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)\right) \\ &\quad - \gamma(u^{n-1} \dots v^N | v^{n-1}) \leq 0 \quad \forall (u^{n-1} \dots v^N) \end{aligned}$$

We assign the dual variables  $\mu(u^n \dots v^N | v^n)$  to the constraints in (A.3.4) and form the Lagrangean,

$$\begin{aligned} \mathcal{L}(y_n, z_n, \gamma, \mu) = & \log \left( \sum_{u^{n-1} \dots v^N} p(u^{n-1} \dots v^N | v^{n-1}) \exp(\gamma(u^{n-1} \dots v^N | v^{n-1})) \right) \\ & + \sum_{u^{n-1} \dots v^N} \mu(u^{n-1} \dots v^N | v^{n-1}) \left( \exp(\lambda y'_n v_1^{n'} y_n + \lambda \alpha z'_n \text{diag}(u^{n-1}) z_n) \right. \\ & + \frac{1}{2} \lambda^2 \Delta t (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n)' (\text{diag}(v_2^n) \Sigma \text{diag}(v_2^n) \\ & + A_n^\mu(u^n \dots v^N | v^n)) (x_{n-1} - y_n - \text{diag}(u^{n-1}) z_n) \\ & \left. - \gamma(u^{n-1} \dots v^N | v^{n-1}) \right). \end{aligned}$$

Since (A.3.4) has Slater's constraint qualification, strong duality holds. To obtain the dual we look at

$$\max_{\mu \geq 0} \min_{y_n, z_n, \gamma} \mathcal{L}(y_n, z_n, \gamma, \mu).$$

For the minimum to be bounded from below we need to impose

$$\sum_{u^{n-1} \dots v^N} \mu(u^n \dots v^N | v^n) = 1, \quad \mu(u^n \dots v^N | v^n) > 0 \quad \forall (u^n \dots v^N).$$

The minimizing strategy satisfying the first order condition reads as,

$$y_n^* |_{x_{n-1}, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1}, \varphi_{n-1}) = v^{n-1}} = G_{n-1}^\mu(v^{n-1}) x_{n-1}, \quad (\text{A.3.5})$$

$$z_n^* |_{x_{n-1}, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1}, \varphi_{n-1}) = v^{n-1}} = F_{n-1}^\mu(v^{n-1}) x_{n-1}, \quad \text{and} \quad (\text{A.3.6})$$

$$\frac{p(u^{n-1} \dots v^N | v^{n-1}) \exp(\gamma^*(u^{n-1} \dots v^N | v^{n-1}))}{\sum_{u^{n-1} \dots v^N} p(u^{n-1} \dots v^N | v^{n-1}) \exp(\gamma^*(u^{n-1} \dots v^N | v^{n-1}))} = \mu(u^{n-1} \dots v^N | v^{n-1})$$

$\forall(u^n \dots v^N)$ . Therefore the dual problem reads as,

$$\begin{aligned} & \max_{\mu} \mathcal{L}(y_n^*(\mu), z_n^*(\mu), \gamma^*(\mu), \mu) & (\text{A.3.7}) \\ \text{subject to} & \sum_{u^{n-1} \dots v^N} \mu(u^n \dots v^N | v^n) = 1, \quad \mu(u^n \dots v^N | v^n) > 0 \quad \forall(u^n \dots v^N). \end{aligned}$$

In particular, letting  $x_{n-1} = x_{n-1}^*$  and denoting the maximizer of (A.3.7) by  $\mu^*$  the optimal primal variables read by Karush–Kuhn–Tucker conditions from equations (A.3.5)-(A.3.6) as,

$$\begin{aligned} y_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v^{n-1}} &= G_{n-1}^{\mu^*}(v^{n-1})x_{n-1}^*, \\ z_n^* |_{x_{n-1}^*, (\Lambda_{n-1}, \sigma_{n-1}, \delta_{n-1})=v^{n-1}} &= F_{n-1}^{\mu^*}(v^{n-1})x_{n-1}^* \end{aligned}$$

for  $n \leq N - 1$ . Plugging back the optimal strategy we obtain,

$$J_{n-1}^{v^{n-1}}(x_{n-1}^*) = \sum_{u^{n-1} \dots v^N} p(u^{n-1} \dots v^N | v^{n-1}) \exp \left( x_{n-1}^*{}' A_n^{\mu^*}(u^{n-1} \dots v^N | v^{n-1}) x_{n-1}^* \right)$$

for  $n \leq N - 1$ . □

## A.4 Proofs for Chapter 5

### Proof of Theorem 5.2.1

We use backward induction to prove the theorem. It follows easily

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \min_{x \in \mathcal{A}_{N-1}(x_{N-1})} \sup_{Q \in \mathcal{P}_{N-1}^v} \delta^{N-1} \mathbb{E}_Q^{N-1, v} [u(V_{N-1}(x))] \\ &= \sup_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1, M_{N-1}}^v\}} \sum_{w \in W} q_w \delta^{N-1} \int u(V_{N-1}^w(x)) \Big|_{x_N=0} \end{aligned}$$

It follows inductively,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \sup_{Q \in \mathcal{P}_{n-1}^v} \delta^{n-1} \mathbb{E}_Q^{n-1,v} \left[ \int u(V_n(x)) \right] \\
&\quad + \mathbb{E}_Q^{n-1,v} \left[ \rho_n^v \circ \dots \circ \rho_{N-1} \left( \sum_{i=n+1}^N \int u(V_n(x)) \right) \right] \\
&= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \sum_{w \in W} q_w \left( \delta^{n-1} \int u(V_{N-1}^w(x)) + J_n^w(x_n) \right)
\end{aligned}$$

for  $n \leq N-1$ . □

### Proof of Corollary 5.2.2

It follows from Theorem 5.2.1 that the terminal value function is given by

$$J_{N-1}^v(x_{N-1}) = x_{N-1}^2 \delta^{N-1} \mathbb{E}_{Q^*}^{N-1,v}[\tilde{\eta}_N] = x_{N-1}^2 a_{N-1}^v,$$

where

$$Q^* |_{(\sigma_{N-1}, \eta_{N-1})=v} = \arg \max_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1, M_{N-1}}^v\}} \mathbb{E}_Q^{N-1,v}[\tilde{\eta}_N].$$

It also follows from Theorem 5.2.1,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \sum_{w \in W} q_w \left( \delta^{n-1} \int u(V_{n-1}^w(x)) + J_n^w(x_n) \right) \\
&= \min_{x_n} \sup_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} (x_{n-1} - x_n)^2 \delta^{n-1} \mathbb{E}_Q^{n-1,v}[\tilde{\eta}_n] + x_n^2 \sum_w q_w a_n^w,
\end{aligned}$$

for  $n \leq N-1$ . Therefore, the unique optimal strategy is given by

$$x_n^* = x_{n-1}^* \frac{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n]}{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n] + \sum_{w \in W} q_w^* a_n^w},$$

where

$$Q^* = \arg \max_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \frac{\delta^{n-1} \mathbb{E}_Q^{n-1,v}[\tilde{\eta}_n] \sum_{w \in W} q_w a_n^w}{\delta^{n-1} \mathbb{E}_Q^{n-1,v}[\tilde{\eta}_n] + \sum_{w \in W} q_w a_n^w},$$



and  $J_{n-1}^v(x_{n-1}^*)$  becomes

$$(x_{n-1}^*)^2 \frac{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n] \sum_{w \in W} q_w^* a_n^w}{\delta^{n-1} \mathbb{E}_{Q^*}^{n-1,v}[\tilde{\eta}_n] + \sum_{w \in W} q_w^* a_n^w} = (x_{n-1}^*)^2 a_{n-1}^v.$$

□

**Proposition A.4.1.** *For a function  $f : W \mapsto \mathbb{R}$ ,  $\theta \in \mathbb{R}_+$  and a reference transition probability distribution  $\mathbb{P}_n^v$ ,*

$$\sup_{Q \in \mathbb{Q}} \sum_{w \in W} q_w f(w) - q_w \frac{1}{\theta} \log \left( \frac{q_w}{p_n^{vw}} \right) = \frac{1}{\theta} \log \left( \sum_{w \in W} p_n^{vw} \exp(\theta f(w)) \right) \quad (\text{A.4.8})$$

*Proof.* Notice that the maximization in (A.4.8) is concave in  $Q$  and it can be solved by forming the Lagrangean:

$$L(\lambda) := \sum_{w \in W} q_w f(w) - q_w \frac{1}{\theta} \log \left( \frac{q_w}{p_n^{vw}} \right) + \lambda \left( \sum_{w \in W} q_w - 1 \right).$$

By a first order condition one obtains the maximizing distribution  $Q^*$ ;

$$q_w^* = \frac{p_n^{vw} \exp(\theta f(w))}{\sum_{w \in W} p_n^{vw} \exp(\theta f(w))},$$

and the objective function of (A.4.8) at  $Q^*$  evaluates as  $\frac{1}{\theta} \log \left( \sum_{w \in W} p_n^{vw} \exp(\theta f(w)) \right)$ .

□

**Corollary A.4.2.** *For any  $x \in \mathcal{A}$  one has*

$$\rho_n^v \circ \rho_{n+1} \cdots \rho_{N-1} \left( \sum_{i=n+1} \delta^{i-1} u(V_i(x)) \right) = \frac{1}{\theta} \log \left( \mathbb{E}_{\mathbb{P}}^{n,v} \left[ \exp \left( \theta \sum_{i=n+1} \int \delta^{i-1} u(V_i(x)) \right) \right] \right),$$

where  $\alpha_n^v$  is the relative entropy index as defined in (5.2.10).

*Proof.* Fix an  $x \in \mathcal{A}$ , then proceed by backwards induction,

$$\rho_{N-1}^v(\delta^{N-1}u(V_N(x))) = \sup_{Q \in \mathbb{Q}} \sum_{w \in W} q_w \int \delta^{N-1}u(V_N^w(x)) - \sum_{w \in W} q_w \frac{1}{\theta} \log \left( \frac{q_w}{p_{N-1}^{vw}} \right)$$

For  $w \in W$ , let  $f$  in Proposition A.4.1 be such that  $f(w) = \int \delta^{N-1}u(V_N^w(x))$ . It follows

$$\rho_{N-1}^v(\delta^{N-1}u(V_N(x))) = \frac{1}{\theta} \log \left( \mathbb{E}_{\mathbb{P}}^{N-1,v} \left[ \exp \left( \theta \int \delta^{N-1}u(V_N(x)) \right) \right] \right).$$

Inductively one obtains,

$$\begin{aligned} \rho_{n-1}^v \circ \dots \circ \rho_{N-1} \left( \sum_{i=n} \delta^{i-1}u(V_i(x)) \right) &= \sup_{Q \in \mathbb{Q}} \sum_{w \in W} q_w \int \delta^{n-1}u(V_n^w(x)) - \sum_{w \in W} q_w \frac{1}{\theta} \log \left( \frac{q_w}{p_{n-1}^{vw}} \right) \\ &\quad + \sum_{w \in W} q_w \frac{1}{\theta} \log \left( \mathbb{E}_{\mathbb{P}}^{n,w} \left[ \exp \left( \theta \sum_{i=n+1} \int \delta^{i-1}u(V_i(x)) \right) \right] \right) \end{aligned}$$

Using proposition A.4.1 with

$$f(w) = \int \delta^{n-1}u(V_n^w(x)) + \frac{1}{\theta} \log \left( \mathbb{E}_{\mathbb{P}}^{n,w} \left[ \exp \left( \theta \sum_{i=n+1} \int \delta^{i-1}u(V_i(x)) \right) \right] \right),$$

it follows that,

$$\rho_{n-1}^v \circ \rho_n \circ \dots \circ \rho_{N-1} \left( \sum_{i=n} \delta^{i-1}u(V_i(x)) \right) = \frac{1}{\theta} \log \left( \mathbb{E}_{\mathbb{P}}^{n-1,v} \left[ \exp \left( \theta \sum_{i=n} \int \delta^{i-1}u(V_i(x)) \right) \right] \right)$$

for  $n \leq N-1$ . □

### Proof of Theorem 5.2.3

Notice that

$$\begin{aligned} \min_{x \in \mathcal{A}_n(z)} \rho_n^v \circ \rho_{n+1} \cdots \rho_{N-1} \left( \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right) &\equiv \\ \min_{x \in \mathcal{A}_n(z)} \exp \left( \theta \rho_n^v \circ \rho_{n+1} \cdots \rho_{N-1} \left( \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right) \right) & \end{aligned}$$

and it follows from Corollary A.4.2,

$$J_n^v(x_n) = \min_{x \in \mathcal{A}_n(z)} \mathbb{E}_{\mathbb{P}}^{n,v} \left[ \exp \left( \theta \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right) \right].$$

We now use backward induction to prove the theorem. It follows easily that

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \min_{x \in \mathcal{A}_{N-1}(x_{N-1})} \mathbb{E}_{\mathbb{P}}^{N-1,v} \left[ \exp \left( \theta \delta^{N-1} u(V_N(x)) \right) \right] \\ &= \sum_{w \in W} p_{N-1}^{vw} \exp \left( \theta \delta^{N-1} \int u(V_N^w(x)) \right) \Big|_{x_N=0} \end{aligned}$$

and by induction,

$$\begin{aligned} J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}_{\mathbb{P}}^{n-1,v} \left[ \exp \left( \theta \delta^{n-1} u(V_n(x)) \right) \exp \left( \theta \sum_{i=n+1}^N \delta^{i-1} u(V_i(x)) \right) \right] \\ &= \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in W} p_{n-1}^{vw} \exp \left( \theta \delta^{n-1} \int u(V_n^w(x)) \right) J_n^w(x_n) \quad n \leq N-1. \end{aligned}$$

□

### Proof of Corollary 5.2.4

It directly follows from Theorem 5.2.3 that the terminal condition satisfies,

$$J_{N-1}^v(x_{N-1}) = \sum_{w \in W} p_{N-1}^{vw} \exp \left( \theta \delta^{N-1} x_{N-1}^2 (w_2 - c/2) \right).$$

One also obtains,

$$J_{n-1}^v(x_{n-1}) = \min_{0 \leq x_n \leq x_{n-1}} \sum_{w \in W} p_{n-1}^{vw} \exp(\theta \delta^{n-1} (x_{n-1} - x_n)^2 (w_2 - c/2)) J_n^w(x_n)$$

for  $n \leq N - 1$ . In every step  $x_n^*$  minimizes a strictly convex function, therefore the optimal strategy is unique.  $\square$

Next proposition follows from Lemma 8 in Skiadas (2013).

**Proposition A.4.3.** *Define  $f : W \mapsto \mathbb{R}$ , let  $\phi$  be the divergence index defined in Section 5.2.3. Fix a reference transition probability  $\mathbb{P}_n^v$  on  $W$ . Assume  $W := \{w_0, w_1, \dots, w_M\}$  for an  $M \in \mathbb{N}$  and  $f(w_0) = \min_{w \in W} f(w)$ . It follows*

$$\sup_{Q \in \mathbb{Q}} \sum_{w \in W} q_w f(w) - p_n^{vw} \phi\left(\frac{q_w}{p_n^{vw}}\right) \quad (\text{A.4.9})$$

attains its maximum by a  $Q \in \mathbb{Q}$  if and only if,

$$\sum_{w \in W} p_n^{vw} \phi'^{-1}(\phi'(0^+) - f(w_0) + f(w)) < 1. \quad (\text{A.4.10})$$

Moreover, if the maximum is attained the maximizing  $Q$  satisfies,

$$\frac{q_w}{p_n^{vw}} = \phi'^{-1}(\beta + f(w)) \quad \forall w \in W, \quad (\text{A.4.11})$$

where  $\beta$  is the unique solution to

$$\sum_{w \in W} p_n^{vw} \phi'^{-1}(\beta + f(w)) = 1 \quad \text{and} \quad \beta > \phi'(0^+) - f(w_0) \quad (\text{A.4.12})$$

*Proof.* For a  $Q \in \mathbb{Q}$  the objective function of equation (A.4.9) can be expressed as,

$$f(w_0) - p_n^{vw_0} \phi\left(\frac{1 - \sum_{i=1}^M q_{w_i}}{p_n^{vw_0}}\right) + \sum_{i=1}^M q_{w_i} (f(w_i) - f(w_0)) - p_n^{vw_i} \phi\left(\frac{q_{w_i}}{p_n^{vw_i}}\right). \quad (\text{A.4.13})$$

The equation of (A.4.13) is maximized by a  $Q \in \mathbb{Q}$  if and only if  $Q$  satisfies the first order condition, so it follows,

$$q_{w_n} = p_n^{vw_n} \phi'^{-1}(\beta + f(w_n)),$$

where  $\beta = \phi' \left( \frac{1 - \sum_{i=1}^M q_{w_i}}{p_n^{vw_0}} \right) - f(w_0)$ . Notice that  $\phi' \left( \frac{1 - \sum_{i=1}^M q_{w_i}}{p_n^{vw_0}} \right) > \phi'(0^+)$ , therefore the number  $\beta$  solving the equations in (A.4.12) implies the inequality in (A.4.10).

Now assume (A.4.10) holds. Then as a function of  $\beta$

$$\sum_{w \in W} p_n^{vw} \phi'^{-1}(\beta + f(w)),$$

is a continuous and an increasing function. When  $\beta$  is approaching  $\phi'(0^+) - f(w_0)$  it is less than 1 by assumption and as the divergence index grows super-linearly there exists a unique  $\beta$  such that equations in (A.4.12) are satisfied. For the  $\beta$  satisfying (A.4.12) we have  $\beta + f(w_n) > \phi'(0^+)$  therefore  $Q$  satisfying (A.4.11) is well defined and this implies the optimality of  $Q$ .  $\square$

### Proof of Theorem 5.2.5

We use backwards induction to prove the theorem. Let  $f$  in Proposition A.4.3 be such that  $f(w) = \delta^{N-1} \int u(V_N^w(x))$  for  $w \in W$ . Since  $\lambda$  is assumed to be sufficiently large we assume,

$$\sum_{w \in W} p_{N-1}^{vw} \delta^{N-1} \int u(V_N^w(x)) - \min_{w \in W} f(w) < \lambda,$$

hence (A.4.10) is satisfied for any  $x \in \mathcal{A}$ . It follows by using the maximizing distribution satisfying the equations (A.4.11) and (A.4.12) we obtain,

$$J_{N-1}^v(x_{N-1}) = \delta^{N-1} \mathbb{E}_{\mathbb{P}}^{N-1,v} \left[ \int u(V_N(x)) \right] + \frac{1}{2\lambda} \delta^{2(N-1)} \text{Var}_{\mathbb{P}}^{N-1,v} \left( \int u(V_N(x)) \right) \Big|_{x_N=0}.$$

Inductively,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \rho_{n-1}^v \circ \dots \circ \rho_{N-1} \left( \sum_{i=n} \delta^{i-1} u(V_i(x)) \right) \\
&= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \sup_{Q \in \mathbb{Q}} \delta^{n-1} \mathbb{E}_Q^{n-1,v} \left[ \int u(V_n(x)) \right] - \alpha_{n-1}^v(Q) \\
&\quad + \mathbb{E}_Q^{n-1,v} \left[ \rho_n \circ \dots \circ \rho_{N-1} \left( \sum_{i=n+1} \delta^{i-1} V_i(x) \right) \right] \\
&= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \mathbb{Q}} \delta^{n-1} \mathbb{E}_Q^{n-1,v} \left[ \int u(V_n(x)) \right] - \alpha_{n-1}^v(Q) + \sum_{w \in W} q_w J_n^w(x_n)
\end{aligned}$$

for  $n \leq N-1$ . Now let  $f(w) = \delta^{n-1} \int u(V_n^w(x)) + J_n^w(x_n)$ . By assumption  $\lambda$  is large enough such that,

$$\sum_{w \in W} p_{n-1}^{vw} \left( \delta^{n-1} \int u(V_n^w(x)) + J_n^w(x_n) \right) - \min_{w \in W} f(w) < \lambda$$

so that (A.4.10) is satisfied. By using the maximizing distribution satisfying the equations of (A.4.11) and (A.4.12) one obtains,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} \delta^{n-1} \mathbb{E}_{\mathbb{P}}^{n-1,v} \left[ \int u(V_n(x)) \right] + \frac{1}{2\lambda} \delta^{2(n-1)} \text{Var}_{\mathbb{P}}^{n-1,v} \left( \int u(V_n(x)) \right) \\
&\quad + \mathbb{E}_{\mathbb{P}}^{n-1,v} [J_n(x_n)] + \frac{1}{2\lambda} \text{Var}_{\mathbb{P}}^{n-1,v}(J_n(x_n)) + \frac{1}{\lambda} \delta^{n-1} \text{Cov}_{\mathbb{P}}^{n-1,v} \left( J_n(x_n), \int u(V_n(x)) \right)
\end{aligned}$$

for  $n \leq N-1$ . □

### Proof of Corollary 5.2.6

It follows directly from Theorem 5.2.5 that

$$J_{N-1}^v(x_{N-1}) = x_{N-1}^2 \mathbb{E}_{\mathbb{P}}^{N-1,v}[\tilde{\eta}_N] + \frac{1}{2\lambda} x_{N-1}^4 \text{Var}_{\mathbb{P}}^{N-1,v}(\tilde{\eta}_N),$$

and

$$\begin{aligned} J_{n-1}^v(x_{n-1}) &= \min_{0 \leq x_n \leq x_{n-1}} (x_{n-1} - x_n)^2 \mathbb{E}_{\mathbb{P}}^{n-1,v}[\tilde{\eta}_n] + \frac{1}{2\lambda} (x_{n-1} - x_n)^4 \text{Var}_{\mathbb{P}}^{n-1,v}(\tilde{\eta}_n) \\ &\quad + \mathbb{E}_{\mathbb{P}}^{n-1,v} [J_n(x_n)] + \frac{1}{2\lambda} \text{Var}_{\mathbb{P}}^{n-1,v}(J_n(x_n)) + \frac{1}{\lambda} (x_{n-1} - x_n)^2 \text{Cov}_{\mathbb{P}}^{n-1,v}(J_n(x_n), \tilde{\eta}_n) \end{aligned}$$

for  $n \leq N - 1$ . □

### Proof of Theorem 5.3.1

Notice that,

$$\begin{aligned} \min_{x \in \mathcal{A}_n(z)} \mu_n^v \circ \mu_{n+1} \dots \mu_{N-1} \left( \sum_{i=n+1}^N V_i(x) \right) &= \min_{x \in \mathcal{A}_n(z)} \mu_n^v \circ \mu_{n+1} \dots \mu_{N-1} (R_n(x)) \\ &\equiv \min_{x \in \mathcal{A}_n(z)} \exp(\theta \mu_n^v \circ \mu_{n+1} \dots \mu_{N-1} (R_n(x))) \end{aligned}$$

where,

$$R_n(x) := \sum_{i=n+1}^N (x_{i-1} - x_i)^2 \tilde{\eta}_i + \frac{1}{2} \theta x_i^2 \sigma_i^2 \Delta t.$$

It is easy to see,

$$\begin{aligned} J_{N-1}^v(x_{N-1}) &= \min_{x \in \mathcal{A}_{N-1}(x_{N-1})} \sup_{Q \in \mathcal{P}_{N-1}^v} \mathbb{E}_Q^{N-1,v} [\exp(\theta R_{N-1}(x))] \\ &= \sup_{Q \in \{P_{N-1,1}^v, \dots, P_{N-1, M_{N-1}}^v\}} \sum_{w \in W} q_w \exp(\theta x_{N-1}^2 (w_2 - c/2)). \end{aligned}$$

It follows by induction,

$$\begin{aligned}
J_{n-1}^v(x_{n-1}) &= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \exp(\theta \mu_{n-1}^v \circ \mu_n \dots \mu_{N-1}(R_{n-1}(x))) \\
&= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \exp\left(\theta \mu_{n-1}^v \left( (x_{n-1} - x_n)^2 \tilde{\eta}_n + \frac{1}{2} \theta x_n^2 \sigma_n^2 \Delta t \right. \right. \\
&\quad \left. \left. + \mu_n \circ \dots \mu_{N-1}(R_n(x)) \right)\right) \\
&= \min_{x \in \mathcal{A}_{n-1}(x_{n-1})} \sup_{Q \in \mathcal{P}_{n-1}^v} \mathbb{E}_Q^{n-1,v} \exp\left(\theta (x_{n-1} - x_n)^2 \tilde{\eta}_n + \frac{1}{2} \theta^2 x_n^2 \sigma_n^2 \Delta t \right. \\
&\quad \left. + \theta \mu_n \circ \dots \mu_{N-1}(R_n(x)) \right) \\
&= \min_{0 \leq x_n \leq x_{n-1}} \sup_{Q \in \{P_{n-1,1}^v, \dots, P_{n-1, M_{n-1}}^v\}} \sum_{w \in W} q_w \exp\left(\theta (x_{n-1} - x_n)^2 (w_2 - c/2) \right. \\
&\quad \left. + \frac{1}{2} \theta^2 x_n^2 w_1^2 \Delta t \right) J_n^w(x_n).
\end{aligned}$$

for  $n \leq N - 1$ . Therefore the minimizing  $x_n^*$  form the optimal strategy.  $\square$

#### Proof of Proposition 5.4.1

The proof follows trivially from the definition of  $\mathcal{A}_n(z)$  and by making the monotone transformation  $\tilde{J}_n^v(x_n) = \nu(J_n^v(x_n))$ .  $\square$



# Bibliography

- [1] Alfonsi, A., Fruth, A. and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions, *Quantitative Finance* 10, 143-157.
- [2] Alfonsi, A., Schied, A. and Slynko, A. (2012). Order book resilience, price manipulation, and the positive portfolio problem, *SIAM Journal on Financial Mathematics* 3, 511-533.
- [3] Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions, *Journal of Risk* 3, 5–39.
- [4] Almgren, R. (2003). Optimal execution with nonlinear impact functions and trading– enhanced risk, *Applied Mathematical Finance* 10, 1-18.
- [5] Almgren, R. (2012). Optimal trading with stochastic liquidity and volatility, *SIAM Journal on Financial Mathematics* 3(1), 163–181.
- [6] Bayraktar, E. and Ludkovski, M. (2011). Optimal trade execution in illiquid markets, *Mathematical Finance* 21, 681–701.
- [7] Bertsimas, D. and Lo, A. (1998). Optimal control of execution costs, *Journal of Financial Markets* 1, 1–50.
- [8] Cheridito, P. and Kupper, M. (2006). Time-consistency of indifference prices and monetary utility functions, *SSRN Preprint*.
- [9] Cheridito, P. and Kupper, M. (2011). Composition of time-consistent dynamic monetary risk measures in discrete time, *International Journal of Theoretical and Applied Finance*, 14(01), 137–162.
- [10] Cheridito, P. and Sepin, T. (2014a): Optimal trade execution under stochastic volatility and liquidity, *Applied Mathematical Finance*, 21(4), 342–362.
- [11] Cheridito, P. and Sepin, T. (2014b): Optimal trade execution with a dark pool and adverse selection, *SSRN Preprint*.

- [12] Cheridito, P. and Sepin, T. (2014c): Portfolio execution with a dark pool under stochastic volatility and liquidity, *SSRN Preprint*.
- [13] Degryse, Hans, Van Achter, Mark and Wuyts, Gunther. (2009). Shedding light on dark liquidity pools, *TILEC Discussion Paper* 2008–39.
- [14] Forsyth, P. A. (2011). A Hamilton Jacobi Bellman approach to optimal trade execution, *Applied Numerical Mathematics* 61, 241–265.
- [15] Ganchev, K., Nevmyvaka, Y., Kearns, M., and Vaughan, J. W. (2010). Censored exploration and the dark pool problem, *Communications of the ACM* 53(5), 99–107.
- [16] Gatheral, J. and Schied, A. (2011). Optimal trade execution under geometric brownian motion in the almgren and chris framework, *International Journal of Theoretical and Applied Finance* 14(3), 353–368.
- [17] Gatheral, J. and Schied, A. (2013). Dynamical models of market impact and algorithms for order execution, *Handbook on Systemic Risk* 579-602. Cambridge University Press.
- [18] Gatheral, J., Schied, A., and Slynko, A. (2012). Transient linear price impact and Fredholm integral equations. *Mathematical Finance* 22, 445–474.
- [19] Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18(2), 141–153.
- [20] Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty, *The American Economic Review*, 91(2), 60–66.
- [21] Klibanoff, P., Marinacci, M. and Mukerji, S. (2005). A smooth model of decision making under ambiguity, *Econometrica*, 73(6), 1849–1892.
- [22] Knight, F.H., (1921). Risk, Uncertainty and Profit, Houghton Mifflin. Boston, MA.
- [23] Kratz, P. (2014): An explicit solution of a nonlinear-quadratic constrained stochastic control problem with jumps: optimal liquidation in dark pools with adverse selection. *Mathematics of Operations Research*.
- [24] Kratz, P., and Schöneborn, T. (2013a): Optimal liquidation in dark pools. *EFA 2009 Bergen Meetings Paper*.
- [25] Kratz, P., and Schöneborn, T. (2013b): Portfolio liquidation in dark pools in continuous time. *Mathematical Finance*, available online.

- [26] Laurelle, S., Lehalle, C. A., and Pages, G. (2011). Optimal split of orders across liquidity pools: a stochastic algorithm approach, *SIAM Journal on Financial Mathematics* 2(1), 1042–1076.
- [27] Maccheroni, F., Marinacci, M. and Rustichini, A. (2006a). Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica*, 74(6), 1447–1498.
- [28] Maccheroni, F., Marinacci, M. and Rustichini, A. (2006b). Dynamic variational preferences, *Journal of Economic Theory*, 128(1), 4–44.
- [29] Makimoto, N., and Yoshihiko, S. (2010). Optimal execution of multiasset block orders under stochastic liquidity, *Institute for Monetary and Economic Studies Discussion Paper Series, Bank of Japan* 2010–E–25.
- [30] Moazeni, S., Coleman, T., Li, Y. (2013). Optimal execution under jump models for uncertain price impact, *Journal of Computational Finance*, 16(4), 1–44.
- [31] Obizhaeva, A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics, *Journal of Financial Markets* 16, 1–32.
- [32] Perold, A. F. (1988). The implementation shortfall: paper versus reality, *Journal of Portfolio Management* 14, 4–9.
- [33] Riedel, F. (2004). Dynamic coherent risk measures, *Stochastic Processes and their Applications*, 112(2), 185–200.
- [34] Schied, A. and Schöneborn, T. (2009). Risk aversion and the dynamics of optimal liquidations strategies in illiquid markets, *Finance and Stochastics* 13(2), 181–204.
- [35] Schied, A., Schöneborn, T., Tehranchi, M. (2010). Optimal basket liquidation for CARA investors is deterministic, *Applied Mathematical Finance* 17(6), 471–489.
- [36] Schied, A. (2013). Robust strategies for optimal order execution in the Almgren-Chriss framework, *Applied Mathematical Finance*, 20(3), 264–286.
- [37] Skiadas, C. (2013). Smooth ambiguity aversion toward small risks and continuous-time recursive utility, *Journal of Political Economy*, 121(4), 775–792.
- [38] Walia, N. (2006). *Optimal Trading: Dynamic Stock Liquidation Strategies*. Senior Thesis. Princeton University. Princeton, NJ.