High-Dimensional Portfolio Management: Taxes, Execution and Information Relaxations

Chun Wang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

High-Dimensional Portfolio Management:

Taxes, Execution and Information Relaxations

Chun Wang

Portfolio management has always been a key topic in finance research area. While many researchers have studied portfolio management problems, most of the work to date assumes trading is frictionless. This dissertation presents our investigation of the optimal trading policies and efforts of applying duality method based on information relaxations to portfolio problems where the investor manages multiple securities and confronts trading frictions, in particular capital gain taxes and execution cost.

In Chapter 2, we consider dynamic asset allocation problems where the investor is required to pay capital gains taxes on her investment gains. This is a very challenging problem because the tax to be paid whenever a security is sold depends on the tax basis, i.e. the price(s) at which the security was originally purchased. This feature results in high-dimensional and path-dependent problems which cannot be solved exactly except in the case of very stylized problems with just one or two securities and relatively few time periods. The asset allocation problem with taxes has several variations depending on: (i) whether we use the *exact* or *average* tax-basis and (ii) whether we allow the *full use* of *losses* (FUL) or the *limited use* of *losses* (LUL). We consider all of these variations in this chapter but focus mainly on the exact and average-cost tax-basis LUL cases since these problems are the most realistic and generally the most challenging. We develop several sub-optimal trading

policies for these problems and use duality techniques based on information relaxations to assess their performances. Our numerical experiments consider problems with as many as 20 securities and 20 time periods. The principal contribution of this chapter is in demonstrating that much larger problems can now be tackled through the use of sophisticated optimization techniques and duality methods based on information-relaxations. We show in fact that the dual formulation of exact tax-basis problems are much easier to solve than the corresponding primal problems. Indeed, we can easily solve dual problem instances where the number of securities and time periods is much larger than 20. We also note, however, that while the average tax-basis problem is relatively easier to solve in general, its corresponding dual problem instances are non-convex and more difficult to solve. We therefore propose an approach for the average tax-basis dual problem that enables valid dual bounds to still be obtained.

In Chapter 3, we consider a portfolio execution problem where a possibly risk-averse agent needs to trade a fixed number of shares in multiple stocks over a short time horizon. Our price dynamics can capture linear but stochastic temporary and permanent price impacts as well as stochastic volatility. In general it's not possible to solve even numerically for the optimal policy in this model, however, and so we must instead search for good sub-optimal policies. Our principal policy is a variant of an open-loop feedback control (OLFC) policy and we show how the corresponding OLFC value function may be used to construct good primal and dual bounds on the optimal value function. The dual bound is constructed using the recently developed duality methods based on information relaxations. One of the contributions of this chapter is the identification of sufficient conditions to guarantee convexity, and hence tractability, of the associated dual problem instances. That said, we do not claim that the only plausible models are those where all dual problem instances are convex. We also show that it is straightforward to include a non-linear temporary price impact as well as return predictability in our model. We demonstrate numerically that good dual bounds can be computed quickly even when nested Monte-Carlo simulations are required to estimate the so-called dual penalties. These results suggest that the dual methodology can be applied in many models where closed-form expressions for the dual penalties cannot be computed.

In Chapter 4, we apply duality methods based on information relaxations to dynamic zero-sum games. We show these methods can easily be used to construct dual lower and upper bounds for the optimal value of these games. In particular, these bounds can be used to evaluate sub-optimal policies for zero-sum games when calculating the optimal policies and game value is intractable.

Contents

List of Figures **v**

List of Tables **vii**

Acknowledgements **ix**

1 Introduction **1**

1.1 Tax-Aware Dynamic Asset Allocation 2

1.2 Dynamic Portfolio Execution with Transaction Cost 6

1.3 Information Relaxations and Dynamic Zero-Sum Games 9

2 Tax-Aware Dynamic Asset Allocation **12**

2.1 Introduction 12

2.2 Problem Formulation 14

2.3 Feasible Investment Policies 20

2.3.1 Tax-Blind Policies 21
2.3.2 The Buy-and-Hold Policy 22
2.3.3 The Rolling Buy-and-Hold Policies 23
2.3.4 Look-Ahead Policies 27
2.4 Evaluating Sub-Optimal Policies 29
2.4.1 Dual Problem Formulation 30
2.4.2 Dual Penalties 33
2.5 Numerical Experiments 34
2.5.1 Security Price and Market State Dynamics 34
2.5.2 Results 35
2.6 Conclusions and Further Research 39

3 Dynamic Portfolio Execution with Transaction Cost **41**

3.1 Introduction 41
3.2 The Portfolio Execution Problem and Model Description 44
3.2.1 Basic Model Description 46
3.2.2 Sub-Optimal Execution Policies 48
3.3 Evaluating Sub-Optimal Policies 50
3.3.1 The Dual Problem Formulation 53

3.3.2 Dual Convexity . . . 55
3.3.3 Using the Dual Formulation to Investigate More Complex Models . . 57
3.4 Numerical Results . . . 59
3.4.1 The Price Impact Dynamics . . . 60
3.4.2 Variance-Covariance Dynamics . . . 63
3.4.3 Results . . . 63
3.5 Extensions and Other Portfolio Execution Problems . . . 66
3.5.1 Computing Dual Bounds When Penalties Cannot Be Computed Explicitly . . . 67
3.5.2 Investigating the Temporary Price Impact Model . . . 72
3.5.3 Including a Non-Linear Temporary Price Impact . . . 73
3.5.4 Predictable State Dynamics . . . 76
3.5.5 Limit Order Book Models . . . 76
3.6 Conclusions and Future Research . . . 79
4 Information Relaxations and Dynamic Zero-Sum Games 84
4.1 Introduction . . . 84
4.2 Dual Bounds for Finite Horizon Zero-Sum Games . . . 85
4.2.1 Obtaining Dual Bounds for the Zero-Sum Game . . . 87

4.2.2 Constructing Dual Bounds for Player A's Dynamic Program 89

4.3 Dual Bounds for Stochastic Shortest Path Games 96

4.4 Conclusions and Further Research 101

Bibliography **102**

A Supplemental Content for Chapter 2 **111**

A.1 Generating LDS Scenarios for the RBH and Lookahead Policies 111

A.2 The Exact-Tax Basis FUL Problem 112

A.2.1 The FUL Model and Problem Formulation 112

A.2.2 Sub-Optimal Policies 114

A.2.3 Dual Bound 114

A.3 The Average Tax-Basis LUL Problem 117

A.3.1 Problem Formulation 117

A.3.2 Suboptimal Policies 119

A.3.3 Dual Bound 121

A.3.4 Numerical Experiments 124

A.4 The Average Tax-Basis FUL Formulation 125

B Supplemental Content for Chapter 3 **127**

B.1 Solving for the Sub-Optimal Policies 127

B.1.1 The Risk-Neutral Policy 128

B.1.2 When the Simple and Risk-Neutral Policies Coincide 131

B.1.3 Computing the OLFC Policy When $\mathbf{s}_j \geq \mathbf{0}$ Is Imposed 132

B.1.4 Computing the OLFC Policy When $\mathbf{s}_j \geq \mathbf{0}$ Is Not Imposed 136

B.1.5 Using Control Variates to Estimate the Primal Bounds 139

B.2 The Dual Problem for the Portfolio Execution Problem 141

B.2.1 Conditions that Guarantee the Positive Definiteness of $\mathbf{Q}$ and $\mathbf{Q}_{OLFC,t}$ 143

B.2.2 Computing Dual Penalties 144

B.2.3 Solving Dual Problem Instances 145

B.3 The Model of Bertsimas, Hummel and Lo 146

B.3.1 Model Description 147

B.3.2 Sub-Optimal Policies 149

B.3.3 The Dual Problem 151

B.3.4 Numerical Results 152

B.4 Additional Calibration Details for Section 3.4 154

B.4.1 Variance-Covariance Dynamics 154

B.4.2 Parameter Values for Top 50 Stocks in the S&P 157

C Review of Duality Based on Information Relaxations 159

List of Figures

3.1 (a) The deterministic component $x_{d,t}$, is assumed to be quadratic in t. (b) The expected temporary price impact of purchasing 1% of the average daily volume, expressed in basis points, is identical for each of the n securities. The dashed line is the time-average temporary price impact. 62

3.2 Certainty equivalent costs 65

3.3 Average cost and mean-standard deviation frontier 66

3.4 Dual bounds obtained using randomized LDS to estimate dual penalties . . . 71

3.5 Certainty equivalent costs for comparing model features 73

List of Tables

2.1 Results for exact tax-basis LUL problem with tax rate $\tau = 20\%$ 37

2.2 Results for exact tax-basis LUL problem with tax rate $\tau = 30\%$ 38

2.3 Results for exact tax-basis LUL problem with tax rate $\tau = 40\%$ 38

4.1 Optimal policies and game values 94

4.2 Suboptimal policies for the SSP matrix game 99

4.3 Game values, dual upper and lower bounds for the SSP matrix game. Means and standard errors (in parentheses) are reported for each of the two possible initial states, $x_0 = 1$ and $x_0 = 2$. 101

A.1 Results for average tax-basis LUL problem with tax rate $\tau = 40\%$ 125

B.1 Simulation results for the portfolio execution problem with non-negativity constraints. 153

B.2 Price impact coefficients for a 10×10 subset of the 25 stocks in BHL. All values have been multiplied by 10^{10}. 154

B.3 Eigenvalue analysis 156

B.4 GARCH(1,1) parameter estimates for the top 6 principal components 157

B.5 Parameter values for top 50 stocks in the S&P. 158

Acknowledgements

Many people say that writing the doctoral dissertation is a long and lonely journey, however I fortunately never feel it that way thanks to all the support that I have received and continue to receive along my journey. First and foremost, I owe the greatest gratitude to my academic advisor Professor Martin Haugh. Martin guided me into the research field and always adhered to high standards. I learned a great deal about the entire research process from him, and this manuscript would never be completed without his thrust, support and humor. I express my sincere appreciation to Professor Garud Iyengar for his invaluable mentorship and many ways in which he helped me. He always make himself available and I benefit tremendously from his insightful guidance and broad research interests.

I am very grateful to my dissertation committee: Professors Emanuel Derman, Tim Leung and Ciamac Moallemi for their careful reading of this manuscript. Their constructive suggestions helped improve this dissertation. Professor Steve Kou deserves my sincere thanks for many inspiring and interesting discussions. I would like to thank Professors Andrew Lo and Costis Maglaras for providing data and calibrated parameters for Chapter 3.

I feel honored to be a member of the IEOR department. I would like to thank all the faculty members. In particular Professors Daniel Bienstock, Donald Goldfarb, Karl Sigman and David Yao for the excellent first-year PhD core courses. I appreciate the convenience provided by alumni Dr. Ning Cai and Dr. Xianhua Peng during the BFS conference at Sydney in 2012. I also thank the former and current staff members, especially Jenny Mak, Jaya Mohanty, Carmen Ng, Darbi Roberts and Krupa Thakore, for providing administrative support and a convenient office environment.

Through out my PhD life at Columbia University, I deeply appreciate the friendship with my PhD fellows: Andrew Ahn, Chen Chen, Ningyuan Chen, Xinyun Chen, Yupeng Chen, Itai Feigenbaum, Jing Guo, Wenqi Hu, Bo Huang, Song Hee Kim, Anran Li, Xin Li, Juan Li, Yan Liu, Yunan Liu, Yin Lu, Yina Lu, Shiqian Ma, Cun Mu, Zhiwei Qin, Aya Wallwater, Zheng Wang, Xingbo Xu, Pengfei Zang, and Haowen Zhong. We have fruitful discussions, enjoy happiness and share pressure together.

I would like to thank my high school classmates Yongtao Cui, Kai Chen, Fan Guo, Fei Liu, Lin Mei, Di Xu and Yang Zhao, for being friends more than 15 years. We crossed the ocean and came to America together. We provide moral support to each other on our journey of pursuing PhD Degrees.

Finally my deepest thanks goes to my family. My grandfather Wang, Yongzhi excited my interest in science from my childhood, and my uncle Wang, Ziwei has shown full support during my entire doctoral study. Their frequent optimistic encouragement strengthens my willpower to overcome frustrations in the past and in the future. My grandmother Mu, Wenping, my aunt Kan, Xian and my sister Wang, Mingsi often comfort and warm my heart when I study far away from home. I thank my maternal grandparents Yin, Zesheng and Wang, Zhen in solely remembrance of them. I feel regretful that it takes me so long to complete this dissertation, and I expect they could share my joy at this moment as well. Last but not the least, I am deeply indebted to my father Wang, Zili and mother Jin, Qinhua for their unconditional love, understanding and support throughout my life. All of them make my humble achievement meaningful.

Chun Wang

April 17, 2014

To my parents:

Jin, Qinhua and Wang, Zili

Chapter 1

Introduction

Dynamic asset allocation and portfolio execution are the two vital components of financial portfolio management, and they have received considerable attention over the past two decades. Most of the work to date assumes trading is frictionless, however investors do confront trading friction in real financial markets. Taxation and execution cost are perhaps the two most significant factors in friction, and they indeed impact the performance of investment. For example, the magnitude of the capital gain taxes could be quite large, typical from 20% to 30%. We investigate the optimal portfolio policy when the investor is subject to capital gain taxes or trading incurs execution cost. Integrating the friction into dynamic portfolio optimization problems leads extraordinary complexity, and the difficulty increases rapidly with number of assets due to the well-known "curse-of-dimensionality", so in general it is not possible to solve for the optimal policy and we must instead search for good sub-optimal policies. But how good are these sub-optimal policies? In this dissertation we primarily focus on portfolio optimization problems and explore the use of the recent developed duality methods based on information relaxations to evaluate how far from optimality these sub-optimal policies can be. In the end we extend these dual methods to dynamic zero-sum games.

1.1 Tax-Aware Dynamic Asset Allocation

Dynamic asset allocation problems have played a central problem in finance since the pioneering work of Samuelson [62], Merton [54] and Hakannson [40]. Since then many researchers have studied the problem of how to dynamically allocate wealth among financial securities in order to optimize a given objective function which is typically some combination of the expected utility of terminal wealth and lifetime consumption. Most work considers problems with frictionless markets, and different problem formulations are typically obtained by varying the price dynamics and / or agent preferences. Relatively little work, however, has focussed on asset allocation in the presence of capital gains taxes. This is not because the problem is unimportant. Indeed the problem of how to efficiently invest and re-balance a portfolio in the presence of capital gains taxes is of considerable interest to investors and academics alike. Rather the problem has received relatively little attention because it is so challenging to solve. This is because the taxes that are owed whenever a security is sold generally depends on its *tax basis*, i.e. the price(s) at which the security was originally purchased. This feature results in high-dimensional and path-dependent problems which can only be solved for extremely simple cases. In Chapter 2, we consider dynamic asset allocation problems where the investor is required to pay capital gains taxes on her investment gains.

Constantinides [20] was among the first to study the asset allocation problem with taxes. He showed that the optimal investment policy can be separated from the tax timing problem if short-sales are allowed and are costless. In particular, when an investor needs to reduce her position in a stock with an embedded capital gain, she prefers to short sell the stock rather than selling from her current holdings[1]. She, therefore, succeeds in re-balancing her portfolio without triggering a tax liability. In practice, of course, short-sales incur collateral costs and may also not be permitted by the tax authorities[2].

Dybvig and Koo [30] studied the asset allocation problem with taxes when short-sales constraints are imposed so that the separation result of Constantinides does not apply. They formulated the problem using the so-called *exact* tax basis and solved a problem with just one risky stock and four time periods. More recently, DeMiguel and Uppal [28] used a stochastic programming approach to solve problems with just one stock and ten time periods, as well as problems with two stocks and seven time periods. Unfortunately, solving larger problems is numerically intractable because the number of state variables and number of constraints grows exponentially with the number of time periods.

An alternative modeling approach is to use the *average* tax basis when determining tax liabilities. The average tax basis for a given security is the weighted average purchase price of the current holdings of the security in the portfolio. The U.S tax code currently

[1]This strategy is sometimes referred to as “shorting the box”.
[2]The 1997 Tax Reform Act in the U.S rules out “shorting the box” transactions.

allows investors to choose between using the exact tax basis and the average tax basis in determining their tax liabilities. While using the average tax basis is generally sub-optimal from the investor's point of view, it requires less record-keeping, and from a modeling point of view, does not result in the exponential explosion of state variables that occurs when the exact tax basis is used. Average tax-basis problems are therefore more amenable to dynamic programming (DP) but are nonetheless still challenging to solve. In particular due to the so-called "curse-of-dimensionality" it is only possible to exactly solve problems with just one or two stocks even when the average tax basis is used.

Dammon et al. [23, 25] were among the first to consider the average tax-basis problem. They considered problems with just one stock and multiple time periods. Dammon et al. [24] and Garlappi et al. [35] considered the case of two stocks. Gallmeyer et al. [34] solved a two-stock problem as well as a problem with just one stock and a put option on that stock. It is worth noting that DeMiguel and Uppal [28] showed that the certainty equivalent loss in wealth was small when using the average tax basis instead of the exact tax basis. Although they only showed this for the relatively small problem sizes they considered, one would expect this observation to hold more generally. Tahar et al. [67] formulated the continuous time version of the model in Dammon et al. [23] and studied properties of the value function in a one-stock infinite horizon problem.

We also note that most of the literature on tax-aware asset allocation assumes the so-called *full use of losses* (FUL) model where the net capital losses in any given period result in an immediate tax rebate. This is in contrast to the *limited use of losses* (LUL) model where the net capital losses do not result in an immediate tax rebate; instead, the net losses can only be used to offset capital gains in future periods. Only the LUL model is consistent with the tax codes encountered in practice; but LUL problems are more challenging to solve than FUL problems since in the LUL case we need to keep track of the sum of prior losses that have not already been used to offset gains. Gallmeyer and Srivastava [33] discussed the impact of the LUL model while more recently, Ehling et al. [31] solved a two-stock average-tax basis LUL problem with 80 time periods[3].

While we consider all variations of tax-basis and use-of-losses assumptions, we will focus mainly on the exact and average-cost tax-basis LUL cases, since these problems are perhaps the most realistic, and generally the most challenging. We develop several sub-optimal trading policies for these problems and use duality techniques based on information relaxations to assess their performances. So instead of trying to solve these problems exactly we settle for *provably* good sub-optimal solutions. Our numerical experiments consider problems with as many as twenty securities and twenty time periods.

The principal contribution of Chapter 2 is to demonstrate that much larger problems can now be tackled through the use of sophisticated optimization techniques and duality methods based on information-relaxations [16, 60]. These dual methods can be used to compute dual bounds on the optimal value function of stochastic dynamic programs. To the best of our knowledge, we are the first to construct valid dual bounds for tax-aware dynamic asset allocation problems. We show that the dual formulations of the exact tax-

[3]Note that their algorithm requires 90 hours with 100 CPUs working in parallel to solve the problem. They do not provide processor or software details.

basis problems are much easier to solve than the corresponding primal problems. We can easily solve dual problem instances where the number of securities and time periods is much larger than twenty. While the average tax-basis (primal) problem is relatively easier to solve than the exact tax basis (primal) problem, the dual of the average tax-basis problem is non-convex and thus more difficult to solve. We therefore propose approaches for bounding these dual problem instances so that valid lower and upper bounds for the average tax-basis problem can still be obtained. While we only consider problems where short-term and long-term capital gains are both taxed at the same rate, it will be clear that we can easily handle asymmetric tax rates.

1.2 Dynamic Portfolio Execution with Transaction Cost

Dynamic portfolio execution is one of the most important practical problems faced by institutional investors and brokerage firms today. This problem requires the purchase or sale of a large number of shares in multiple stocks over a short time horizon. The large number of shares and short time horizon can result in a very significant market impact that can in turn greatly increase the cost of trading. This of course is in contrast to traditional dynamic portfolio optimization problems where trading is generally assumed to be frictionless. Because of the potentially large trading costs, it is necessary to model market impact explicitly when determining portfolio execution policies.

The literature on controlling execution costs is extensive and we can barely do justice to it here. Most of the early work, beginning with Bertsimas and Lo [11] and Almgren and

Chriss [6], considered single-stock execution problems which they formulated as dynamic control problems. Other related work in this spirit includes, for example, Almgren [4], Huberman and Stanzl [44] and Gatheral and Schied [37]. Generally these papers consider the *macro* or *scheduling* component of the execution problem, that is, the problem of deciding how to slice the order, when the execution algorithm should trade and in what size. These models might, for example, take a time horizon of 1 day and break this period up into 5-minute time periods. Assuming a trading day of 6.5 hours duration, this yields a finite-horizon control problem with 78 time periods.

In the last decade, the growth of electronic exchange markets has led to the modeling of limit order book dynamics. These models, beginning with Obizhaeva and Wang [56], and extended by Alfonsi, Fruth and Schied [2] and Predoiu, Shaikhet and Shreve [58], model the market impact and the *decay* of market impact at the limit order book level. Often formulated as continuous-time problems, the resulting stochastic control problems generally yield deterministic execution policies due to the assumption of relatively simple price dynamics. These papers can also be viewed as considering the scheduling component of the execution policy although they are inspired by micro foundations in that they model the transient behavior of the limit order book. They assume that trades are executed as market orders whose price impact depends on the shape of the order book at the time of execution. In contrast, Cont and Kukanov [21] and Huitema [46] consider the use of both limit and market orders. Other papers related to execution in limit order books include, for example, Cont, Stoikov and Talreja [22] and Gatheral [36], while Toth et al. [69] focus on market impact rather than execution.

More recently so-called *dark pools* have grown in importance as an alternative form of trading. There is no order book in dark pool venues and buyers/sellers are matched electronically without revealing any information. Trades in dark pools generally have a smaller price impact but order execution is uncertain. Recent work on optimal execution in these venues include for example Kratz and Schoneborn [48], and Buti et al. [17], both of which consider the use of a classical exchange and dark pools for optimal execution.

While most research has focussed on the single-stock execution problem, there has also been some work on the *portfolio* execution problem. Early work in this direction, which again focuses on the macro component of the problem, includes Bertsimas, Hummel and Lo [10] and Almgren and Chriss [5]. More recent work includes Huberman and Stanzl [45], Schied, Schoneborn and Tehranci [63] and Lim and Wimonkittiwat [51]. Giesecke, Tsoukalas and Wang [70] consider the portfolio execution problem at the limit order book level and formulate their control problem as an equivalent static convex optimization problem. In another recent paper, Moallemi and Saglam [55] consider general dynamic portfolio optimization problems, including portfolio execution problems, and propose optimizing over linear rebalancing rules in order to find good sub-optimal policies. Their optimization problem is therefore static in nature and amenable to standard convex optimization algorithms. Their paper is quite similar in spirit to our work in that they also focus on developing good sub-optimal policies and also use duality based on information relaxations to demonstrate the effectiveness of these policies.

In Chapter 3 we consider the portfolio execution problem of a possibly risk-averse agent who needs to account for temporary and permanent market impact costs, as well as other market features such as stochastic liquidity and return predictability among others. We formulate the problem, search for good sub-optimal policies, and again use the duality method based on information relaxations to evaluate these sub-optimal policies. We also recognize that even "realistic" models are still only an approximation to the true market dynamics, and we therefore study how these dual methods can also be used to estimate how robust a given policy is to departures from the assumed model. An additional contribution of this chapter is that valid and tight dual bounds can still be computed efficiently even when the dual penalties are not explicitly available and need to be estimated using Monte-Carlo.

1.3 Information Relaxations and Dynamic Zero-Sum Games

In Chapter 2 and Chapter 3, we use dual methods based on information relaxations to assess the performance of sub-optimal policies for tax-aware dynamic asset allocation problem and dynamic portfolio execution problem respectively. The use of duality techniques to construct good bounds for dynamic control problems began independently with Haugh and Kogan [41] and Rogers [59] in the context of optimal stopping problems and the pricing of American options. See also Andersen and Broadie [7] and Jamshidian [47]. These techniques were then extended to multiple optimal stopping problems by Meinshausen and Hambly [53] and Schoenmakers [64]. Bender [8] and Bender, Schoenmakers and Zhang [9] provided further extensions of this work. A significant development came with Brown, Smith and Sum [16] (hereafter denoted by BSS) and Rogers [60], who independently extended these duality techniques to more general stochastic dynamic programs. This has now become a very significant research area. More recent applications and developments can be found in Brown and Smith [15], Desai, Farias and Moallemi [29], Lai, Margot and Secomandi [49], Lai et al [50], Moallemi and Saglam [55] and Haugh and Lim [42]. The latter paper makes the point that some of the ideas behind these dual techniques are actually not so recent and date back to earlier work, including Davis and Karatzas [26] in the case of American options, and Davis and Zervos [27] for linear-quadratic control. Chandramouli and Haugh [18] also showed that the earlier duality results for multiple optimal stopping problems could easily be derived using the more general framework of BSS [16]. We provide a brief review of the duality theory based on information relaxations in Appendix C. We refer the reader to BSS [16] for the theoretical details underpinning the dual methods.

In Chapter 4 which is the last part of this dissertation, we apply these duality methods to dynamic two-person zero-sum games. Our extension of the information relaxation technology to dynamic zero-sum games is motivated in part by Beveridge and Joshi [12] who generalized the dual optimal stopping work of Haugh and Kogan [41] and Rogers [59] to zero-sum optimal stopping games. The main contribution of this chapter is to demonstrate that the more general dual approach of BSS [16] and Rogers [60] can also be easily applied to dynamic zero-sum games. While the results are easy to prove, there are many interesting applications including pursuit-evasion games, heads-up poker and many two-person computer games.

Our ultimate goal is to apply these dual techniques to very complex zero-sum games which cannot be solved to optimality.

Chapter 2

Tax-Aware Dynamic Asset Allocation

2.1 Introduction

The challenges of the tax-aware dynamic asset allocation problem root in the taxation code. In order to compute capital gains or losses, the taxation code needs to specify the tax basis to which the selling price of a security will be compared. The *exact* tax basis is defined as the price at which the security was originally purchased. This feature results in path-dependent problems because the investor is required to keep track of the basis price for every single transaction along the investment, and therefore problems cannot be solved exactly except in the case of just one or two securities and relatively few time periods. An alternative modeling approach is to use the *average* tax basis, which is defined as the weighted average purchase price of the current holdings of a security, to tackle with the path dependency. Average tax-basis problem requires less record-keeping. However, due to "curse-of-dimensionality" it is still hard to be solved exactly if there are many securities in the portfolio. The taxation code also needs to specify how the realized capital losses can be used to offset gains. Most of the literature asumes the so-called *full use of losses* (FUL) model where the investor faces no restrictions on the use of losses and the net capital losses generate a tax rebate which can be immediately invested. This is in contrast to the *limited use of losses* (LUL) model where the investor can only use realzied capital losses to offset current realized capital gains and unused capital losses can be carried forward to future trading dates. Only the LUL model is consistent with the tax codes in practice, however, LUL problems are more difficulty to solve than FUL problems becuase the accumulated unused losses has to be tracked.

The tax-aware asset allocation problem has several variations depending on the tax basis adopted and the rule of using losses. We consider all of these variations but focus on the exact and average tax-basis LUL cases because they are the most realistic and most challenging. We investigate several sub-optimal trading policies and use the duality techniques based on information relaxations to evaluate the quality of their performances. The remainder of this chapter is organized as follows. In Section 2.2 we formulate the problem, focusing mainly on the exact tax-basis LUL case. We describe our sub-optimal policies for tackling this problem in Section 2.3. Section 2.4 describes how we use duality techniques based on information relaxations to construct dual bounds for these problems. We describe our numerical results in Section 2.5 and conclude in Section 2.6. In Appendix A.1 we give additional details on how we solve for the sub-optimal policies while in Appendix A.2 we discuss the exact tax-basis FUL problem in further detail. In Appendices A.3 and A.4, we consider the average tax-basis LUL and FUL problems, respectively, and describe how the non-convexity of the corresponding dual problem instances can be addressed.

2.2 Problem Formulation

We consider the problem of a risk-averse agent who can trade in multiple assets over a fixed time horizon. The agent is also required to pay taxes on any realized capital gains that she incurs. Our framework is an extension of the multiple assets model of Dybvig and Koo [30] and DeMiguel and Uppal [28]. We consider a discrete-time economy with equally spaced trading dates $t = 0, 1, ..., T$. The security market consists of a risk-free cash account and K risky securities. We let b_t denote the cash-account holding at time t and assume it earns an after-tax risk-free return of r_0 per period. We let $p_{t,k}$ denote the time-t price of the k-th risky security and let $n_{j,t,k}$ denote the number of units of this security that was bought at time j for $j = 0, ..., t$ and still held *after* trading at time t for $t = 0, ..., T$. The *exact* tax-basis of these $n_{j,t,k}$ units is then given by $p_{j,k}$[1]. We define the time-t stock price vector $\mathbf{p}_t := [p_{t,1} \ \ldots \ p_{t,K}]'$ and time-t stock holding vector $\mathbf{n}_{j,t} := [n_{j,t,1} \ \ldots \ n_{j,t,K}]'$ for $j = 0, ..., t$. We also assume that borrowing and short-sales are prohibited so that the aforementioned separation result of Constantinides [20] does not apply.

In most of the previous literature[2], the tax aware portfolio selection problem is formulated using the FUL model where the realized capital losses earn an immediate tax rebate. In

[1] We discuss the average tax-basis problem in Appendices A.3 and A.4.

[2] See, e.g. Dybvig and Koo [30], Dammon et al. [23, 24, 25], Garlappi et al. [35], DeMiguel and Uppal [28], Gallmeyer et al. [34] and Tahar et al. [67].

practice, however, tax codes only allow for the limited use of losses in that an investor can only use realized capital losses to offset realized capital gains. Unused capital losses can then be carried forward to offset future capital gains. In this chapter we will focus mainly on the more challenging LUL model.[3] Let g_t denote the time t capital gains and let l_t denote the accumulated unused realized capital losses after trading at time t. Then the LUL tax rule can be modeled as follows:

$$c_t = \sum_{j=0}^{t-1} (\mathbf{p}_t - \mathbf{p}_j)'(\mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}) + l_{t-1} \tag{2.2.1}$$

$$g_t = \max\{c_t, 0\} \tag{2.2.2}$$

$$l_t = \min\{c_t, 0\}. \tag{2.2.3}$$

Since $\mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}$ denotes the number of units of securities that were purchased at time j and sold at time t, the tax basis for these securities is $\mathbf{p}_j$; therefore, c_t is the total realized capital gains or losses from trading at time t after offsetting by the accumulated unused realized capital losses l_{t-1} carried forward from the previous periods. When c_t is positive, the taxable capital gains g_t is set to c_t; otherwise c_t is negative and represents the total unused realized losses. We assume that the unused realized losses can be carried forward all the way to the final period T. We also assume that both long-term and short-term capital gains are taxed at the same rate τ. Note, however, that we could relax both of these assumptions by introducing additional state variables.

The goal is to maximize the agent's expected utility of terminal after-tax wealth b. We

[3]We discuss the FUL model in Appendices A.2 and A.4.

assume that the investor's utility function, $U(\cdot)$, belongs to the CRRA class of utility function so that

$$U(b) := \begin{cases} \frac{b^{1-\gamma}}{1-\gamma} & \gamma > 0 \text{ and } \gamma \neq 1 \\ \ln(b) & \gamma = 1 \end{cases} \tag{2.2.4}$$

where γ is the coefficient of relative risk aversion. The tax-aware asset allocation problem can then be formulated as follows:

$$\max_{(b_t,\mathbf{n}_{j,t},g_t,l_t)\in\mathcal{F}_t,\ t=0,\ldots,T} \quad \mathbb{E}_0\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \tag{2.2.5}$$

$$\text{s.t.} \quad b_0 + \mathbf{p}_0'\mathbf{n}_{0,0} = w_0 \tag{2.2.6}$$

$$b_t + \sum_{j=0}^{t}\mathbf{p}_t'\mathbf{n}_{j,t} + \tau g_t = b_{t-1}r_0 + \sum_{j=0}^{t-1}\mathbf{p}_t'\mathbf{n}_{j,t-1} \qquad t \geq 1 \tag{2.2.7}$$

$$\mathbf{n}_{t,t} \geq \mathbf{0} \qquad t \geq 0 \tag{2.2.8}$$

$$\mathbf{n}_{j,t-1} \geq \mathbf{n}_{j,t} \geq \mathbf{0} \qquad t \geq 1, j < t \tag{2.2.9}$$

$$b_t \geq 0 \qquad t \geq 0 \tag{2.2.10}$$

$$g_t + l_t = \sum_{j=0}^{t-1}(\mathbf{p}_t - \mathbf{p}_j)'(\mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}) + l_{t-1} \qquad t \geq 1 \tag{2.2.11}$$

$$g_t \geq 0 \qquad t \geq 1 \tag{2.2.12}$$

$$l_t \leq 0,\ l_0 = 0 \qquad t \geq 1 \tag{2.2.13}$$

and security price and state variable dynamics

where we use $\{\mathcal{F}_t\}_{t=0,\ldots,T}$ to denote the filtration generated by the security price vectors as well as any other state variables in the model. The investor is assumed to have an initial wealth w_0; therefore, constraint (2.2.6) is the initial budget constraint. The budget constraints for $t \geq 1$ are in (2.2.7) where the right-hand side is the pre-trade wealth at time t and the sum of the first two terms on the left-hand side represents the post-trade wealth after paying the capital gains taxes τg_t. The decision variables $\mathbf{n}_{j,t}$ allow the agent to track the exact tax basis of each share in the portfolio. Short-selling is ruled out by constraints (2.2.8) and (2.2.9). In addition, (2.2.9) also reflects the fact that any shares purchased at time j can only be sold in later periods. Constraint (2.2.10) prohibits borrowing via the cash account.

The taxable capital gain g_t is required to be non-negative so the agent can never receive a tax rebate in (2.2.7). Since paying additional taxes is always sub-optimal, the optimal solution of (2.2.5) will never simultaneously have $g_t > 0$ and $l_t < 0$. We can therefore replace the LUL tax rule described in (2.2.1)-(2.2.3) with constraints (2.2.11)-(2.2.13). We also assume an initial unused loss $l_0 = 0$.

We assume the after-tax return r_0 of the risk-free cash account is constant over time. The gross return of the k-th risky security between times $t-1$ and t is denoted by $r_{t,k}$. We assume this return is stochastic and define the time t return vector, $\mathbf{r}_t := [r_{t,1} \ ... \ r_{t,K}]'$. Without loss of generality, the initial stock prices are set to one, i.e., $\mathbf{p}_0 = [1 \ ... \ 1]'$ and we assume the price dynamics are given by

$$\mathbf{p}_t = \mathbf{p}_{t-1} \cdot \mathbf{r}_t, \tag{2.2.14}$$

where $\cdot$ denotes the componentwise multiplication. In contrast to most of the literature which assumes IID price dynamics, we assume $\mathbf{r}_{t+1}$ is driven by a Markovian state vector $\mathbf{z}_t$ so that $\mathbf{r}_{t+1}$ is independent of $\mathbf{r}_j$ for all $j \leq t$ conditional on $\mathbf{z}_t$.

In order to simplify later problem formulations we define $\mathbf{x}_t := [b_t\ \mathbf{n}'_{0,t}\ ...\ \mathbf{n}'_{t,t}\ g_t\ l_t]$ and $\mathbf{p}_{0:t} := [\mathbf{p}'_0\ \ldots\ \mathbf{p}'_t]'$ which is a vector of the prices of all risky securities up to and including time t. $\mathbf{x}_{t-1}$ and $\mathbf{p}_{0:t}$ completely describe the positions and tax basis of the agents's portfolio before trading at time t. We denote the set of feasible trades at time t by $\mathbb{X}_t$. Thus,

$$\mathbb{X}_0 := \left\{\mathbf{x}_0 \mid \mathbf{x}_0 \in \mathcal{F}_0 \text{ satisfies constraints (2.2.6), (2.2.8) and (2.2.10) at } t = 0\right\}$$

$$\mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t}) := \left\{\mathbf{x}_t \mid \mathbf{x}_t \in \mathcal{F}_t \text{ satisfies constraints (2.2.7)–(2.2.13) at time t}\right\}, \quad t \geq 1.$$

Unfortunately, it is impossible to solve the exact tax-basis LUL problem due to the large number of state variables, constraints and path-dependence induced by the need to keep track of the tax basis for each security. Instead we seek good sub-optimal policies which is the subject of Section 2.3. But first we discuss the problem where the agent does not have to pay capital gains taxes. The solution to this no-tax problem will be of use to us in constructing some of our sub-optimal policies as well as constructing dual bounds on the optimal value function for the exact tax-basis LUL problem.

The No-Tax Problem Formulation

In the no-tax problem capital gains are not taxed so that $\tau = 0$. The portfolio optimization problem can then be solved using dynamic programming (DP) as long as the dimension of the exogenous state vector $\mathbf{z}_t$ is sufficiently small.

Let $V_t^N(w_t, \mathbf{z}_t)$ denote the time-t optimal value function for the no-tax problem. Although the current wealth w_t is a state variable, it is well known that wealth can be factored out when the utility function belongs to CRRA class, i.e.

$$V_t^N(w_t, \mathbf{z}_t) = \frac{1}{1-\gamma} w_t^{1-\gamma} \phi_t(\mathbf{z}_t), \tag{2.2.15}$$

where $\phi_t(\mathbf{z}_t)$ is recursively defined with $\phi_T(\mathbf{z}_T) = 1$, and

$$\frac{1}{1-\gamma}\phi_t(\mathbf{z}_t) = \max_{\tilde{b}_t, \tilde{\mathbf{n}}_t} \mathbb{E}_t\left[\frac{1}{1-\gamma}(\tilde{b}_t r_0 + \mathbf{r}'_{t+1}\tilde{\mathbf{n}}_t)^{1-\gamma}\phi_{t+1}(\mathbf{z}_{t+1})\right] \tag{2.2.16}$$

$$\text{s.t.} \quad \tilde{b}_t + \mathbf{1}'\tilde{\mathbf{n}}_t = 1 \tag{2.2.17}$$

$$\tilde{b}_t \geq 0, \;\; \tilde{\mathbf{n}}_t \geq \mathbf{0}. \tag{2.2.18}$$

$\tilde{b}_t$ and $\tilde{\mathbf{n}}_t$ denote the post-trade *fractions* of wealth w_t invested in the cash account and risky securities, respectively. The time-t conditional expectation in (2.2.16) is taken over the next period's return vector $\mathbf{r}_{t+1}$ and the state vector, $\mathbf{z}_{t+1}$. The budget constraint is given by (2.2.17) and (2.2.18) are the no borrowing and no short-sales constraints respectively. In this case, the dimension of the state space is equal to the dimension of the state vector $\mathbf{z}_t$ and so we can solve V_t^N numerically over a fine grid of possible values of $\mathbf{z}_t$ if the dimension of $\mathbf{z}_t$ is sufficiently small. We can then use V_t^N as an approximate value function for the look-ahead policy of Section 2.3.4, and also as a basis for constructing dual-feasible penalties in Section 2.4.

The following lemma compares the set of feasible policies for the exact tax basis LUL problem to those for the no-tax problem.

Lemma 2.2.1. *The set of feasible policies for the exact tax-basis LUL problem is a subset of the set of feasible policies for the no-tax problem. Therefore,* $V_t^N \geq V_t^{LUL}$.

Proof. Suppose taxes in the LUL problem are not paid to the tax authority; instead they are invested in a special risk-free security that also earns the same risk-free rate r_0. The proceeds of this investment are made available to the agent at maturity T. In this case we see that any feasible policy for this adjusted exact tax-basis LUL problem is also feasible for the no-tax problem. In particular the optimal value function for this adjusted problem $V_t^{LUL,\, adj}$ satisfies $V_t^{LUL,\, adj} \leq V_t^N$. But we clearly have $V_t^{LUL} \leq V_t^{LUL,adj}$ and so the result follows. □

Note that the result in Lemma 2.2.1 does *not* hold in general for FUL problems.

2.3 Feasible Investment Policies

In this section we discuss several methods for constructing feasible policies for the exact tax-basis LUL problem that we formulated in Section 2.2. We note that each of these policies can be simulated so that it is straightforward to obtain unbiased estimates of their performances. Given that these policies are feasible, it is clear that these estimates are lower or *primal* bounds on the optimal value function, V_0^{LUL}. We begin with policies based on the solution to the no-tax problem.

2.3.1 Tax-Blind Policies

Under the tax-blind policy, the agent simply re-balances her portfolio in every period so that her after-tax portfolio agrees with[4] the optimal portfolio of the no-tax model. In order to implement this policy, we first define a sufficiently fine grid of points representing the support of $\mathbf{z}_t$. We then solve the DP (2.2.16) numerically on this grid and obtain the optimal post-trade asset fractions $(\tilde{b}_t(\mathbf{z}_t), \tilde{\mathbf{n}}_t(\mathbf{z}_t))$, for each time t and each grid value, $\mathbf{z}_t$. We only need to solve the no-tax problem once and then store the solution for future use.

The fractions $(\tilde{b}_t(\mathbf{z}_t), \tilde{\mathbf{n}}_t(\mathbf{z}_t))$ for the realized state $\mathbf{z}_t$ are computed using linear interpolation. If she needs to sell shares in a certain stock, we assume she does so in equal proportions from each holding or tax-basis of that security. These trades are transacted under the LUL assumption so that taxes are paid at each time t if there are net realized capital gains at that time after offsetting by any accumulated unused losses. Note that trading this way is more efficient (and therefore more realistic) than assuming the portfolio is liquidated at each time t before paying any taxes due and then trading to the no-tax solution.

While typically ruled out in practice, it is common in the literature to allow so-called

[4]We say two portfolios "agree" if they both hold the same *fractions* of wealth in each security.

wash-sales where the agent is allowed to sell and immediately repurchase shares in any stock. This is advantageous for a portfolio with embedded capital losses because the wash sales transaction realizes the capital losses so they are immediately available for offsetting any realized gains. We will refer to the tax-blind policy which always avails of wash-sales as the *modified tax-blind* policy. Note that we allow wash-sales in the rolling buy-and-hold and look-ahead policies of Sections 2.3.3 and 2.3.4.

2.3.2 The Buy-and-Hold Policy

Under the buy-and-hold policy, the investor can trade at time $t = 0$, but is not allowed to trade again until time T when her portfolio is liquidated and capital gain taxes, τg_T, are paid. Therefore, the agent's portfolio selection problem at $t = 0$ is given by

$$\begin{aligned} \max_{b_0,\ \mathbf{n}_{0,0}} \quad & \mathbb{E}_0\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \\ \text{s.t.} \quad & b_0 + \mathbf{p}_0'\mathbf{n}_{0,0} = w_0 \\ & b_0 \geq 0 \\ & \mathbf{n}_{0,0} \geq \mathbf{0} \\ & b_T = b_0 r_0^T + \mathbf{p}_T'\mathbf{n}_{0,0} - \tau g_T \\ & g_T \geq (\mathbf{p}_T - \mathbf{p}_0)'\mathbf{n}_{0,0} \\ & g_T \geq 0 \end{aligned} \tag{2.3.1}$$

where the last two constraints ensure the LUL tax rule is observed.

2.3.3 The Rolling Buy-and-Hold Policies

In contrast to the tax-blind policy, the buy-and-hold portfolio is chosen taking taxes into account but it cannot react to changes in the market conditions by re-balancing the portfolio during the investment horizon. The rolling buy-and-hold (RBH) policy can partly overcome these shortcomings. At each time t, the agent following the RBH policy assumes that she can trade at time t, but subsequently that she will not be allowed to trade again until time T when her portfolio will be liquidated. Therefore, at time t, she selects her portfolio by solving the problem

$$\begin{aligned} \max_{\mathbf{x}_t, b_T, g_T} \quad & \mathbb{E}_t\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \\ \text{s.t.} \quad & \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t}) \\ & b_T = b_t r_0^{T-t} + \sum_{j=0}^{t} \mathbf{p}_T' \mathbf{n}_{j,t} - \tau g_T \\ & g_T \geq \sum_{j=0}^{t} (\mathbf{p}_T - \mathbf{p}_j)' \mathbf{n}_{j,t} + l_t \\ & g_T \geq 0 \end{aligned} \tag{2.3.2}$$

where $\mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$ denotes the set of feasible trades at time t.

We obtain an unbiased estimate for the value function of the RBH policy by averaging the performance over I sample paths. Over each path i and each time instant t along the path, we need to compute the optimal trade by solving (2.3.2). The expectation in the objective function of (2.3.2) cannot be evaluated analytically so instead we approximate it by taking samples. In particular, we use low-discrepancy sequences (LDS) and quasi-Monte-Carlo to generate M scenarios $\mathbf{p}_T^{(m)}, m = 1, ..., M$ for the random time T security prices.[5] At time t, we choose the RBH trades by solving the following approximation to (2.3.2):

$$\max_{\mathbf{x}_t,\ b_T^{(m)},\ g_T^{(m)}} \quad \frac{1}{M}\sum_{m=1}^{M}\frac{(b_T^{(m)})^{1-\gamma}}{1-\gamma} \tag{2.3.3}$$

$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$$

$$b_T^{(m)} = b_t r_0^{T-t} + \sum_{j=0}^{t} \mathbf{p}_T^{(m)\prime} \mathbf{n}_{j,t} - \tau g_T^{(m)} \qquad 1 \le m \le M \tag{2.3.4}$$

$$g_T^{(m)} \ge \sum_{j=0}^{t} (\mathbf{p}_T^{(m)} - \mathbf{p}_j)' \mathbf{n}_{j,t} + l_t \qquad 1 \le m \le M \tag{2.3.5}$$

$$g_T^{(m)} \ge 0 \qquad 1 \le m \le M \tag{2.3.6}$$

where $\tau g_T^{(m)}$ is the capital gains tax paid after liquidating the portfolio at time T in the m-th scenario and $b_T^{(m)}$ is the resulting after-tax cash. The optimization problem (2.3.3) has $2M + K(t+1) + 3$ decision variables and $2M + 2$ constraints [6]. Since the objective function (2.3.3) is concave and all constraints are linear, (2.3.3) is a convex optimization problem, and in principle can be solved using standard solvers.[7]

We need the number of scenarios M for the time T security prices to be large in order to adequately approximate the expectation in (2.3.2). In the case of 20 risky securities and 20 time periods we chose $M = 100,000$ scenarios. MOSEK then required 2.5 minutes[8] to solve

[5]Further details are in Appendix A.1.

[6]We include the main constraints (2.2.7), (2.2.11), (2.3.4) and (2.3.5) in our constraint count but exclude the non-negativity constraints (2.2.8)–(2.2.10), (2.2.12), (2.2.13) and (2.3.6) since these latter constraints have little impact on the solution time. Note that constraints (2.2.7)–(2.2.13) are implicitly present in $\mathbb{X}_t$.

[7]When γ is rational, (2.3.3) can be reformulated as a second-order cone program [3].

[8]The computations were done with Matlab using the convex non-linear optimization in MOSEK on a Windows 7 computer with 3.6GHz Intel i7-4770 4 cores CPU and 16 GB of RAM. We used this computer for all computations reported in this chapter.

(2.3.3) for all trading periods on a given sample path. If one uses $N = 5,000$ sample paths to estimate the value of the RBH policy, the total running time is then approximately 9 days! It is clear we need an alternative approach to solving the approximate RBH problem.

To find such an approach, recall that constraints (2.3.5) and (2.3.6) formulate the LUL tax rule as

$$g_T^{(m)}(\mathbf{x}_t) = \max\left\{ \sum_{j=0}^{t} (\mathbf{p}_T^{(m)} - \mathbf{p}_j)'\mathbf{n}_{j,t} + l_t, 0 \right\}. \tag{2.3.7}$$

We can smooth the max function in (2.3.7) and approximate the LUL function $g_T^{(m)}$ by the function

$$\tilde{g}_T^{(m)}(\mathbf{x}_t) = \frac{1}{\theta} \ln\left(\exp\left(\theta\bigg(\sum_{j=0}^{t} (\mathbf{p}_T^{(m)} - \mathbf{p}_j)'\mathbf{n}_{j,t} + l_t \bigg) \right) + 1 \right) \tag{2.3.8}$$

where $\theta > 0$ is a fixed scalar. Note that the function $\tilde{g}_T^{(m)}(\cdot)$ is always non-negative, and is therefore consistent with the no tax-rebate property of the LUL tax rule. In addition, it is easy to check that $\tilde{g}_T^{(m)}(\mathbf{x})$ converges to $g_T^{(m)}(\mathbf{x})$ as θ goes to infinity. Therefore, $\tilde{g}_T^{(m)}(\cdot)$ serves as a good approximation to $g_T^{(m)}(\cdot)$ for sufficiently large θ. Our motivation for using $\tilde{g}_T^{(m)}(\cdot)$ instead of $g_T^{(m)}(\cdot)$ is twofold. First, we have analytical expressions for the first and second order derivatives of $\tilde{g}_T^{(m)}$ that can be passed to the convex optimization solver. Second, we can significantly reduce the problem size by expressing $b_T^{(m)}$ as a function of $\mathbf{x}_t$ using $\tilde{g}_T^{(m)}(\mathbf{x}_t)$ and (2.3.4), and then eliminating the constraints (2.3.4), (2.3.5) and (2.3.6). The resulting optimization problem[9] is

$$\max_{\mathbf{x}_t} \quad \frac{1}{M(1-\gamma)}\sum_{m=1}^{M}\left(b_t r_0^{T-t} + \sum_{j=0}^{t}\mathbf{p}_T^{(m)\prime}\mathbf{n}_{j,t} - \tau\tilde{g}_T^{(m)}(\mathbf{x}_t)\right)^{1-\gamma} \tag{2.3.9}$$
$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$$

where there are now a total of $K(t+1)+2$ variables and only 2 constraints (2.2.7) and (2.2.11) (implicitly present in $\mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$[10]. Note that the size of the problem (2.3.9) is now independent of M.

Let $f_t(\mathbf{x}_t)$ denote the objective function in (2.3.9). We solve it using the following sequential quadratic programming (SQP) approach.

1. Choose a starting point, $\bar{\mathbf{x}}_t = [\bar{b}_t \ \bar{\mathbf{n}}_{0,t}' \ ... \ \bar{\mathbf{n}}_{t,t}' \ \bar{g}_t \ \bar{l}_t]$.

2. Approximate $f_t(\mathbf{x}_t)$ with a second order Taylor expansion $\bar{f}_t(\mathbf{x}_t)$ about $\bar{\mathbf{x}}_t$, and solve

$$\max_{\mathbf{x}_t} \quad \bar{f}_t(\mathbf{x}_t) := f_t(\bar{\mathbf{x}}_t) + \nabla f_t(\bar{\mathbf{x}}_t)'(\mathbf{x}_t - \bar{\mathbf{x}}_t) + \frac{1}{2}(\mathbf{x}_t - \bar{\mathbf{x}}_t)'\nabla^2 f_t(\bar{\mathbf{x}}_t)(\mathbf{x}_t - \bar{\mathbf{x}}_t) \tag{2.3.10}$$
$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$$

where $\nabla f_t(\bar{\mathbf{x}}_t)$ and $\nabla^2 f_t(\bar{\mathbf{x}}_t)$ are the gradient vector and Hessian matrix of f_t, respectively, evaluated at $\bar{\mathbf{x}}_t$. Recall that we have analytical expressions for both $\nabla f_t(\bar{\mathbf{x}}_t)$ and $\nabla^2 f_t(\bar{\mathbf{x}}_t)$. Let $\mathbf{x}_t^{opt}$ be the optimal solution to (2.3.10).

[9] We note that the non-linear function $\tilde{g}_T^{(m)}(\mathbf{x}_t)$ in (2.3.9) implies that the problem (2.3.9) cannot be formulated as a second-order cone program.

[10] Again not counting the non-negativity constraints.

3. Evaluate the objective function f_t at $\mathbf{x}_t^{opt}$, and stop if we have converged to within a given error tolerance. Otherwise set $\bar{\mathbf{x}}_t = \mathbf{x}_t^{opt}$ and return to step 2.

We used an absolute error tolerance of 10^{-5} in our SQP algorithm. Depending on the level of risk aversion, γ, this corresponds to a relative error tolerance between 10^{-5} and 10^{-6}. We typically found that the SQP approach converged after only two or three iterations. The time required to solve the quadratic programming problem (2.3.10) is significantly smaller than that required to solve the general convex programming problem (2.3.3). Returning to our earlier example of 20 stocks and 20 time periods with $M = 100,000$ scenarios, the SQP approach requires approximately 3 seconds to solve all 20 problems on one sample path[11]; computing the RBH policy along 5,000 paths to estimate the value function therefore takes approximately 4 hours. This corresponds to approximately a 98% improvement in the solution time.

Note that the optimization problem (2.3.1) required for the buy-and-hold policy is a special case of the RBH problem and so it can be solved using the SQP approach described here.

2.3.4 Look-Ahead Policies

We also considered an *h-step look-ahead* policy where the investor selects the time t trades assuming that she will next trade at time $t + h$, and approximates the continuation value at time $t + h$ by the no-tax model's value function V_t^N. Thus, the time t trades are computed

[11]Calculations are done with MOSEK's QP solver called from Matlab.

by solving the optimization problem

$$\max_{\mathbf{x}_t, w_{t+h}, g_{t+h}} \quad \mathbb{E}_t\Big[V_{t+h}^N(w_{t+h}, \mathbf{z}_{t+h})\Big] \tag{2.3.11}$$

$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$$

$$w_{t+h} = b_t r_0^h + \sum_{j=0}^{t} \mathbf{p}'_{t+h}\mathbf{n}_{j,t} - \tau g_{t+h} \tag{2.3.12}$$

$$g_{t+h} \geq \sum_{j=0}^{t} (\mathbf{p}_{t+h} - \mathbf{p}_j)'\mathbf{n}_{j,t} + l_t \tag{2.3.13}$$

$$g_{t+h} \geq 0. \tag{2.3.14}$$

Note that we use the terminal utility function V_T^N in place of V_{t+h}^N whenever $t + h > T$. The time t conditional expectation in (2.3.11) is taken over the stock prices $\mathbf{p}_{t+h}$ and state vector, $\mathbf{z}_{t+h}$, at time $t + h$. The performance of the h-step look-ahead is clearly a function of h. In our numerical experiments, we found that taking $h = 6$ yielded the best results.

We solve for the h-step look-ahead policy using a technique similar to the one we used to solve for the RBH policy. We approximate the conditional expectation in (2.3.11) using M scenarios for the vector of stock prices $\mathbf{p}_{t+h}^{(m)}$ and state vector $\mathbf{z}_{t+h}^{(m)}$ conditional on the information at t. Thus, our approximation for (2.3.11) is given by

$$\max_{\mathbf{x}_t, w_{t+h}^{(m)}, g_{t+h}^{(m)}} \quad \frac{1}{M} \sum_{m=1}^{M} \frac{\left(w_{t+h}^{(m)}\right)^{1-\gamma} \phi_{t+h}\left(\mathbf{z}_{t+h}^{(m)}\right)}{1-\gamma} \tag{2.3.15}$$

$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t})$$

$$w_{t+h}^{(m)} = b_t r_0^h + \sum_{j=0}^{t} \mathbf{p}_{t+h}^{(m)\prime} \mathbf{n}_{j,t} - \tau g_{t+h}^{(m)} \qquad 1 \leq m \leq M \tag{2.3.16}$$

$$g_{t+h}^{(m)} \geq \sum_{j=0}^{t} (\mathbf{p}_{t+h}^{(m)} - \mathbf{p}_j)' \mathbf{n}_{j,t} + l_t \qquad 1 \leq m \leq M \tag{2.3.17}$$

$$g_{t+h}^{(m)} \geq 0 \qquad 1 \leq m \leq M \tag{2.3.18}$$

where we use the fact that $V^N(w,z) = \frac{1}{1-\gamma} w^{(1-\gamma)} \phi(\mathbf{z})$ (see (2.2.15)). Recall that we pre-compute the function $\phi_{t+h}(\cdot)$ on a grid in advance and then use linear interpolation to evaluate $\phi_{t+h}(\mathbf{z})$ for values of $\mathbf{z}$ that are not on the grid. The optimization problem (2.3.15) has the same structure as the optimization problem (2.3.3) for rolling RBH policy and so we can employ the same smoothing techniques and SQP approach to solve it efficiently.

2.4 Evaluating Sub-Optimal Policies

The performance of any feasible policy for the exact tax-basis LUL problem in (2.2.5) is clearly a lower bound on the optimal value function V_0^{LUL}. However, since V_0^{LUL} is not computable, we cannot assess the quality of any feasible policy. We can overcome this by computing valid upper bounds for the value function. We showed in Section 2.2 that the no-tax value function V^N is an upper bound for V^{LUL}; however, this bound is typically very weak, and therefore provides relatively little information about the quality of the sub-optimal feasible policies. In this section we show how to construct tighter upper (or dual) bounds using the recently developed information relaxations approach of Rogers [60], and in particular, BSS [16].

2.4.1 Dual Problem Formulation

Let $\mathbf{p} := [\mathbf{p}_0' \ ... \ \mathbf{p}_T']'$, $\mathbf{z} := [\mathbf{z}_0' \ ... \ \mathbf{z}_T']'$ and $\mathbf{x} := [\mathbf{x}_0' \ ... \ \mathbf{x}_T']'$ denote the entire sequence of security prices, market states and trade decision vectors, respectively. A trading policy can be interpreted as a function $\mathbf{x}(\mathbf{p}, \mathbf{z})$ that maps sequences of prices $\mathbf{p}$ and market states $\mathbf{z}$ to a sequence of trading decisions $\mathbf{x}$. A feasible trading policy is one where each individual decision $\mathbf{x}_t$ is in the set of feasible trades, $\mathbb{X}_t$. A feasible adapted policy is a feasible policy that is $\mathcal{F}_t$-adapted. Let $\mathcal{X}^L$ denote the set of all feasible adapted policies for the exact tax-basis problem. The agent wants to compute a feasible adapted policy that maximizes the expected utility of the after-tax terminal wealth. Thus, the optimization problem (2.2.5) can be reformulated as

$$V_0^{LUL} = \max_{\mathbf{x} \in \mathcal{X}^L} \left\{ \mathbb{E}_0 \left[\frac{b_T^{1-\gamma}}{1-\gamma} \right] \right\}. \tag{2.4.1}$$

We will call a function $\pi(\mathbf{x})$ a *dual feasible penalty* if $\mathbb{E}_0[\pi(\mathbf{x}(\mathbf{p}, \mathbf{z}))] \leq 0$ for all feasible $\mathcal{F}_t$-adapted policies, $\mathbf{x}(\mathbf{p}, \mathbf{z}) \in \mathcal{X}^L$. It is then clear that

$$V_0^{LUL} \leq \max_{\mathbf{x} \in \mathcal{X}^L} \left\{ \mathbb{E}_0 \left[\frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) \right] \right\}.$$

Suppose at time $t = 0$ the investor has *perfect information* of all future security prices and market states before she makes any trading decisions. It then follows that

$$V_0^{LUL} \leq \max_{\mathbf{x} \in \mathcal{X}^L} \left\{ \mathbb{E}_0 \left[\frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) \right] \right\} \leq \mathbb{E}_0 \left[\max_{\mathbf{x} \in \mathbb{X}} \left\{ \frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) \right\} \right] =: V_{up} \qquad (2.4.2)$$

where $\mathbb{X}$ denotes the set of feasible trades under the assumption of perfect information. BSS [16] and Rogers [60] establish that strong duality holds in (2.4.2), i.e. there exists dual-feasible penalties for which $V_{up} = V_0^{LUL}$. We defer the discussion of our particular choice for the dual penalty $\pi(\mathbf{x})$ until Section 2.4.2; we only note at this point that we will only ever need to consider penalties that are linear in the actions $\mathbf{x}$.

The dual bound of (2.4.2) leads itself to estimation via Monte-Carlo simulation: we simply simulate I sample paths of the security prices and market states, and on each path we solve the maximization problem inside the expectation in (2.4.2). If $V_{up}^{(i)}$ is the optimal solution of the problem on the i-th path, then $\sum_{i=1}^{I} V_{up}^{(i)}/I$ is an unbiased estimate of V_{up}. Moreover, we can estimate V_{up} to any desired accuracy by choosing a large enough number of paths, I.

Suppose now we have simulated one sample path for the security prices and state variables. The inner maximization problem in (2.4.2) is a deterministic optimization problem. Since $b^{(1-\gamma)}/(1-\gamma)$ is concave and we only consider dual penalties $\pi(\mathbf{x})$ that are linear in $\mathbf{x}$, the inner problem is a convex optimization problem with linear constraints. Moreover, we can significantly reduce the number of constraints in (2.2.9) via the variable transformation $\hat{\mathbf{n}}_{j,t} := \mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}$ so that $\hat{\mathbf{n}}_{j,t}$ denotes the number of shares with tax basis $\mathbf{p}_j$ sold at time

t. The inner maximization problem (2.4.2) then takes the form

$$\max_{b_t,\ \mathbf{n}_{t,t},\ \hat{\mathbf{n}}_{j,t},\ g_t,\ l_t} \quad \frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) \tag{2.4.3}$$

$$\begin{aligned}
\text{s.t.} \quad & b_0 + \mathbf{p}_0' \mathbf{n}_{0,0} = w_0 & & \\
& b_t + \mathbf{p}_t' \mathbf{n}_{t,t} + \tau g_t = b_{t-1} r_0 + \sum_{j=0}^{t-1} \mathbf{p}_t' \hat{\mathbf{n}}_{j,t} & t \geq 1 & \quad (2.4.4) \\
& \mathbf{n}_{t,t} - \sum_{j=t+1}^{T} \hat{\mathbf{n}}_{t,j} = \mathbf{0} & t \geq 0 & \\
& g_t + l_t = \sum_{j=0}^{t-1} (\mathbf{p}_t - \mathbf{p}_j)' \hat{\mathbf{n}}_{j,t} + l_{t-1} & t \geq 1 & \\
& \mathbf{n}_{t,t} \geq \mathbf{0} & t \geq 0 & \\
& \hat{\mathbf{n}}_{j,t} \geq \mathbf{0} & t \geq 1,\ j < t & \\
& b_t \geq 0 & t \geq 0 & \\
& g_t \geq 0 & t \geq 1 & \\
& l_t \leq 0,\ l_0 = 1, & t \geq 1. &
\end{aligned}$$

Note that there are $(T+1)(KT+2K+2)/2 + 2T + 1$ variables and $T(K+2)+1$ constraints in this dual *inner* problem. Moreover, the constraints in the optimization problem (2.4.3) are quite sparse; only a small subset of the decision variables appears in each of the constraints, and the objective function in (2.4.3) is separable. Most convex optimization algorithms, including the MOSEK non-linear convex optimization solver that we used, can take advantage of these two properties.

For our 20 stock and 20 time period example, each dual problem instance has $4,682$

decision variables and 441 constraints, and it can be solved in less than 0.03 seconds. Given that the dual problem size only grows quadratically in K and T, we can easily solve dual problem instances with a much larger number of securities and time periods. This is perhaps a surprising observation given the reputation of the (primal) tax-problem for being an extremely challenging problem to solve.

2.4.2 Dual Penalties

For our LUL exact tax-basis problem we consider a gradient penalty of the form originally proposed by Brown and Smith [15]. Recall that we use $\mathcal{X}^L$ to denote the set of feasible adapted policies for the exact tax-basis LUL problem. We also let $\mathcal{X}^N$ denote the set of feasible adapted policies for the no-tax problem. Given a sample path of security prices $\mathbf{p}$ and market states $\mathbf{z}$, let $\tilde{\mathbf{x}}^*(\mathbf{p}, \mathbf{z})$ denote the optimal feasible adapted trading policy for the no-tax problem, and let $\tilde{w}_T(\tilde{\mathbf{x}}^*(\mathbf{p}, \mathbf{z}))$ denote the corresponding terminal wealth that results from following this policy. We define our gradient penalty as follows.

$$\pi(\mathbf{x}) := \nabla_{\mathbf{x}} U(\tilde{w}_T(\tilde{\mathbf{x}}^*(\mathbf{p}, \mathbf{z})))'(\mathbf{x} - \tilde{\mathbf{x}}^*(\mathbf{p}, \mathbf{z})) \tag{2.4.5}$$

where $\nabla_{\mathbf{x}} U(\tilde{w}_T(\mathbf{x}))$ denotes the gradient of the terminal utility with respect to $\mathbf{x}$. Note that $\pi(\mathbf{x})$ is clearly linear in the decision vector $\mathbf{x}$.

Since we can view the no-tax problem of (2.2.15) as a static convex optimization problem over the trading *policy* $\mathbf{x}$, the first order conditions imply that

$$\mathbb{E}\big[\nabla_{\mathbf{x}} U(\tilde{w}_T(\tilde{\mathbf{x}}^*(\mathbf{p},\mathbf{z})))'(\mathbf{x}-\tilde{\mathbf{x}}^*(\mathbf{p},\mathbf{z}))\big] \leq 0 \tag{2.4.6}$$

for any feasible adapted trading policy $\mathbf{x} \in \mathcal{X}^N$. From Lemma 2.2.1 it follows that $\mathcal{X}^L \subseteq \mathcal{X}^N$ so that (2.4.6) also holds for all $\mathbf{x} \in \mathcal{X}^L$. Thus, the gradient penalty (2.4.5) is dual feasible for the exact tax-basis LUL problem.

2.5 Numerical Experiments

We consider a stylized exact tax-basis LUL problem where the investor can invest in 20 risky securities as well as the cash account over a time horizon of $T = 20$ periods corresponding to a horizon of 20 years. Without loss of generality we assume an initial wealth of $w_0 = 1$, and we set tax rates of $\tau = 20\%$, 30% and 40%. We consider coefficients of relative risk aversion $\gamma \in \{1.5, 3, 5\}$. In Appendix A.3.4 we report numerical results for the LUL average-cost tax basis problem.

2.5.1 Security Price and Market State Dynamics

We assume $r_0 = 1.01$ so that the net annual after-tax risk-free rate is 1%. Following Lynch [52], we define the returns and market state variable dynamics as follows.

$$\ln \mathbf{r}_{t+1} = \boldsymbol{\alpha} + \boldsymbol{\beta} z_t + \tilde{\boldsymbol{\epsilon}}_{t+1} \tag{2.5.1}$$

$$z_{t+1} = \lambda z_t + \tilde{\eta}_{t+1} \tag{2.5.2}$$

where $\mathbf{r}_{t+1}$ denotes the return vector for the risky securities over the period $[t, t+1]$ and $\boldsymbol{\epsilon}_t$ are IID multivariate normal random vectors with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_\epsilon$. The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}_\epsilon$ are chosen so that all securities have a Sharp ratio of 0.2, and the expected annual returns are uniformly distributed between 2% and 10%. We assume security returns are equi-correlated with a correlation coefficient of either $\rho = 0.4$ or $\rho = 0.9$.

We assume z_t is 1-dimensional and without loss of generality, we assume that it has unit variance. The market-state-sensitivity parameter $\boldsymbol{\beta}$ is then in units of standard deviation of z_t. We set $\boldsymbol{\beta}$ in such a way that a value of $z_t = 1$ increases the annual expected return on each stock by 30%, e.g. from 10% per year to 13%. The random variable η_t in (2.5.2) is assumed to be IID Normal with mean 0 and variance $\sigma_\eta^2 = 1 - \lambda^2$. We set $\lambda = 0.6$.

2.5.2 Results

We simulated $I = 5,000$ sample paths and implemented all policies described in Section 2.3 on each sample path. The dual problem was also solved on the same 5,000 paths.

Let $\bar{u}$ denote the average utility of a feasible policy, or a dual bound on the optimal value. Since $\bar{u}$ is difficult to interpret, we instead report the annualized certainty equivalent return, $\mathrm{CE}(\bar{u})$, in our numerical results. $\mathrm{CE}(\bar{u})$ is defined as the constant annualized return, r_{ce}, that yields the same average utility, i.e. $\frac{1}{1-\gamma}\big(w_0(1+r_{ce})^T\big)^{1-\gamma} = \bar{u}$, or equivalently

$$\mathrm{CE}(\bar{u}) = \left(\frac{((1-\gamma)\bar{u})^{\frac{1}{1-\gamma}}}{w_0} \right)^{\frac{1}{T}} - 1. \tag{2.5.3}$$

We report the mean and 95% confidence intervals (CI) for the CE($\bar{u}$).

We used the realized utility of the no-tax model as a control variate in order to reduce the number of Monte-Carlo paths that we required for estimating accurate primal and dual bounds. In particular, we estimate the expected utility $\bar{u}$ as

$$\bar{u} = \frac{1}{I}\sum_{i=1}^{I}\left(U\big(b_T(\mathbf{x}(\mathbf{p}^{(i)},\mathbf{z}^{(i)}))\big) + \beta\left[V_0^N(w_0,\mathbf{z}_0) - U\big(\tilde{w}_T(\tilde{\mathbf{x}}(\mathbf{p}^{(i)},\mathbf{z}^{(i)}))\big)\right]\right) \tag{2.5.4}$$

where I is the number of Monte-Carlo paths, $\mathbf{p}^{(i)}$ and $\mathbf{z}^{(i)}$ are the sequences of security prices and market states on the i-th path, $\mathbf{x}(\mathbf{p}^{(i)},\mathbf{z}^{(i)})$ is the sequence of decisions made by the sub-optimal policy under consideration, $\tilde{\mathbf{x}}(\mathbf{p}^{(i)},\mathbf{z}^{(i)})$ is the optimal feasible adapted trading policy for the no-tax problem and V_0^N is the optimal expected utility for the no-tax problem.

Our experimental results are displayed in Tables 2.1, 2.2 and 2.3 corresponding to values of $\tau = 20\%$, 30% and 40%, respectively. The estimated CE returns for the tax-blind, modified tax-blind, buy-and-hold, RBH and h-step look-ahead policies are denoted by V^{tb}, V^{mth}, V^{bh}, V^{rbh} and V^{hl}, respectively. In the case of the h-step look-ahead policy we chose a value of $h = 6$. The dual bound, V_d^g, was obtained using the gradient penalty in (2.4.5) while V_0^N is the CE return of the no-tax problem and, as explained earlier, is also an upper bound on the true optimal CE return.

We note that the best primal bounds are obtained from the h-period look-ahead policy which outperforms the RBH policy by 1 to 5 basis-points per annum. The RBH policy is the second best of the policies under consideration and outperforms the other policies by as much as 15 basis points per annum or as little as 1 or 2 basis points. The modified

Table 2.1: Results for exact tax-basis LUL problem with tax rate $\tau = 20\%$

Parameters			Sub-Optimal Policies					Upper Bounds	
γ	ρ	(%)	V^{tb}	V^{mtb}	V^{bh}	V^{rbh}	V^{hl}	V_d^g	V_0^N
1.5	0.4	CE return	3.33	3.41	3.23	3.42	3.47	3.64	4.01
		95% C.I	(3.33, 3.34)	(3.41, 3.41)	(3.20, 3.27)	(3.40, 3.44)	(3.44, 3.48)	(3.63, 3.64)	
1.5	0.9	CE return	2.09	2.11	2.07	2.16	2.20	2.31	2.58
		95% C.I	(2.08, 2.09)	(2.11, 2.12)	(2.05, 2.09)	(2.15, 2.18)	(2.18, 2.21)	(2.31, 2.31)	
3	0.4	CE return	2.23	2.26	2.11	2.26	2.29	2.41	2.61
		95% C.I	(2.23, 2.24)	(2.26, 2.26)	(2.09, 2.15)	(2.24, 2.28)	(2.28, 2.31)	(2.41, 2.42)	
3	0.9	CE return	1.55	1.56	1.51	1.59	1.61	1.70	1.84
		95% C.I	(1.55, 1.55)	(1.56, 1.56)	(1.49, 1.52)	(1.58, 1.60)	(1.60, 1.61)	(1.70, 1.70)	
5	0.4	CE return	1.68	1.72	1.63	1.72	1.74	1.82	1.94
		95% C.I	(1.68, 1.69)	(1.72, 1.72)	(1.61, 1.64)	(1.71, 1.73)	(1.73, 1.75)	(1.82, 1.82)	
5	0.9	CE return	1.33	1.34	1.29	1.35	1.36	1.44	1.54
		95% C.I	(1.33, 1.34)	(1.34, 1.34)	(1.28, 1.31)	(1.34, 1.35)	(1.36, 1.37)	(1.44, 1.44)	

tax-blind policy in particular is comparable to the RBH policy when τ is just 20%. This is not surprising since we know that the tax-blind policies are optimal in the limit as τ goes to zero. It may be surprising that our best policy, i.e. the h-period look-ahead policy, only outperforms the tax-blind and buy-and-hold policies by 6 to 36 basis points per annum with the actual number depending on γ and τ. However, this can be explained by the fact that even these latter policies are optimal within their class, and therefore, expected to still perform reasonably well.

Turning to the dual bounds, we note that in all cases V_d^g is significantly superior to the no-tax bound, V_0^N. In fact the duality gap, i.e. the difference between the estimated best CE return (corresponding to the h-period lookahead policy) and V_d^g, ranges from just 8 to 22 basis points per annum. If the true optimal CE is approximately at the midpoint of the duality gap, we could conclude that the h-step look-ahead policy is only 4 to 11 basis points per annum from the optimal CE return, at least for the numerical experiments we have considered here.

Table 2.2: Results for exact tax-basis LUL problem with tax rate $\tau = 30\%$

Parameters			Sub-Optimal Policies					Upper Bounds	
γ	ρ	(%)	V^{tb}	V^{mtb}	V^{bh}	V^{rbh}	V^{hl}	V_d^g	V_0^N
1.5	0.4	CE return	2.96	3.10	2.99	3.17	3.21	3.43	4.01
		95% C.I	(2.96, 2.97)	(3.10, 3.11)	(2.95, 3.02)	(3.14, 3.19)	(3.18, 3.22)	(3.42, 3.44)	
1.5	0.9	CE return	1.82	1.84	1.89	2.00	2.02	2.19	2.58
		95% C.I	(1.82, 1.83)	(1.83, 1.84)	(1.87, 1.92)	(1.98, 2.01)	(2.00, 2.03)	(2.19, 2.19)	
3	0.4	CE return	2.03	2.09	1.98	2.12	2.15	2.31	2.61
		95% C.I	(2.03, 2.04)	(2.09, 2.10)	(2.95, 2.01)	(2.10, 2.14)	(2.13, 2.17)	(2.30, 2.32)	
3	0.9	CE return	1.41	1.42	1.42	1.49	1.51	1.63	1.84
		95% C.I	(1.41, 1.41)	(1.42, 1.42)	(1.41, 1.44)	(1.48, 1.50)	(1.50, 1.52)	(1.63, 1.63)	
5	0.4	CE return	1.58	1.60	1.55	1.62	1.65	1.76	1.94
		95% C.I	(1.57, 1.58)	(1.60, 1.61)	(1.53, 1.56)	(1.60, 1.64)	(1.63, 1.66)	(1.75, 1.76)	
5	0.9	CE return	1.23	1.24	1.24	1.29	1.31	1.40	1.54
		95% C.I	(1.23, 1.23)	(1.24, 1.24)	(1.23, 1.25)	(1.29, 1.30)	(1.30, 1.31)	(1.40, 1.40)	

Table 2.3: Results for exact tax-basis LUL problem with tax rate $\tau = 40\%$

Parameters			Sub-Optimal Policies					Upper Bounds	
γ	ρ	(%)	V^{tb}	V^{mtb}	V^{bh}	V^{rbh}	V^{hl}	V_d^g	V_0^N
1.5	0.4	CE return	2.58	2.75	2.72	2.90	2.94	3.21	4.01
		95% C.I	(2.58, 2.59)	(2.75, 2.76)	(2.68, 2.75)	(2.88 2.92)	(2.92, 2.96)	(3.21, 3.22)	
1.5	0.9	CE return	1.55	1.57	1.72	1.83	1.84	2.05	2.58
		95% C.I	(1.54, 1.55)	(1.56, 1.57)	(1.70, 1.75)	(1.81, 1.85)	(1.83, 1.85)	(2.05, 2.05)	
3	0.4	CE return	1.80	1.89	1.84	1.97	2.00	2.18	2.61
		95% C.I	(1.80, 1.81)	(1.89, 1.89)	(1.81, 1.86)	(1.95, 1.99)	(1.99, 2.01)	(2.18, 2.18)	
3	0.9	CE return	1.25	1.26	1.33	1.39	1.41	1.56	1.84
		95% C.I	(1.24, 1.25)	(1.25, 1.26)	(1.32, 1.35)	(1.38, 1.40)	(1.40, 1.41)	(1.56, 1.56)	
5	0.4	CE return	1.43	1.48	1.47	1.54	1.57	1.68	1.94
		95% C.I	(1.42, 1.43)	(1.47, 1.48)	(1.46, 1.49)	(1.53, 1.56)	(1.55, 1.58)	(1.68, 1.69)	
5	0.9	CE return	1.13	1.14	1.19	1.24	1.25	1.35	1.54
		95% C.I	(1.13, 1.13)	(1.14, 1.14)	(1.18, 1.20)	(1.23, 1.25)	(1.24, 1.25)	(1.35, 1.35)	

2.6 Conclusions and Further Research

In this chapter we have considered the challenging problem of tax-aware dynamic portfolio allocation. We have developed several sub-optimal trading policies for the exact tax-basis LUL problems and constructed lower and upper bounds on the certainty equivalent returns of these policies. It is clear that similar policies can also be constructed for other variations of these tax problems. Our principal contribution has been to demonstrate that much larger problems than previously considered can now be tackled through the use of sophisticated optimization techniques and duality methods based on information-relaxations. Moreover, the dual formulations of exact tax-basis problems are much easier to solve than the corresponding primal problems and it is quite straightforward to solve dual problems where the number of securities and time periods is much larger than in the problems we considered in Section 2.5. To the best of our knowledge we are also the first to successfully use these duality methods for tax-aware asset allocation. We also consider the relatively easier average tax-basis problem in Appendix A.3 and A.4 but note that dual problem instances in this case are non-convex. We propose solution approaches for these problem instances so that we can still obtain valid upper bounds.

There are several possible directions for future research. First, we would like to identify better primal policies. We believe that it will be possible to leverage recent work on sub-optimal control and approximate dynamic programming to construct near optimal policies. A second direction is to improve our understanding of the (approximately) optimal trading policies for these problems. In particular, when do these policies trade? Is there an easy-tocharacterize no-trade zone? How do the answers to these questions depend on the particular version of the tax problem that we are addressing?

A third direction concerns portfolio allocation problems with the average tax-basis. While the primal version of this problem is significantly easier (albeit still challenging) when compared to the exact tax-basis problems, the corresponding dual problem are non-convex, and therefore, very difficult to solve. We have outlined in Appendix A.3 an approaches for constructing valid bounds. We successfully tested one of these approaches using the `BARON` solver but further experimentation will be required before we can definitively conclude that this approach is viable for all average-cost tax basis dual problems.

In addition, there are other tax-related problems, e.g. tax-aware index tracking, that could also be solved using the approach outlined in this chapter.

Chapter 3

Dynamic Portfolio Execution with Transaction Cost

3.1 Introduction

We study a dynamic portfolio execution problem which incorporates risk aversion, stochastic return covariances, and temporary and permanent linear price impacts that are also stochastic and time-varying. We propose a variant of an open-loop feedback control (OLFC) policy for solving the resulting portfolio execution problem and use duality methods based on information relaxations to demonstrate how well this policy performs, at least for the parameter settings that we consider. We also show that it is straightforward to include a non-linear temporary price impact as well as return predictability. Our model therefore allows us to capture a much broader range of market features than can be captured by those models that insist on explicit calculation of the optimal policy.

In order to apply the aforementioned duality methods it is necessary to use an approximate value function to compute so-called *penalties* which penalize the decision-maker for violating the non-anticipativity constraints. These penalties, however, require us to compute conditional expectations of the approximate value function. Ideally we can compute these expectations analytically but in many circumstances this may not be possible. This does not limit the applicability of the dual methodology, however, since it is known that being able to compute an unbiased estimate of the conditional expectations is sufficient for computing valid dual bounds. We study the use of suitably randomized low-discrepancy sequences (LDS) to estimate these expectations *efficiently*, and show that using randomized LDS to estimate these expectations can still yield tight dual bounds and at only a very modest increase (approx 10%) in computational work.

The remainder of this Chapter is organized as follows. We define the basic portfolio execution problem and describe our model in Section 3.2. We also describe there the sub-optimal policies that we will consider, focusing in particular on a variant of an OLFC policy. In Section 3.3 we show how the duality theory of BSS [16] can be applied in the context of our portfolio execution problem. We present numerical results in Section 3.4 where we consider a portfolio of 50 stocks and 78 time periods, representing, as stated earlier, a trading frequency of once every 5 minutes. In Section 3.5 we consider how the dual methodology can be extended to other portfolio execution problems including the problem where there is also a non-linear temporary price impact. We also include numerical results comparing the performance of the dual bound when the penalties are computed analytically with the performance when the dual penalties are computed via an unbiased Monte-Carlo simulation. We conclude in Section 3.6. Appendix B.1 contains the various calculations and derivations required for implementing the OLFC and other policies. Appendix B.2 analyzes the particular dual problem arising from our portfolio execution application. Appendix B.4 contains additional calibration details for the numerical results of Section 3.4.

In Appendix B.3 we consider the model of Bertsimas, Hummel and Lo [10] (BHL) who considered the problem of an agent needing to purchase shares in multiple securities assuming a linear temporary price impact and return predictability. The presence of no-sales constraints, which in general are binding due to return predictability, implies that it is hard to solve for optimal policy even numerically. They proposed instead an OLFC policy and conjectured that such a policy should be close to optimal. Using their calibrated model parameters, we use the duality techniques to confirm that the OLFC policy is indeed very close to optimal. While this appendix is stand-alone we include it for two reasons: (i) it provides another set of numerical experiments (in a portfolio execution context) demonstrating the use of dual methods to confirm the conjecture that a sub-optimal policy is close to optimal and (ii) it provides a simple demonstration of how duality can be used to determine *in advance* whether or not a particular feature, in this case cross-price impacts, are worth accounting for in a portfolio execution policy. We believe the duality methodology is particularly suited for answering such *need-to-model* questions.

3.2 The Portfolio Execution Problem and Model Description

We now describe the basic portfolio execution problem of a possibly risk-averse agent who we assume needs to *purchase* a fixed number of shares in a fixed number of assets. We note that we could just as easily handle the problem where the agent needs to *sell* a portfolio of securities as well as the case where the agents needs to purchase one subset of securities and sell another subset.

We assume time is discrete and runs from $t = 0$ to $t = T$ for a total of $T+1$ time periods. There are n different assets that are traded in the market and the agent needs to purchase a fixed number of shares in each of the n assets between $t = 0$ and $t = T$. At time t the agent observes the non-impact asset price vector $\tilde{\mathbf{p}}_t = [\tilde{p}_t^{(1)} \ldots \tilde{p}_t^{(n)}]'$, and then determines the decision vector, $\mathbf{s}_t = [s_t^{(1)} \ldots s_t^{(n)}]'$, where $s_t^{(i)}$ is the number of shares of the i-th asset purchased at time t. We will also use $\mathbf{s}_{u:v}$ to denote the $n(v-u+1) \times 1$ vector of decision variables, $[\mathbf{s}'_u \ldots \mathbf{s}'_v]'$. However, we will simply write $\mathbf{s}$ for $\mathbf{s}_{0:T}$. We use $\mathbf{p}_t = [p_t^{(1)} \ldots p_t^{(n)}]'$ to denote the time t vector of transaction prices and note that in general, $\mathbf{p}_t \neq \tilde{\mathbf{p}}_t$ due to the market impact of trading. We let $\mathbf{w}_t = [w_t^{(1)} \ldots w_t^{(n)}]'$ where $w_t^{(i)}$ denotes the remaining number of shares in asset i that must be purchased in periods $t, \ldots, T$. We define the execution cost as the difference between the actual cost, $\sum_{t=0}^{T} \mathbf{p}'_t \mathbf{s}_t$, and the benchmark cost, $\tilde{\mathbf{p}}'_0 \mathbf{w}_0$, which would prevail if the agent could purchase everything at time $t = 0$ without any market impact. The execution cost is therefore given by

$$\sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t - \tilde{\mathbf{p}}_0' \mathbf{w}_0. \tag{3.2.1}$$

In order to model risk-aversion we assume that the agent has an exponential utility function, with parameter γ. We assume that $\gamma > 0$ to reflect the fact that we are defining utility over *costs* rather than *wealth* as is usually the case in portfolio optimization problems. The portfolio execution problem is then stated as

$$\min_{\mathbf{s}_t \in \mathcal{F}_t, t=0,\ldots,T} \quad \mathbb{E}_0 \left[\exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t \right) \right] \tag{3.2.2}$$

$$\text{s.t.} \quad \text{price and state variable dynamics}$$

$$\sum_{t=0}^{T} \mathbf{s}_t = \mathbf{w}_0 \tag{3.2.3}$$

$$\text{and any other constraints} \tag{3.2.4}$$

where we use $\{\mathcal{F}_t\}_{t=0,\ldots,T}$ to denote the filtration generated by the price vectors as well as any other state variables in the model. Note also that the objective function (3.2.2) does not include $\tilde{\mathbf{p}}_0' \mathbf{w}_0$ because it is a constant and therefore can be factored out. Several comments are in order:

(i) Risk-neutrality can easily be modeled by taking the limit in (3.2.2) as γ goes to 0.

(ii) Recall the *CARA* property of exponential utility which, in the context of a standard dynamic portfolio optimization problem, implies that the optimal dollar value invested in risky assets does *not depend* on the current level of wealth. This property is typically viewed as a serious weakness of exponential utility. We see no problem with this assumption in the context of portfolio execution problems, however, because such problems often have a time horizon of just a few hours or at most just a few days.

(iii) Notwithstanding the previous point, we only use exponential utility as a mechanism for trading off execution cost with execution risk. In particular, other utility functions could also be used although the solution approach of Appendices B.1.3 and B.1.4 would no longer be applicable.

3.2.1 Basic Model Description

Our main model allows for stochastic and time-varying price impacts as well as stochastic variance-covariance return dynamics. These are important features in practice but are generally ignored in the academic literature. While we suspect that superior or more sophisticated proprietary models may be used by some industry participants, our model is sufficiently rich to demonstrate the broad applicability of the dual methodology as outlined in Section 3.3.

We assume the non-impact price, $\tilde{\mathbf{p}}_t$, follows a random walk where the dollar return, $\mathbf{r}_{t+1}$, between times t and $t+1$, is normally distributed with mean $\mathbf{0}$ and conditional covariance matrix, $\mathbf{\Sigma}_t$. A trading volume of $\mathbf{s}_t$ incurs a permanent price impact of $\mathbf{A}_t\mathbf{s}_t$ and a temporary price impact of $\mathbf{B}_t\mathbf{s}_t$. This results in dynamics of the form

$$\mathbf{p}_t = \tilde{\mathbf{p}}_t + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t \tag{3.2.5}$$

$$\tilde{\mathbf{p}}_{t+1} = \tilde{\mathbf{p}}_t + \mathbf{r}_{t+1} + \mathbf{A}_t\mathbf{s}_t \tag{3.2.6}$$

$$\mathbf{r}_{t+1} = \mathbf{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_{t+1} \tag{3.2.7}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \mathbf{s}_t \tag{3.2.8}$$

for $t = 0, \ldots, T$ and with the understanding that $\mathbf{w}_{T+1} = \mathbf{0}$. The $\boldsymbol{\epsilon}_t$'s in (3.2.7) are IID standard normal random vectors.

In general we allow $\mathbf{\Sigma}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$ to be stochastic. This will allow us to model the tendency for markets to be more liquid around the open and close and less liquid in between. The assumption of a linear permanent price impact is commonly made in the academic literature where it has been justified by no-arbitrage considerations; see Huberman and Stanzl [44] and Gatheral [36]. Nonetheless we note that permanent price impact is generally assumed to be better approximated by a square-root function in practice. We note that Guéant [39] recently provided a justification for this without introducing arbitrage into the model. Later in Section 3.5 we will show that a non-linear temporary price impact can also be included in our model.

When the agent chooses $\mathbf{s}_t$ at time t, we assume that he knows $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{\Sigma}_t$ and $\tilde{\mathbf{p}}_t$. We note that it would also be straightforward to model the possibly more realistic situation where the agent is not assumed to know $\tilde{\mathbf{p}}_t$ when choosing $\mathbf{s}_t$. We will also assume that we can compute $\mathbb{E}_t[\mathbf{A}_j]$ and $\mathbb{E}_t[\mathbf{B}_j]$ for all $j \geq t$. In Section 3.4 we will describe the specific dynamics for $\mathbf{\Sigma}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$ that we used in our numerical experiments.

In practical applications the agent may also need to include various types of portfolio constraints including no-sales constraints

$$\mathbf{s}_t \geq \mathbf{0} \text{ for } t = 0, ..., T \tag{3.2.9}$$

as well as sector-balance constraints. These constraints are linear and therefore, as we shall see later, do not impact the convexity of the primal or dual problems that we consider. We do not consider portfolio balance constraints but note that it is generally straightforward to include them when attempting to construct good sub-optimal policies. We also note that the assumption of a risk-averse utility function implicitly incorporates some form of portfolio balance constraints.

3.2.2 Sub-Optimal Execution Policies

In general it is not possible to find the optimal policy for this portfolio execution problem and so instead we seek good sub-optimal policies. For the particular model of Section 3.2.1 we will consider three different policies:

The Simple Policy: Here the agent buys the same quantity of shares in each of the $T+1$ time periods so that $\mathbf{s}_t = \mathbf{w}_0/(T+1)$.

The Risk-Neutral Policy: In this case the agent ignores the *risk* of the execution costs and uses the trading policy which minimizes the expected execution cost. If the dynamics of $\mathbf{A}_t$ and $\mathbf{B}_t$ are sufficiently tractable then this trading policy can be determined via dynamic programming when there are no portfolio constraints. The details are in Appendix B.1.1. In Appendix B.1.2 we also show that the simple and risk-neutral policies coincide if $\mathbf{A}_t$ and $\mathbf{B}_t$ are martingales and if $\mathbf{A}_t$ and $\mathbf{B}_t$ are symmetric. We note that this condition is of course very unlikely to hold in practice.

The OLFC Policy: At each time t the agent assumes that $\mathbf{\Sigma}_t$ will remain constant thereafter in which case he assumes the return series, $\{\mathbf{r}_j\}_{j=t+1}^{T}$, is IID MVN $(\mathbf{0}, \mathbf{\Sigma}_t)$. The agent also assumes the price impact matrices evolve deterministically and replaces $\mathbf{A}_j$ and $\mathbf{B}_j$ with their time-t conditional expectations, $\mathbb{E}_t[\mathbf{A}_j]$ and $\mathbb{E}_t[\mathbf{B}_j]$, respectively, for $j \geq t$. The OLFC policy therefore solves the following problem at each time $t = 0, \ldots, T$:

$$
\begin{aligned}
\min_{\mathbf{s}_{t:T} \in \mathcal{F}_t} \quad & \mathbb{E}_t\left[\exp\left(\gamma \sum_{j=t}^{T} \mathbf{p}_j' \mathbf{s}_j\right)\right] && (3.2.10) \\
\text{s.t.} \quad & \mathbf{p}_j = \tilde{\mathbf{p}}_j + \mathbb{E}_t[\mathbf{A}_j]\mathbf{s}_j + \mathbb{E}_t[\mathbf{B}_j]\mathbf{s}_j \quad \text{for } j = t, \ldots, T \\
& \tilde{\mathbf{p}}_{j+1} = \tilde{\mathbf{p}}_j + \mathbf{r}_{j+1} + \mathbb{E}_t[\mathbf{A}_j]\mathbf{s}_j \quad \text{for } j = t, \ldots, T \\
& \mathbf{r}_j = \mathbf{\Sigma}_t^{1/2} \boldsymbol{\epsilon}_j \quad \text{for } j = t+1, \ldots, T \\
& \sum_{j=t}^{T} \mathbf{s}_j = \mathbf{w}_t \\
& \mathbf{s}_j \geq \mathbf{0}. \quad \text{for } j = t, \ldots, T
\end{aligned}
$$

Let V_t^{ol} denote the optimal value of (3.2.10) and let $\mathbf{s}_{t:T}^{ol,t} := [\mathbf{s}_t^{ol,t\,\prime} \ \ldots \ \mathbf{s}_T^{ol,t\,\prime}]'$ denote the corresponding optimal solution. The OLFC policy implements $\mathbf{s}_t^{ol,t}$ at time t and ignores $\mathbf{s}_{t+1}^{ol,t}, \ldots, \mathbf{s}_T^{ol,t}$. Further details on solving this problem are in Appendix B.1.3. We also note that if we remove the constraint $\mathbf{s}_j \geq \mathbf{0}$, then an analytic solution to the OLFC policy can be found very quickly via dynamic programming. This analytic solution will form the basis for constructing dual penalties as discussed in Section 3.3. Details on this solution can be found in Appendix B.1.4.

A Risk-Neutral OLFC Policy: As mentioned above, depending on the dynamics of $\mathbf{A}_t$ and $\mathbf{B}_t$, the Risk-Neutral policy may not be computable. In this case we can compute the risk-neutral OLFC policy by taking the limit of (3.2.10) as γ goes to 0. The details are in Appendix B.1.3.

3.3 Evaluating Sub-Optimal Policies

Given a feasible suboptimal policy to the portfolio execution problem in (3.2.2), we can construct an unbiased upper bound, V_{ub}, to the optimal value function, V^*, by simulating multiple paths of the policy and taking the sample average of the realized utility. We can also use dual methods to estimate a lower bound, V_{lb}. Appendix C contains a review of these dual methods that were originally developed independently by BSS [16] and Rogers [60]. In this section we state the dual problem formulation of (3.2.2) and briefly discuss the tractability of this dual problem and its implications for modeling various features of the portfolio execution problem.

Let $\tilde{V}_t(\mathbf{s}_{0:t-1})$ be some approximation to the time t optimal value function of the portfolio execution problem. While we typically consider a value function to be a function of the model's state variables, in our duality context it is much more convenient to explicitly recognize the dependence of $\tilde{V}_t$ only on $\mathbf{s}_{0:t-1}$. As reviewed in Appendix C, weak duality implies that

$$V_{lb} \;:=\; \mathbb{E}_0\left[\min_{\mathbf{s}\in\mathbb{S}}\left\{\exp\left(\gamma\sum_{t=0}^{T}\mathbf{p}_t'\mathbf{s}_t\right)+\sum_{t=0}^{T-1}\left(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})]-\tilde{V}_{t+1}(\mathbf{s}_{0:t})\right)\right\}\right] \qquad (3.3.1)$$

yields a lower bound on V^* where $\mathbb{S}$ is the decision space defined by constraints (3.2.3) and (3.2.4). Moreover, strong duality states that if we take $\tilde{V} = V^*$ then $V_{lb} = V^*$. This suggests that the closer $\tilde{V}_t(\mathbf{s}_{0:t-1})$ is to the optimal value function the tighter the dual bound will be. Note that it is the *shape* of $\tilde{V}$ that is important rather than the absolute level of $\tilde{V}$ because adding any constant to $\tilde{V}$ will have no impact on the dual bound as it will cancel out on the right-hand-side of (3.3.1). In general for a given $\tilde{V}_t$ we cannot compute V_{lb} in closed form but in principle, an unbiased estimate of it can be computed via Monte-Carlo: we simply simulate M paths of the exogenously specified noise processes and on each path we solve the minimization problem inside the expectation in (3.3.1). If $V_{lb}^{(i)}$ is the optimal solution of the problem on the i-th path, then $\sum_{i=1}^{M} V_{lb}^{(i)}/M$ is an unbiased estimate of V_{lb}.

Two key issues arise in applying the duality methodology. First, while $V_{lb}^{(i)}$ can be obtained as the solution of a deterministic dynamic program, solving this DP may, like the primal problem, also be difficult. Instead we prefer to solve it as a static optimization problem but to do this we would generally prefer this optimization problem to be convex.

Even if the term $\exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right)$ is convex in $\mathbf{s}$ there is no guarantee that the objective function will be convex because of the penalty term $\sum_{t=0}^{T-1} \left(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})\right)$. We can overcome this problem, however, by using an approximation $\tilde{V}_{t+1}$ that is linear in $\mathbf{s}_t$. It is important to note that using such an approximation still yields a valid dual bound. This approach was introduced by Brown and Smith [15] and further details are provided in Section 3.3.1 below.

The second issue that arises is the possibility of not being able to compute an analytic expression for $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})]$. Note that these conditional expectations are required to evaluate the objective function inside the expectation in (3.3.1). As discussed in Section 3.3.1 below, with our choice of $\tilde{V}_t$, we were able to compute these conditional expectations analytically. We will also see in Section 3.5, however, that even if we cannot compute these expectations in closed form then we can instead use unbiased estimates of them and still obtain valid dual bounds. In a series of numerical experiments we will see that the resulting dual bounds remain very tight and that they are not much more expensive to compute than the bounds we obtain when the conditional expectations are available in closed form. We conclude then that the second issue can also often be overcome.

3.3.1 The Dual Problem Formulation

Under the model assumptions of Section 3.2.1, we see from (3.3.1) that a dual problem instance takes the form

$$\begin{aligned} \min_{\mathbf{s}\in\mathbb{S}} \quad & \exp\Big(\gamma\sum_{t=0}^{T}\mathbf{p}_t'\mathbf{s}_t\Big) \ + \ \sum_{t=0}^{T-1}\Big(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})\Big) \\ \text{s.t.} \quad & \mathbf{p}_t = \tilde{\mathbf{p}}_t + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t \\ & \tilde{\mathbf{p}}_{t+1} = \tilde{\mathbf{p}}_t + \mathbf{r}_{t+1} + \mathbf{A}_t\mathbf{s}_t \end{aligned} \tag{3.3.2}$$

where the $\mathbf{r}_t$'s, $\mathbf{A}_t$'s and $\mathbf{B}_t$'s have been simulated according to the true model dynamics and are all known to the decision-maker, and $\mathbb{S}$ denotes the constraint region defined by (3.2.3) and (3.2.9). We would like to take $\tilde{V}_t$ to be a linearized version of V_t^{ol} as we expect that the OLFC policy will typically provide a good approximation to V_t^*. In this case, however, it is difficult to compute $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})]$. While we can overcome this problem using nested Monte-Carlos as discussed in Section 3.5, we prefer instead to use a *modified* version of V_t^{ol} which we denote by V_t^{mol}. The precise definition of V_t^{mol} can be found in Appendix B.2.2 but we note in particular that we can compute an analytic expression for $\mathbb{E}_t[V_{t+1}^{mol}(\mathbf{s}_{0:t})]$.

In order to define the particular choice of $\tilde{V}_t$ that we use in (3.3.2) we first define

$$\hat{V}_{t+1}(\mathbf{s}_{0:t}) := \exp\left(\gamma\sum_{j=0}^{t}\mathbf{p}_j'\mathbf{s}_j\right)\ V_{t+1}^{mol} \tag{3.3.3}$$

$$= \exp\left(\gamma\sum_{j=0}^{t}\mathbf{p}_j'\mathbf{s}_j + \frac{\gamma}{2}\mathbf{w}_{t+1}'\tilde{\mathbf{G}}_{t+1}\mathbf{w}_{t+1} + \big(\tilde{\mathbf{p}}_t + \mathbf{A}_t\mathbf{s}_t\big)'\mathbf{w}_{t+1}\right)\exp\left(\gamma\mathbf{r}_{t+1}'\mathbf{w}_{t+1}\right) \tag{3.3.4}$$

where we have used (B.2-6) to substitute for V_{t+1}^{mol} in (3.3.3) and then used (3.2.6) to obtain (3.3.4). We note that the dependence of $\hat{V}_{t+1}(\cdot)$ on $\mathbf{s}_{0:t}$ also appears via $\mathbf{w}_{t+1}$ since $\mathbf{w}_{t+1} = \mathbf{w}_0 - \sum_{j=0}^{t} \mathbf{s}_j$. Note also that we want to include the first term on the right-hand side of (3.3.3) as this term is present in the value function (see also the remark at the end of Appendix C) of any policy but we omitted it from (3.2.10) as it had no impact on the OLFC *policy* at time $t+1$.

The difficulty with taking $\tilde{V}_{t+1}(\mathbf{s}_{0:t}) = \hat{V}_{t+1}(\mathbf{s}_{0:t})$ as given by (3.3.4) is that the resulting objective function in (3.3.2) will not in general be convex, even if the first term in (3.3.2) is convex. Instead we linearize $\hat{V}_{t+1}$. Recall that $\mathbf{s}_{0:t} = [\mathbf{s}_0' \ldots \mathbf{s}_t']'$ denotes the $n(t+1) \times 1$ vector of decision variables corresponding to the first $t+1$ time periods and let $\tilde{\mathbf{s}}_{0:t}$ be a fixed $n(t+1) \times 1$ vector. We then take $\tilde{V}_{t+1}(\cdot)$ to be a first order Taylor expansion of $\hat{V}_{t+1}(\cdot)$ about $\tilde{\mathbf{s}}_{0:t}$. In particular, we take

$$\tilde{V}_{t+1}(\mathbf{s}_{0:t}) \ := \ \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t}) + \nabla \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})'(\mathbf{s}_{0:t} - \tilde{\mathbf{s}}_{0:t}) \tag{3.3.5}$$

and then use this in (3.3.2), noting that the linearity of $\tilde{V}_{t+1}$ in $\mathbf{s}_{0:t}$ will preserve convexity of the objective function in (3.3.2) if the exponential term there is itself convex.

The question that now arises is how to choose $\tilde{\mathbf{s}}_{0:t}$? Intuitively we would like $\tilde{\mathbf{s}}_{0:t}$ to be as close as possible to the true optimal trade sequence. But of course we don't know the optimal trade sequence, so instead we will take $\tilde{\mathbf{s}}_{0:t} = [\mathbf{s}_0^{ol,0\,\prime} \ldots \mathbf{s}_t^{ol,t\,\prime}]'$, the trade sequence from the unconstrained OLFC policy. This implies that $\tilde{\mathbf{s}}_{0:t}$ is path-dependent so that each dual problem instance will use a different $\tilde{\mathbf{s}}_{0:t}$ to construct the $\tilde{V}_{t+1}$. This requires no additional work, however, since the $\mathbf{s}_t^{ol,t}$'s will have already been computed when simulating the OLFC policy.

In order to compute $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] = \mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})] + \mathbb{E}_t[\nabla \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})'](\mathbf{s}_{0:t} - \tilde{\mathbf{s}}_{0:t})$ we need $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ and $\mathbb{E}_t[\nabla \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$ to solve the dual instance in (3.3.2). Using (3.3.4) we obtain

$$\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})] = \exp\left(\gamma \sum_{j=0}^{t} \mathbf{p}_j' \tilde{\mathbf{s}}_j + \frac{\gamma}{2} \tilde{\mathbf{w}}_{t+1}' \tilde{\mathbf{G}}_{t+1} \tilde{\mathbf{w}}_{t+1} + \left(\tilde{\mathbf{p}}_t + \mathbf{A}_t \tilde{\mathbf{s}}_t\right)' \tilde{\mathbf{w}}_{t+1}\right) \mathbb{E}_t\left[\exp\left(\gamma \mathbf{r}_{t+1}' \tilde{\mathbf{w}}_{t+1}\right)\right] \tag{3.3.6}$$

where $\tilde{\mathbf{w}}_{t+1} = \mathbf{w}_0 - \sum_{j=0}^{t} \tilde{\mathbf{s}}_j$. The first exponential term on the right-hand-side of (3.3.6) is $\mathcal{F}_t$-measurable (since the OLFC policy is $\mathcal{F}_t$-adapted and $\tilde{\mathbf{G}}_{t+1}$ is $\mathcal{F}_t$-measurable) and conditional on $\mathcal{F}_t$, $\mathbf{r}_{t+1}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_t$. We can therefore compute $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ analytically and, while the calculations are somewhat tedious, we can also do the same for $\mathbb{E}_t[\nabla \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$. This means in particular that we can compute an analytic expression for the dual penalty term, $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})$.

3.3.2 Dual Convexity

The exponential term in (3.3.2) is studied in Appendix B.2 where we establish sufficient conditions on the $\mathbf{A}_t$'s and $\mathbf{B}_t$'s that guarantee its convexity for all dual problem instances. For example, we show that if $\mathbf{A}_t = \mathbf{A}$, a constant matrix, and that $\mathbf{A} + \mathbf{A}'$ is positive definite and $\mathbf{B}_t + \mathbf{B}_t'$ is positive semi-definite for all t, then every dual problem instance will be convex. These conditions guarantee the positive definiteness of $\mathbf{A}_t + \mathbf{A}_t' + \mathbf{B}_t + \mathbf{B}_t'$ which seems reasonable from an economic standpoint.

If the $\mathbf{A}_t$'s are stochastic, however, then economic considerations may still be employed to impose some structure on their dynamics. For example, it seems reasonable to assume that there should always exists a solution to the OLFC problem. Otherwise the OLFC decision-maker would believe (possibly incorrectly admittedly since he makes simplifying assumptions on the model dynamics) that arbitrage opportunities exist. This seems very unlikely in practice and so we could therefore insist on dynamics for the $\mathbf{A}_t$'s that guarantee the positive-definiteness of the $\mathbf{Q}_{OLFC,t}$'s as defined in (B.1-19) of Appendix B.1.3

More generally, however, it is not clear that we can use economic considerations to justify the convexity of each dual problem instance when the $\mathbf{A}_t$'s are stochastic. After all, consider the example where there is no temporary price impact and the agent needs to purchase shares in just one security over the course of a day. If the $\mathbf{A}_t$'s are stochastic it is possible on some given realization of the model uncertainty over the course of a day, that the price impact is very large in the early periods and approximately zero in the later periods. In the corresponding dual problem instance (with zero dual penalties, say) this could result in arbitrage profits since the agent could drive up the stock price over the course of the day by purchasing many shares when the price impact is large and then selling them all at the end of the day when the price impact is small. The presence of such arbitrage-profits would imply the non-convexity of the corresponding dual problem instance. The primal problem could still be well-posed, however, since the decision-maker does not get to see the realization of the day's uncertainty at time $t = 0$ and therefore would not be in a position to profit from the aforementioned strategy.

In general then, it seems possible for a plausible model to result in some dual problem instances that are non-convex. In that case we would need to bound the optimal objective function of these particular dual problems in order to still obtain a valid dual bound. This should be possible in the presence of portfolio constraints such as no-sales constraints, since in that case every dual problem would have a compact decision-space and therefore have a bounded optimal solution. Nonetheless, this would require more work as we would need to check for convexity of each dual problem and then solve to optimality those problems that are convex and bound the optimal solution to the non-convex problems. In the numerical results of Section 3.4, we have limited ourselves to a model specification where convexity of all problem instances is guaranteed. But we do acknowledge that this appears to be a strong assumption in models with stochastic price impacts.

3.3.3 Using the Dual Formulation to Investigate More Complex Models

We mention at this point just how useful the dual methodology can be. Suppose, for example, that we begin with a simplified model where the non-impact price, $\tilde{\mathbf{p}}_t$, follows a simple random walk. Suppose also that we have established that all of the corresponding dual problems are convex and therefore very tractable. Consider now adding the following features to this simplified model: (i) we replace the random walk dynamics with stochastic variance-covariance dynamics and (ii) we introduce new state variables to model return predictability. Then the dual problem instances of this more complex model will remain convex as long as the stochastic volatility and state variable dynamics are not influenced by the decision variables. This follows immediately once we recognize that these new model features only affect the *distribution* of dual problem instances; in particular they do not change the convexity properties of the dual problem instances. Another way to see this is to consider the dual problem instance in (3.3.2). We see $\mathbf{r}_{t+1}$ appear in the constraints for this problem but Σ_t does not. Similarly, if we had a state vector, Z_t, that induced predictability in the returns, then Z_t would also not appear in the dual problem formulation and so it would not influence the convexity of the dual problems. It would, however, influence the distribution of the dual problems and therefore the actual value of the dual bound.

Note also that in this more complex model we are free to use the original $\tilde{V}_t$ from the simplified model in order to construct dual penalties and a dual bound. Similarly we are also free to use a sub-optimal policy for the simplified model to construct a primal bound for the more complex model. If we then find that the resulting duality gap is small then we know that the new features (as currently calibrated) have little influence and can be safely ignored. We would argue then that the dual methodology can also be employed to determine whether or not certain features, that are known to exist in the market, are sufficiently important as to require explicit modeling. We do precisely this in Section 3.5.2 where we investigate whether or not it's necessary to include a stochastic component in our temporary price impact model.

3.4 Numerical Results

In this section we analyze the performance of the primal and dual bounds on a stylized example where the agent needs to purchase $100,000$ shares in each of $n = 50$ securities over $T + 1 = 78$ time periods. The 78 time periods are intended to reflect the fact that there are 78 five-minute periods in a trading day of 6.5 hours. We note that each dual problem instance has a total of $50 \times 78 = 3,900$ decision variables and that computing the OLFC policy at each time t involves $(T + 1 - t) \times 50$ decision variables. Of course in practice, portfolio execution problems can involve several hundred assets and computing good sub-optimal policies in (almost) real time and evaluating these policies via duality can itself be a significant challenge. If, for example, a portfolio execution problem has 500 assets then each dual problem instance will have $500 \times 78 = 39,000$ decision variables and solving a large number of these dual problem instances would therefore be computationally demanding. We note, however, that there is no need to compute dual bounds in real time as they are not required to implement a given execution policy. There is therefore no problem with computing dual bounds off-line.

Returning to our example, we take our 50 stocks to be the top 50 stocks by market capitalization in the S&P 500 as of October 12 2011. We take the initial price vector, $\tilde{\mathbf{p}}_0$, to be the initial prices of these 50 stocks on that date. We consider 10 different values of the risk aversion parameter: $\gamma = k \times 10^{-7}$ for $k = 1, \ldots, 10$. For each value of γ, we simulate $5,000$ sample paths and implement the various sub-optimal policies on each sample path. We use only 100 sample paths to estimate the dual bounds as this smaller number was found to estimate the dual bounds with an accuracy comparable to that of the primal bounds. We also used control variates as a variance reduction technique and the specific details can be found in Appendix B.1.5. We note that obtaining the primal bound required solving $5,000 \times 78$ optimization problems whereas obtaining the dual bound required solving just 100 optimization problems. The computational bottleneck was therefore in estimating the primal bound.

3.4.1 The Price Impact Dynamics

We assume the price impact coefficients of a stock are inversely proportional to its average daily trading volume. In particular, we take the permanent price impact coefficient matrix $\mathbf{A}_t = \mathbf{A}$ to be a constant diagonal matrix so that there is no cross price impact. The diagonal elements of $\mathbf{A}$ are obtained by assuming that the purchase of 10% of the average daily volume of a stock will incur a permanent price impact of 10 basis points of the stock's initial price. In the case of *Apple Inc*, for example, this implies a permanent price impact coefficient of 1.7830×10^{-7}. With an initial price of \$407.33, we therefore see that if all $100,000$ shares are purchased immediately then the permanent price impact will be $1.7830 \times 10^{-7} \times 10^5 =$ \$0.01783 per share.

We assume that the temporary price impact matrix $\mathbf{B}_t$ is driven by a single factor, x_t, according to

$$\mathbf{B}_t = \max(x_t, 0)\mathbf{B} \tag{3.4.1}$$

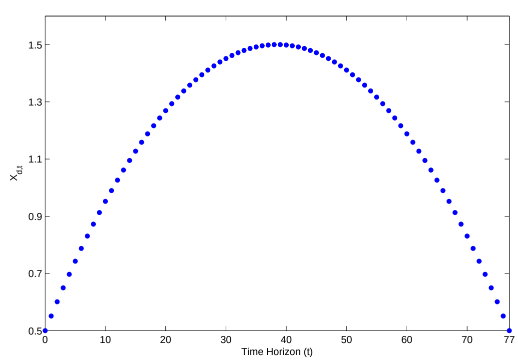
$$x_t = x_{d,t} + x_{s,t} \tag{3.4.2}$$

$$x_{s,t} = \rho x_{s,t-1} + \eta_t \tag{3.4.3}$$

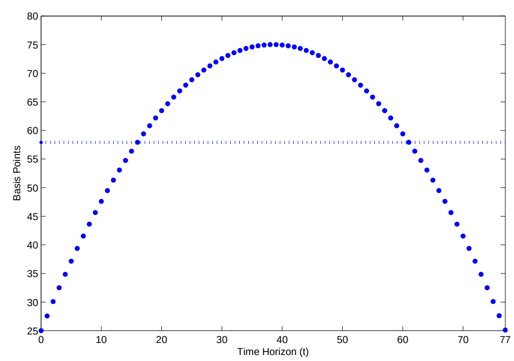
where $\mathbf{B}$ is a positive-definite matrix and the max operator in (3.4.1) ensures $\mathbf{B}_t$ is positive semi-definite for all t. The state variable x_t is composed of a deterministic component, $x_{d,t}$, and a stochastic component, $x_{s,t}$. The $x_{d,t}$'s are chosen to reflect the fact that the market impact tends to be smaller nearer the open and the close of the trading day. We have assumed $x_{s,t}$ is an AR(1) process so that η_t's in (3.4.3) are IID normal random variables with mean 0 and variance σ_η^2. The AR(1) assumption allows us to capture mean-reverting and clustering effects so that periods of low (high) liquidity tend to be followed by periods of low (high) liquidity. We emphasize here, however, that this is a stylized example and in practice considerably more care would be required to specify the dynamics of $\mathbf{B}_t$.

We assume that $\mathbf{B}$ is also a diagonal matrix and represents a temporary price impact of 50 basis points when $x_t = 1$ and 1% of the average daily volume of a stock is purchased immediately. Since $\mathbf{A}$ and $\mathbf{B}$ are both diagonal matrices we have assumed here that there are no cross-price impacts. We set $x_{s,0} = 0$, $\rho = 0.6$ and $\sigma_\eta^2 = 0.25^2 \times (1 - \rho^2)$. If we ignore the max operator in (3.4.1) then these parameter values correspond to a stationary distribution of $\mathbf{B}_t$ where the diagonal elements have a standard deviation equivalent to a temporary price impact of 12.5 basis points.

Figure 3.1(a) plots $x_{d,t}$ as a function of t while Figure 3.1(b) shows the expected temporary price impact of purchasing 1% of the average daily volume shares at time t as a function of t and expressed in basis points. Because this is a stylized example we simply assumed that $x_{d,t}$ is a simple quadratic function as this is perhaps the simplest way to model the fact that the price impact tends to be lower near the open and the close. Of course we could have just as easily assumed an alternative and possibly more accurate function for $x_{d,t}$. It is also not necessary to restrict ourselves to a scalar process, x_t, and we expect a higher dimensional process would be more realistic in practice.



(a) Deterministic Component, $x_{d,t}$



(b) Expected Temporary Price Impact

Figure 3.1: (a) The deterministic component $x_{d,t}$, is assumed to be quadratic in t. (b) The expected temporary price impact of purchasing 1% of the average daily volume, expressed in basis points, is identical for each of the n securities. The dashed line is the time-average temporary price impact.

Returning now to the case of *Apple Inc*, the purchase of all 100,000 shares at time $t = 0$ incurs a temporary price impact of $0.5 \times 8.9150 \times 10^{-6} \times 10^5 = \$0.4458$ per share. Therefore the total execution cost for this trade is approximately $(0.01783 + 0.4458) \times 100,000 = \$46,358$. The assumption that $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices is made because of the difficulty of calibrating cross-price impacts but we do note that in some circumstances it may be worthwhile accounting for them. Table B.5 in Appendix B.4.2 contains the initial prices, average daily volume and price impact coefficients that we assumed for the 50 stocks.

Finally we note that dynamics of $\mathbf{A}_t$ and $\mathbf{B}_t$ assumed here satisfy conditions (i) and (ii) of Appendix B.2.1. We are therefore guaranteed that all dual problem instances will be convex.

3.4.2 Variance-Covariance Dynamics

We assume that $\mathbf{\Sigma}_t$ follows an O-GARCH model as in Alexander [1] so that

$$\mathbf{\Sigma}_t = \mathbf{F}\mathbf{\Omega}_t\mathbf{F}' + \mathbf{\Upsilon} \tag{3.4.4}$$

where $\mathbf{\Omega}_t$ is a diagonal matrix, $\mathbf{F}$ is a matrix of factor loadings and $\mathbf{\Upsilon}$ is a diagonal matrix of idiosyncratic variances. The diagonal elements in $\mathbf{\Omega}_t$ are assumed to follow independent GARCH(1,1) processes. Further details on this model and its calibration can be found in Appendix B.4.1.

3.4.3 Results

In our numerical results we report both the average execution cost of each policy as well the certainty equivalent (CE) execution cost. Given an average utility, $\hat{u}$, calculated as the average utility across the simulated sample paths, the CE cost is defined as the execution cost (in basis points), c_e, which yields the same utility with certainty if all shares are purchased immediately. That is, c_e solves

$$\hat{u} = \exp(\gamma \tilde{\mathbf{p}}_0' \mathbf{w}_0 (1 + c_e/10,000)). \tag{3.4.5}$$

The CE cost is much easier to interpret than the average utility and so we prefer to report the former together with the average execution cost. We note of course that the performance of any given policy is measured by the CE cost and *not* the average execution cost which we would naturally expect to increase with the level of risk aversion, γ.

In Figure 3.2 we have plotted the CE costs for the simple, OLFC and risk-neutral OLFC policies. We have also plotted the CE cost for the dual bound constructed using the penalties based on the modified OLFC policy as explained in Section 3.3.1. In this case we see that there is no discernible gap between the OLFC primal bound and the dual bound which implies that the OLFC policy is actually very close to optimal. For low levels of risk aversion the simple and risk-neutral OLFC policies lose only around 5 basis points with respect to the OLFC policy but this number increases (as expected) to approximately 60 basis points as γ increases.

We have also plotted the average execution cost as a function of γ in Figure 3.3(a). We see that the risk-neutral OLFC policy has the lowest average cost which is as expected since the risk-neutral agent only cares about minimizing execution cost and is indifferent to execution risk. Since the simple policy is deterministic and the risk-neutral OLFC policy does not depend on γ we see that their expected execution costs do not vary with γ. As

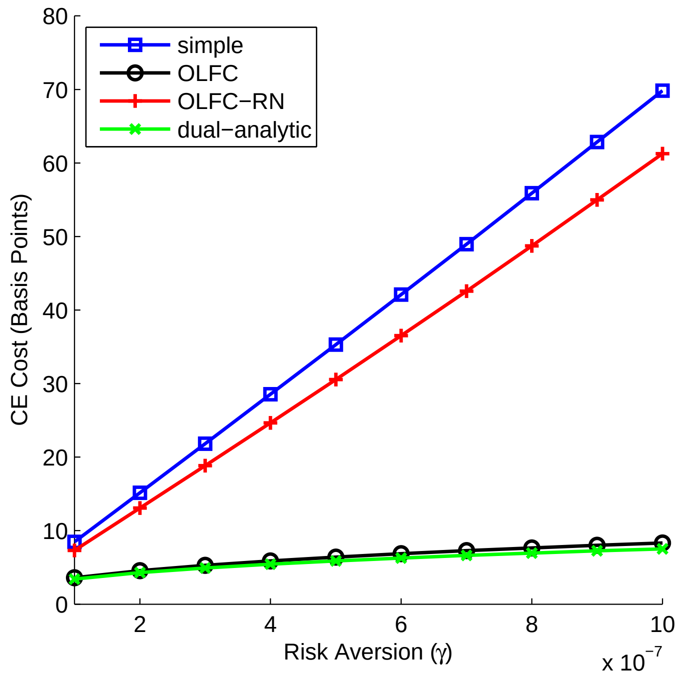
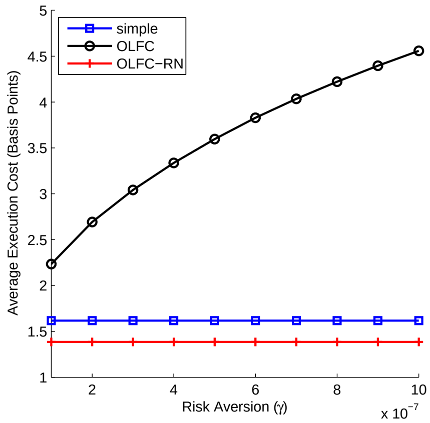


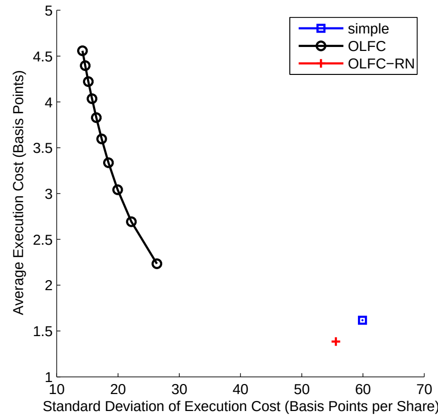
Figure 3.2: Certainty equivalent costs

γ increases the agent becomes more risk averse and therefore prefers to buy more shares in earlier time periods. This results in a higher price impact which is reflected by the higher average execution cost for the OLFC policy.

In Figure 3.3(b) we have plotted the mean-standard deviation frontier corresponding to the OLFC policy. As expected, we see that a higher average execution cost is accompanied by a lower standard deviation. We also see that the simple policy and risk-neutral OLFC policy have low average execution costs but high standard deviations.



(a) Average Execution Cost



(b) Mean-Standard Deviation Frontier

Figure 3.3: Average cost and mean-standard deviation frontier

3.5 Extensions and Other Portfolio Execution Problems

In this section we briefly describe how our model can be extended to include a non-linear temporary price impact as well as predictable state variables with linear dynamics. We also discuss limit-order book models and explain how the dual technology can also handle these problems. We begin, however, by considering the case where dual penalties cannot be calculated explicitly. We will see that we can still quickly compute good dual bounds in that case using Monte-Carlo methods. While this is not a "model extension" it clearly broadens the range of penalties that can be used when constructing dual bounds.

3.5.1 Computing Dual Bounds When Penalties Cannot Be Computed Explicitly

We noted in Section 3.3.1 that we needed to compute terms of the form $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ and $\mathbb{E}_t[\nabla\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$ in order to solve a given dual problem instance as in (3.3.2). BSS [16] showed that if we could instead only compute unbiased (conditional on $\mathcal{F}_T$) estimates of these terms then we would still obtain a valid, albeit more conservative, dual bound. The downside of this approach is that we require Monte-Carlo simulation to estimate these conditional expectations and so constructing the dual bound therefore requires nested Monte-Carlo's which can be very demanding from a computational standpoint.

In this section we will use suitably randomized *low-discrepancy sequences* (LDS)[1] to perform the nested simulations and we will compare the resulting dual bound to the dual bound that we obtained when the conditional expectations are computed in closed form. We will see that we can compute LDS-based dual bounds that are virtually indistinguishable from the original dual bounds and that these new bounds only require 10% additional work.

We note that each of the two conditional expectations may be written in the form $\theta = \mathbb{E}[g(\mathbf{r})]$ for some function $g(\cdot)$ and where $\mathbf{r}$ is an n-dimensional multivariate normal random vector with mean vector $\mathbf{0}$ and covariance matrix Σ. The precise form of $g(\cdot)$ is clear from (3.3.6) in the case of $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ and is also easy to determine in the case of $\mathbb{E}_t[\nabla\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$.

[1]See, for example, Chapter 5 of Glasserman [38] for an introduction to LDS and the randomization technique we describe here.

Using the inverse-transform approach to Monte-Carlo we can then easily rewrite θ as

$$\theta = \mathbb{E}[g(\mathbf{U})] \tag{3.5.1}$$

where $\mathbf{U}$ is an n-dimensional vector of independent $U(0,1)$ random variables. We could perform the integration in (3.5.1) using LDS which have very nice theoretical properties and which often work particularly well for high-dimensional integrals. The problem with doing this is that the resulting estimate of θ is deterministic and in particular will have some (although presumably small) bias. As a result we could no longer conclude that our estimate of V_{lb} would indeed be an unbiased lower bound. We can overcome this problem by randomizing the LDS in the following manner. In particular, we set

$$\hat{\theta} := \frac{1}{M_u} \sum_{j=1}^{M_u} \left(\frac{1}{M_l} \sum_{i=1}^{M_l} g\big((\mathbf{l}_i + \mathbf{U}_j) \text{mod } \mathbf{1}\big) \right) \tag{3.5.2}$$

where $(\mathbf{l}_1, ..., \mathbf{l}_{M_l})$ is a series of n-dimensional low discrepancy points, and $(\mathbf{U}_1, ..., \mathbf{U}_{M_u})$ is a series of independent n-dimensional uniform random vectors. Note also that the mod-operator in (3.5.2) applies component-wise and that the inner summation in (3.5.2) uses the same uniform vector, $\mathbf{U}_j$, for all M_l samples. In practice M_u can be very small, e.g. 5 or 10, whereas M_l might be on the order of 10^4 or higher. That $\hat{\theta}$ is an unbiased estimator for θ follows from the fact that if $\mathbf{U}_j$ is an independent n-dimensional uniform random vector then so too is $(\mathbf{l}_i + \mathbf{U}_j)\text{mod } \mathbf{1}$ for all i. If we set $\hat{\theta}_j = \frac{1}{M_l} \sum_{i=1}^{M_l} g\big((\mathbf{l}_i + \mathbf{U}_j)\text{mod } \mathbf{1}\big)$ then we see that $\hat{\theta} = \frac{1}{M_u}\sum_{j=1}^{M_u}\hat{\theta}_j$ is the mean of M_u IID random variables, and so confidence intervals can be constructed in the usual way if so desired.

Numerical Results

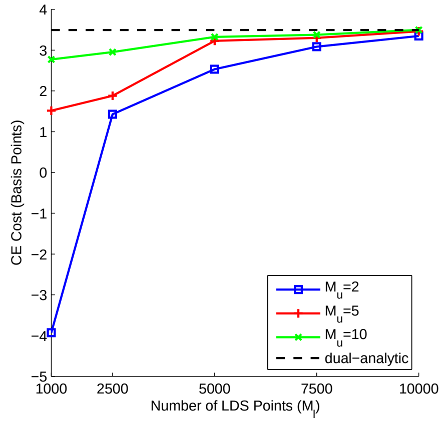
We considered the same portfolio execution problem that we studied in Section 3.4. In Figure 3.4 we plot the performance of the dual bound as estimated using our LDS scheme for various values of M_u, M_l and γ. Our LDS was an n-dimensional *Sobol* sequence that we generated using *Matlab's* LDS functionality. We skipped the first 1000 points, retained every 101^{st} point thereafter and also applied the so-called Matousek-Affine-Owen scrambling scheme.

The dashed lines in Figure 3.4 represent the dual bound that was computed using the analytic expressions for $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ and $\mathbb{E}_t[\nabla\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$. These dashed lines are therefore the same bounds that we calculated in Section 3.4. As expected, we see that the new LDS-based dual bounds are conservative so that they are all at or below the corresponding dashed line. It is also not surprising that as we increase M_l and M_u the bounds improve to the point that with $M_l = 10,000$ and $M_u = 10$ the bound is virtually indistinguishable from the dashed lines. (In fact in a couple of cases, e.g. $\gamma = 7 \times 10^{-7}$ and $M_u = 2$, the LDS-based bound was marginally higher than the dashed line but the difference was well within the Monte-Carlo's standard error. Recall that even the dashed line has some statistical error as it is computed as the average over 100 dual problem instances. The LDS-based bounds use the same 100 dual paths and so these bounds are also exposed to this error as well as the nested simulation error of (3.5.2).)

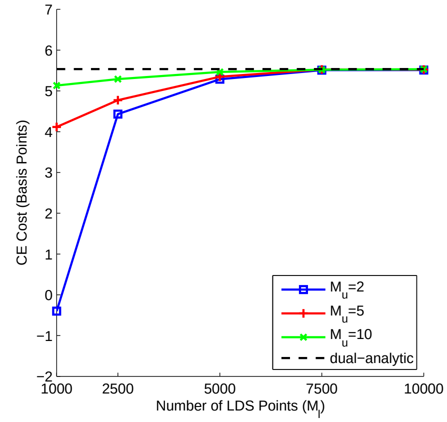
The surprising feature of these results is how little additional work is required. For example, it typically took approximately 150 seconds to solve a dual problem instance when we used the analytic expressions for the dual penalties. When we used the LDS to estimate these penalties with $M_l = 10,000$ and $M_u = 10$ the running time increased by approximately 15 seconds for a relative increase of just 10%. This was possible because it is straightforward to generate and store in advance all of the $(\mathbf{l}_i + \mathbf{U}_j) \text{mod } \mathbf{1}$ vectors. Moreover the $g(\cdot)$ function was easy to evaluate as it could be computed explicitly. Of course, the use of exponential utility would force us to consider only those distributions that have a moment generating function.

Other Applications

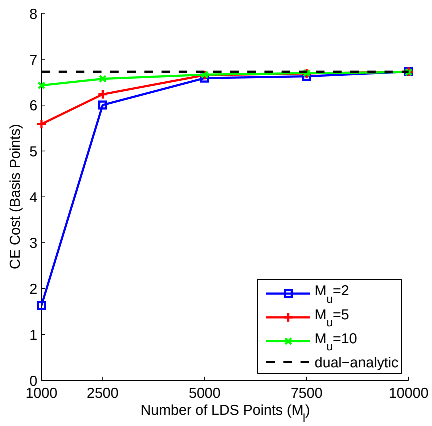
The results of Figure 3.4 suggest that the randomized LDS approach can be used efficiently to generate very good dual bounds. One possible application would be for solving portfolio execution problems where we suspect that the random return vectors are only approximately normally distributed. In particular, we may suspect that the true return vectors have fatter tails. In that case we could, for example, use the same OLFC policy (which assumed normal returns) as before but in order to evaluate it using the dual formulation we would need to estimate $\mathbb{E}_t[\hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})]$ and $\mathbb{E}_t[\nabla \hat{V}_{t+1}(\tilde{\mathbf{s}}_{0:t})']$ using the *true* return distributions. If these terms cannot be computed analytically, then we could use the approach outlined in this



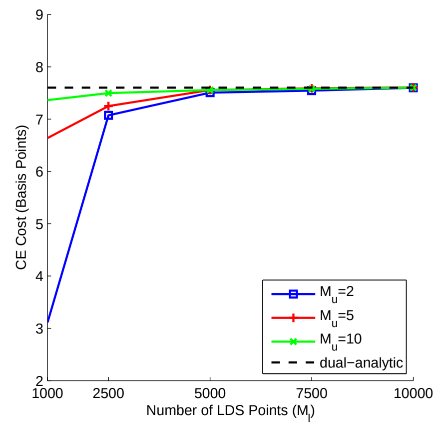
(a) $\gamma = 1 \times 10^{-7}$



(b) $\gamma = 4 \times 10^{-7}$



(c) $\gamma = 7 \times 10^{-7}$



(d) $\gamma = 10 \times 10^{-7}$

Figure 3.4: Dual bounds obtained using randomized LDS to estimate dual penalties

section. If the resulting duality gap, i.e. the difference between the primal and dual bounds, was sufficiently small then we would know that we could safely ignore the non-normality of returns. Otherwise, we would need to adapt our policy to take this non-normality into account.

3.5.2 Investigating the Temporary Price Impact Model

In Section 3.3.3 we discussed how the dual methodology could be used to determine whether or not certain market features require explicit modeling. We demonstrate this idea here by investigating the features of our temporary price impact model from Section 3.4.1. Suppose the agent suspects that only the time-varying feature of the temporary price impact is important and that the stochastic component can be safely ignored by the execution policy. To investigate this conjecture we consider two additional policies. The first policy, which we call the OLFC-TV policy, assumes the temporary price impact is time-varying but deterministic. In particular the OLFC-TV policy assumes the temporary price impact at time t is given by $\mathrm{E}_0[\mathbf{B}_t]$. The second policy, which we call the OLFC-C policy, assumes the temporary price impact is constant and equal to its time average across all time periods. Note that the "true" temporary price impact model remains as described by (3.4.1) to (3.4.3).

The performance of these policies is shown in Figure 3.5 together with the original OLFC policy and our dual bound. Nor surprisingly we see that the OLFC policy outperforms the OLFC-TV policy which in turn outperforms the OLFC-C policy. It is interesting to note the degree of out-performance, however, particularly for higher levels of risk aversion where

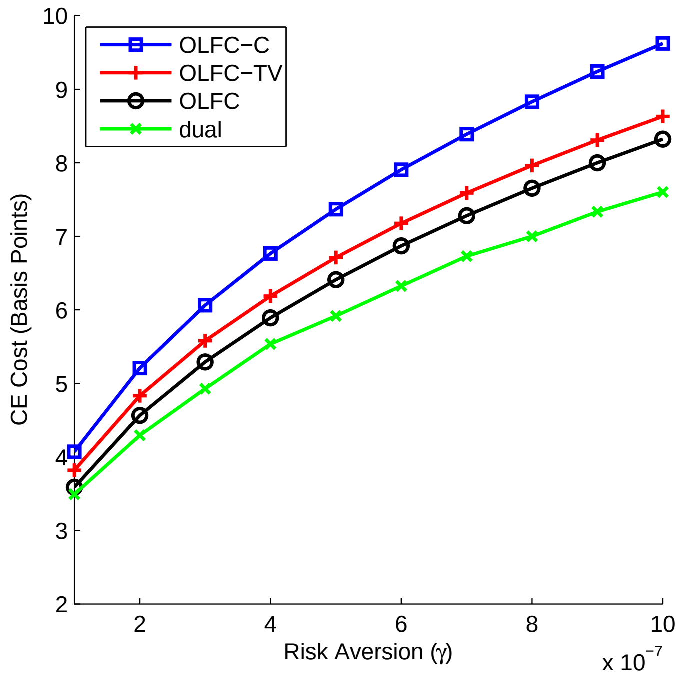


Figure 3.5: Certainty equivalent costs for comparing model features

the OLFC-TV policy performs almost as well as the OLFC policy and much better than the OLFC-C. In this case we could argue that the agent's conjecture was justified although to draw this conclusion more generally we would need a more careful specification and calibration of the temporary price impact model.

3.5.3 Including a Non-Linear Temporary Price Impact

In Section 3.2, we introduced a model with linear price impacts but of course in practice the true price dynamics are generally more complex. In this section we therefore investigate the case where the price dynamics also include a temporary non-linear price impact. (We cannot include a permanent non-linear price impact in our modeling framework as this will introduce the possibility of arbitrage; see Huberman and Stanzl [44], for example.) We let $h_t(\cdot)$ denote this non-linear temporary price impact function at time t so that (3.2.5) is now replaced by

$$\mathbf{p}_t = \tilde{\mathbf{p}}_t + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t + h_t(\mathbf{s}_t). \tag{3.5.3}$$

The Primal Problem

If the agent ignores the non-linear impact he will implement a portfolio execution policy that he perceives as being a good policy for the linear-price impact model of Section 3.2. The original OLFC policy of Section 3.2 could play the role of this "good" policy. Note that the dual methodology can be used to determine whether or not the non-linear price impact is significant and therefore needs to be accounted for in the policy. In particular, if we find the OLFC policy is close to optimal even in the presence of the non-linear price impact, then the latter need not be modeled explicitly.

In practice, of course, we would expect the original OLFC policy to be far from optimal implying that the non-linear price impact is significant. In that case the agent should explicitly account for the non-linear impact in constructing the OLFC policy. This would be quite straightforward and we could formulate the problem analogously to (B.1-18). We could also find similar expressions to (3.5.5) and (3.5.6) below for the gradient and Hessian of the OLFC objective function and it would be reasonable to insist that this Hessian be positive definite for all $\mathbf{s}$ in order to guarantee the convexity of the OLFC problem.

The Dual Problem

Calculating dual bounds for the non-linear impact model proceeds exactly as in Section 3.3.1. Many dual problem instances are simulated and each results in an optimization problem of the form

$$\begin{aligned} \min_{\mathbf{s}\in\mathbb{S}} \quad & \exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right) + \sum_{t=0}^{T-1} \left(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})\right) \\ \text{s.t.} \quad & \mathbf{p}_t = \tilde{\mathbf{p}}_t + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t + h_t(\mathbf{s}_t) \\ & \tilde{\mathbf{p}}_{t+1} = \tilde{\mathbf{p}}_t + \mathbf{r}_{t+1} + \mathbf{A}_t\mathbf{s}_t \end{aligned} \tag{3.5.4}$$

where $\tilde{V}_{t+1}$ is some suitably linearized approximate value function. The analog to (B.2-1) under the non-linear temporary price impact model is

$$\begin{aligned} f(\mathbf{s}) \;&:=\; \exp\left(\gamma \sum_{t=0}^{T} \left(\tilde{\mathbf{p}}_0 + \sum_{i=0}^{t-1} \mathbf{A}_i\mathbf{s}_i + \sum_{i=1}^{t} \mathbf{r}_i + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t + h_t(\mathbf{s}_t)\right)' \mathbf{s}_t\right) \\ &=\; \exp(\gamma \tilde{\mathbf{p}}_0' \mathbf{w}_0)\ \exp\left(\gamma \left(\frac{1}{2}\mathbf{s}'\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0}'\mathbf{s} + \tilde{h}(\mathbf{s})\right)\right). \end{aligned}$$

The gradient vector and Hessian matrix of f are

$$\nabla f(\mathbf{s}) = \gamma f(\mathbf{s})(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0} + \nabla\tilde{h}(\mathbf{s})) \tag{3.5.5}$$

$$Hf(\mathbf{s}) = \gamma f(\mathbf{s})\left(\mathbf{Q} + H\tilde{h}(\mathbf{s}) + \gamma(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0} + \nabla\tilde{h}(\mathbf{s}))(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0} + \nabla\tilde{h}(\mathbf{s}))'\right) \tag{3.5.6}$$

where $\tilde{h}(\mathbf{s}) := \sum_{t=0}^{T} h_t'(\mathbf{s}_t)\mathbf{s}_t$, $\nabla\tilde{h}(\mathbf{s})$ is the gradient of $\tilde{h}$ and $H\tilde{h}(\mathbf{s})$ is the Hessian matrix of $\tilde{h}$. A sufficient condition to guarantee that $Hf(\mathbf{s})$ is positive definite (so that f is therefore convex) is that $\tilde{h}$ is convex and that $\mathbf{Q}$ is positive definite.

3.5.4 Predictable State Dynamics

Some of the earlier papers such as Bertsimas and Lo [11] and Bertsimas et al. [10] (some details are available in Appendix B.3) allow for a predictable component in the return dynamics. In particular, they assumed the price dynamics depended on a state variable, X_t, which itself had linear dynamics. Assuming the agent is risk-neutral, they were able to solve recursively for the optimal solution in the same manner as the analysis of Appendix B.1.4.

We could include a similar state variable or vector in our model. Except for some specific circumstances, we would not be able to solve for the optimal feed-back control policy but we could still solve for the OLFC policy in that case. It should also be clear that we can again use a modified version of the OLFC value function to compute dual penalties and therefore obtain valid dual bounds. We do not pursue this any further as allowing for predictable price dynamics is generally of less interest than the accurate modeling of price impact and liquidity effects.

3.5.5 Limit Order Book Models

Building on the earlier single-stock work of Obizhaeva and Wang [56], Alfonsi et al. [2] and Predoiuet al [58], Giesecke, Tsoukalas and Wang [70] formulate the portfolio execution problem at the limit order book (LOB) level. Here we briefly discuss this model and note that the dual methodology may also be applied in this context.

Their model assumes a two-sided block-shaped order book with infinite depth and time-invariant density. The mid price is a random walk with zero drift. A buy (sell) order $\mathbf{s}_t^+$ ($\mathbf{s}_t^-$) is executed as a market order against available inventory in the ask (bid) side of the LOB, creating a linear temporary price impact that moves the best ask (bid) price away from the mid-price. The order also creates a linear permanent price by moving the mid-price up (down). Newly arriving limit orders replenish the inventory and cause the distance between the best ask (bid) price and the mid-price to decay exponentially. The execution problem is then formulated as

$$\max_{\mathbf{s}_t^+, \mathbf{s}_t^- \in \mathcal{F}_t, t=0,\ldots,T} \quad \mathbb{E}_0\left[-\exp\left(-\gamma \sum_{t=0}^{T}\left((\mathbf{p}_t^b - \mathbf{B}^b\mathbf{s}_t^-)'\mathbf{s}_t^- - (\mathbf{p}_t^a + \mathbf{B}^a\mathbf{s}_t^+)'\mathbf{s}_t^+\right)\right)\right] \quad (3.5.7)$$

$$\text{s.t.} \quad \mathbf{p}_t^a = \tilde{\mathbf{p}}_0 + \sum_{j=1}^{t}\mathbf{r}_j + \frac{1}{2}\boldsymbol{\delta}_t + \mathbf{A}\sum_{j=0}^{t-1}(\mathbf{s}_j^+ - \mathbf{s}_j^-) + \mathbf{d}_t^a \quad (3.5.8)$$

$$\mathbf{p}_t^b = \tilde{\mathbf{p}}_0 + \sum_{j=1}^{t}\mathbf{r}_j - \frac{1}{2}\boldsymbol{\delta}_t + \mathbf{A}\sum_{j=0}^{t-1}(\mathbf{s}_j^+ - \mathbf{s}_j^-) + \mathbf{d}_t^b \quad (3.5.9)$$

$$\mathbf{d}_t^a = (\mathbf{d}_{t-1}^a + \mathbf{B}^a\mathbf{s}_{t-1}^+ - \mathbf{A}(\mathbf{s}_{t-1}^+ - \mathbf{s}_{t-1}^-))\exp(-\boldsymbol{\rho}^a\Delta t) \quad (3.5.10)$$

$$\mathbf{d}_t^b = (\mathbf{d}_{t-1}^b - \mathbf{B}^b\mathbf{s}_{t-1}^- - \mathbf{A}(\mathbf{s}_{t-1}^+ - \mathbf{s}_{t-1}^-))\exp(-\boldsymbol{\rho}^b\Delta t) \quad (3.5.11)$$

$$\sum_{t=0}^{T}(\mathbf{s}_t^+ - \mathbf{s}_t^-) = \mathbf{w}_0 \text{ and } \mathbf{s}_t^+, \mathbf{s}_t^- \geq \mathbf{0} \quad (3.5.12)$$

where $\mathbf{p}_t^a$ and $\mathbf{p}_t^b$ are the best ask/bid prices in the LOB at time t, $\tilde{\mathbf{p}}_0$ is the initial mid-price and $\boldsymbol{\delta}_t$ is the bid-ask spread. $\mathbf{A}$ is the permanent price impact coefficient matrix while $\mathbf{B}^a$ and $\mathbf{B}^b$ are the temporary price impact coefficients for buy and sell orders, respectively.

Each of these matrices are assumed to be deterministic. In addition $\mathbf{B}^a$ and $\mathbf{B}^b$ are diagonal matrices with diagonal elements equal to the inverse of the order book density on the ask and bid sides, respectively. $\mathbf{d}_t^a$ and $\mathbf{d}_t^b$ track the deviation of the current ask and bid prices from their steady state levels and they decay exponentially at constant speeds $\boldsymbol{\rho}^a$ and $\boldsymbol{\rho}^b$ due to order book replenishment. Δt is the length of a time period. The only uncertainty comes from the returns, $\mathbf{r}_j$, which are IID normally distributed with zero mean and constant covariance matrix. The objective is to maximize an exponential utility function over the terminal wealth which is equivalent to minimizing the exponential utility function over cost in our model.

Giesecke et al. [70] show that the optimal execution policy is deterministic and find an equivalent quadratic formulation for the problem. Such a result is quite standard in the portfolio optimization literature and indeed the equivalence of our formulations in Appendices B.1.3 and B.1.4 are in the same spirit. But once we begin to relax some of the assumptions in this model then solving for the optimal policy even numerically becomes a very challenging task and we again find ourselves in the situation of needing to construct and evaluate good sub-optimal policies. For example if we introduce a state vector, Z_t, that drives the dynamics of the $\mathbf{r}_t$'s and the temporary price impact matrices then a dual problem instance will take the form

$$\begin{aligned} \max_{\mathbf{s}} \quad & -\exp\left(-\gamma \sum_{t=0}^{T} \left((\mathbf{p}_t^b - \mathbf{B}_t^b \mathbf{s}_t^-)' \mathbf{s}_t^- - (\mathbf{p}_t^a + \mathbf{B}_t^a \mathbf{s}_t^+)' \mathbf{s}_t^+\right)\right) \\ & + \sum_{t=0}^{T-1} \left(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})\right) \end{aligned} \tag{3.5.13}$$

subject to (3.5.8) to (3.5.12) but with $\mathbf{B}^a$ and $\mathbf{B}^b$ replaced by $\mathbf{B}^a_j$ and $\mathbf{B}^b_j$ where $j = t$ or $t-1$ as appropriate. Note again that Z_t does not appear explicitly in any dual problem instance but instead it changes the distribution of these instances. We see that the dual problem in (3.5.13) is similar to the dual problem of Section 3.3 and that we could approach this problem in a similar manner.

3.6 Conclusions and Future Research

Any realistic model of portfolio execution should be able to handle features such as stochastic variance-covariance dynamics, return-predictability, time-of-day effects as well as stochastic liquidity / price-impacts and possible risk aversion. Any model which includes these features will not in general be analytically tractable and so it will be necessary instead to construct good sub-optimal policies. It is important that these sub-optimal policies can be properly evaluated and in this chapter we have demonstrated the use of duality methods to do this. In particular, our model is capable of capturing all of the above effects and our OLFC policy appears to be capable of generating very good primal bounds while a variation of the associated OLFC value function also leads to very tight dual bounds. (Of course we have only shown this to actually be the case for the parameter settings we considered, but we suspect it to be true more generally for realistic parameter settings.) While these dual methods have become quite standard in a relatively short period of time, implementing them is not straightforward in general since the dual optimization problems can be quite complex and establishing convexity or non-convexity may also require some effort.

An additional contribution of this chapter is that valid and tight dual bounds can still be computed efficiently even when the dual penalties are not explicitly available and need to be estimated via Monte-Carlo. The numerical experiments of Section 3.5.1 where we used randomized low-discrepancy sequences have demonstrated the validity of this broadly applicable approach.

We have also demonstrated some useful properties of the dual problem in the context of portfolio execution problems. For example, we noted that the convexity of the dual problem does not depend on variance-covariance dynamics or state variable dynamics as long as these dynamics are not influenced by the execution policy. We have also noted how the duality technology can be used to determine in advance if a market feature, e.g. the cross-price impact of Appendix B.3 or the non-linear price-impacts of Section 3.5.3, need to be accounted for by the portfolio execution policy.

There are several possible future research directions and these include conducting a more detailed calibration and empirical evaluation of the model we have proposed in this chapter. Another direction is to use duality to study the impact of using misspecified return distributions. For example, the assumption of (conditional) multivariate normality has been made throughout this chapter and throughout the literature on single-stock and portfolio execution. This of course is due to the tractability of the normal distribution. In practice, however, we might expect conditional returns to only be approximately normal. We could still construct feasible execution policies in this case by *pretending* that everything was normally distributed after which we could use the true model to simulate sample paths to estimate valid primal and dual bounds. Of course when computing the penalty terms we require expectations of the linearized approximate value functions and these expectations must be evaluated under the *true* model. If they are not available explicitly then they can be estimated via nested Monte-Carlo's as described in Section 3.5.

Another potentially interesting direction is the problem of multi-portfolio execution where firms need to solve the portfolio execution problem simultaneously for multiple portfolios rather than just one portfolio. This raises the question as to how we should prioritize the portfolios. We could, for example, solve each portfolio execution problem separately and ignore all knowledge of the other portfolio trades that need to be executed. We would expect this to be far from optimal in general although it would be necessary to first define what is meant by "optimal". Towards this end, we would also need to determine how to allocate the total execution costs fairly across each of the portfolios which is difficult when impact costs are non-linear. The multi-portfolio execution problem is non-trivial and depending on the solution technique, can draw on concepts from cooperative or non-cooperative game theory. It was first studied by O'Cinneide, Scherer and Xu [19] who proposed optimizing the total "social welfare" without accounting for fairness. Stubbs and Vandenbussche [66] study the pros and cons of both the cooperative and non-cooperative solution techniques in a static rather than dynamic context. Once the multi-portfolio execution problem has been properly formulated, however, then it will be necessary to determine good execution policies. The duality technology might be very useful in assessing the quality of these policies.

A final direction for future research is understanding just how plausible it is to impose conditions that guarantee the convexity of all dual problem instances. There should be little difficulty in justifying this assumption when the price impacts are assumed to be deterministic. Huberman and Stanzl [44] provide such a justification in the single stock case, for example. When the price impacts are stochastic, however, insisting on the positive-definiteness of the matrix, $\mathbf{Q}$, in (B.2-3) on all dual problem instances is not so easy to justify from an economic standpoint. From an intuitive viewpoint it seems plausible to insist that $\mathbf{Q}$ should be positive-definite in some *average* sense (as implied by the positive definiteness of (B.1-19), for example) but there is no reason to assume that all realizations of $\mathbf{Q}$ need to be positive-definite to avoid the absence of arbitrage.

If we were to consider models where dual convexity was not guaranteed then it should still be possible to obtain a valid dual bound. In particular, on each dual sample path we first check if $\mathbf{Q}$ is positive-definite. If it is, then we solve this dual problem instance as before. If $\mathbf{Q}$ is not positive-definite, then the dual optimization problem is difficult to solve in general. But we can still bound the optimal value of such a problem instance. For example, the exponential term in (3.3.2) is bounded below by zero and if the constraint set is a bounded polyhedron (as will typically be the case in practice) then LP techniques could be used to bound the second term in (3.3.2). If relatively few of the $\mathbf{Q}$'s fail to be positive-definite then we may still be able to obtain a good dual bound in this manner. Another possibility would be to form the concave dual (see Boyd and Vanderberghe [13], for example) of the non-convex dual instance. The optimal solution (or indeed a good feasible solution) to this concave problem will also provide a valid lower bound although it may be difficult to compute.

Ultimately we believe the goal is to construct realistic models that can include the various market effects that are found in practice, that can handle hundreds of securities or more, that are straightforward to calibrate and for which good sub-optimal policies can easily be found. Because it is impossible to ever know the true market dynamics, we believe these sub-optimal policies need to be robust to deviations from the assumed model and that duality based on information relaxations can play a key role in assessing this robustness.

Acknowledgements

We thank Costis Maglaras for providing us with the 5-minute price data on the S&P 500 securities that we used in Section 3.4 and Andrew Lo for providing the calibrated model parameters for the numerical experiments in Appendix B.3. We are also grateful to Costis Maglaras and Garud Iyengar for helpful comments and discussions.

Chapter 4

Information Relaxations and Dynamic Zero-Sum Games

4.1 Introduction

Zero-sum game is a very useful framework with lots of interesting applications including pursuit-evasion games, heads-up poker, callable bond and many two-person computer games. In this chapter we apply the duality methods based on information relaxations for stochastic dynamic problems to dynamic two-person zero-sum games. Our ultimate goal is to apply these dual techniques to very complex zero-sum games which cannot be solved to optimality.

The remainder of this chapter is organized as follows. We define the problem and establish the dual theory for finite horizon zero-sum games in Section 4.2. In Section 4.3 we outline how the approach can be extended to discounted infinite horizon problems as well as more general stochastic shortest path games. We conclude in Section 4.4.

4.2 Dual Bounds for Finite Horizon Zero-Sum Games

We begin with a finite horizon zero-sum game played by two players, A and B. We index time according to $t = 0, 1, ..., T$. and the evolution of information is described by a filtration $\mathbb{F} = \{\mathcal{F}_0, ..., \mathcal{F}_T\}$ with $\mathcal{F}_T = \mathcal{F}$. We make the simplifying assumption that both A and B have access to the same information $\mathcal{F}_t$ at time t and by insisting that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$ for all $t < T$ we model the assumption that the two players do not forget previously known information. We assume $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so player A and B know nothing initially. The state vector at time t is denoted by x_t. We let $\mathcal{A}_t(x_t)$ and $\mathcal{B}_t(x_t)$ be the sets of time-t actions that are feasible for players A and B, respectively, given the current state is x_t. We assume $\mathcal{A}_t(x_t)$ and $\mathcal{B}_t(x_t)$ are finite sets and that x_t evolves according to

$$x_{t+1} = f_t(x_t, a_t, b_t, \omega_{t+1}), \quad t = 0, ..., T-1 \tag{4.2.1}$$

where $a_t \in \mathcal{A}_t(x_t)$ and $b_t \in \mathcal{B}_t(x_t)$ are the actions of A and B, respectively, at time t, and ω_{t+1} is an $\mathcal{F}_{t+1}$-measurable random disturbance. After A and B implement their actions at time t, B makes a payment of $\tilde{g}_t(x_t, a_t, b_t)$ to A. The game ends at time T when B pays $g_T(x_T)$ to A. We assume[1] that each $\tilde{g}_t$ is $\mathcal{F}_t$-measurable.

[1] This assumption is not too restrictive. Suppose for example the true amount B pays to A is $\bar{g}(x_t, a_t, b_t, \omega_{t+1})$ so that it depends on the as yet unobserved disturbance, ω_{t+1}. Then we can replace it

It is useful to make the distinction between pure and mixed policies. A pure policy chooses an action deterministically whereas a mixed policy chooses an action randomly. In particular at time t a mixed policy for player A is a probability distribution $u_t = \{u_t^{(a_t)} \,|\, a_t \in \mathcal{A}(x_t)\}$ over A's feasible action set $\mathcal{A}(x_t)$ such that action a_t is selected with probability $u_t^{(a_t)}$. We note that a pure policy is of course also a mixed policy. We use $\mathbf{u} = (u_0, ..., u_{T-1})$ to represent player A's policy, and similarly define the probability distribution $v_t = \{v_t^{(b_t)} \,|\, b_t \in \mathcal{B}(x_t)\}$ and the policy $\mathbf{v} = (v_0, ..., v_{T-1})$ for player B. If $\mathbf{u}$ or $\mathbf{v}$ is a pure policy, we can simply write them as the action sequences $\mathbf{a} = (a_0, ..., a_{T-1})$ or $\mathbf{b} = (b_0, ..., b_{T-1})$. A joint policy $\{\mathbf{u}, \mathbf{v}\}$ is $\mathcal{F}_t$-adapted if each u_t and v_t is $\mathcal{F}_t$-measurable. We denote by $\mathcal{A}^{\mathbb{F}}$ and $\mathcal{B}^{\mathbb{F}}$ the set of all $\mathcal{F}_t$-adapted policies for A and B, respectively.

We let $g_t(x_t, u_t, v_t)$ denote the time-t *expected* payment from B to A when they use the mixed policies u_t and v_t. Then[2]

$$g_t(x_t, u_t, v_t) = \sum_{a_t \in \mathcal{A}(x_t)} \sum_{b_t \in \mathcal{B}(x_t)} u_t^{(a_t)} v_t^{(b_t)} \tilde{g}_t(x_t, a_t, b_t), \quad t = 0, ..., T-1 \tag{4.2.2}$$

and $g_t(x_t, u_t, v_t)$ is $\mathcal{F}_t$-measurable because each $\tilde{g}_t(x_t, a_t, b_t)$ is $\mathcal{F}_t$-measurable. The total expected payment from B to A under these policies is then

$$g(\mathbf{u}, \mathbf{v}) = g_T(x_T) + \sum_{t=0}^{T-1} g_t(x_t, u_t, v_t). \tag{4.2.3}$$

with $\tilde{g}_t(x_t, a_t, b_t) := \mathbb{E}[\bar{g}(x_t, a_t, b_t, \omega_{t+1})|\mathcal{F}_t]$ where the expectation is taken over ω_{t+1}, and thus $\tilde{g}_t(x_t, a_t, b_t)$ is $\mathcal{F}_t$-measurable.

[2] If A uses a pure policy, a_t, and B uses a mixed policy, v_t, the expected payment (4.2.2) reduces to $g_t(x_t, a_t, v_t) = \sum_{b_t \in \mathcal{B}(x_t)} v_t^{(b_t)} \tilde{g}_t(x_t, a_t, b_t)$. We note that $g_t(x_t, u_t, b_t)$ and $g_t(x_t, a_t, b_t)$ can be calculated in a similar manner and that these expressions are all $\mathcal{F}_t$-measurable.

From A's perspective, $g(\mathbf{u},\mathbf{v})$ is a reward and so A seeks an optimal $\mathcal{F}_t$-adapted policy to maximize this reward. From B's perspective, $g(\mathbf{u},\mathbf{v})$ is a cost and so B seeks an optimal $\mathcal{F}_t$-adapted policy to minimize this cost. Let $J_0(x_0)$ denote the optimal (equilibrium) value of this zero-sum game. Then a classic result due to Shapley [65] states that $J_0(x_0)$ exists and satisfies

$$J_0(x_0) = \sup_{\mathbf{u}\in\mathcal{A}^{\mathbb{F}}} \inf_{\mathbf{v}\in\mathcal{B}^{\mathbb{F}}} \mathbb{E}_0[g(\mathbf{u},\mathbf{v})] = \inf_{\mathbf{v}\in\mathcal{B}^{\mathbb{F}}} \sup_{\mathbf{u}\in\mathcal{A}^{\mathbb{F}}} \mathbb{E}_0[g(\mathbf{u},\mathbf{v})] \tag{4.2.4}$$

where we have used $\mathbb{E}_t[\cdot]$ to denote $\mathbb{E}[\cdot|\mathcal{F}_t]$. We assume the optimal game value $J_0(x_0)$ is bounded in the following discussion. Let $J_t(x_t)$ denote the time-t optimal value function of the game and note that we can compute it recursively according to

$$J_T(x_T) = g_T(x_T) \tag{4.2.5}$$

$$J_t(x_t) = \inf_{v_t} \sup_{u_t} \left\{ g_t(x_t,u_t,v_t) + \mathbb{E}_t\big[J_{t+1}(f(x_t,u_t,v_t,\omega_{t+1}))\big] \right\}, \quad t = 0,...,T-1 \tag{4.2.6}$$

with dynamics in (4.2.1).

4.2.1 Obtaining Dual Bounds for the Zero-Sum Game

If we fix B's policy to be $\hat{\mathbf{v}} \in \mathcal{B}^{\mathcal{F}}$ (which may be sub-optimal from B's perspective), A's problem reduces to

$$J_0(x_0;\hat{\mathbf{v}}) = \sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{F}}} \mathbb{E}_0[g(\mathbf{a},\hat{\mathbf{v}})]. \tag{4.2.7}$$

where we have restricted A to choosing a pure policy rather than a mixed policy. This presents no problem since it is well known that if B's policy is fixed then A will have an optimal pure[3] policy. Problem (4.2.7) can be solved via the following DP:

$$J_T(x_T; \hat{\mathbf{v}}) = g_T(x_T) \tag{4.2.8}$$

$$J_t(x_t; \hat{\mathbf{v}}) = \sup_{a_t} \Big\{ g_t(x_t, a_t, \hat{v}_t) + \mathbb{E}_t\big[J_{t+1}(f(x_t, a_t, \hat{v}_t, \omega_{t+1}))\big] \Big\}, \quad t = 0, ..., T-1 \tag{4.2.9}$$

with dynamics given by (4.2.1) and the understanding that b_t is the (generally random) action obtained when B follows $\hat{\mathbf{v}}$. Note that the optimal solution, $J_0(x_0; \hat{\mathbf{v}})$, obtained from (4.2.8) and (4.2.9) is an upper bound on the value of the game, $J_0(x_0)$. Moreover we have equality, i.e. $J_0(x_0; \hat{\mathbf{v}}) = J_0(x_0)$, if $\hat{\mathbf{v}} = \mathbf{v}^*$ where $\mathbf{v}^*$ is the optimal policy of B obtained from solving (4.2.5) and (4.2.6). In general, however, it is hard to even solve for $J_0(x_0; \hat{\mathbf{v}})$ but we can use the duality methods of BSS (and outlined in Section 4.2.2 below) to find an upper bound, say $\overline{J}_0(x_0; \hat{\mathbf{v}}, z(\hat{\mathbf{v}}))$, on $J_0(x_0; \hat{\mathbf{v}})$ where $z(\hat{\mathbf{v}})$ is a dual-feasible penalty in the language of BSS.

By a similar argument we can find a lower bound, $\underline{J}_0(x_0; \hat{\mathbf{u}}, z(\hat{\mathbf{u}}))$, on $J_0(x_0; \hat{\mathbf{u}})$, the optimal solution to B's problem when A's policy is fixed at $\hat{\mathbf{u}}$. Of course we also have $J_0(x_0; \hat{\mathbf{u}}) \leq J_0(x_0)$ since in general A is free to choose any feasible policy in $\mathcal{A}^{\mathbb{F}}$. Combining these observations we have

$$\underline{J}_0(x_0; \hat{\mathbf{u}}, z(\hat{\mathbf{u}})) \leq J_0(x_0; \hat{\mathbf{u}}) \leq J_0(x_0) \leq J_0(x_0; \hat{\mathbf{v}}) \leq \overline{J}_0(x_0; \hat{\mathbf{v}}, z(\hat{\mathbf{v}})). \tag{4.2.10}$$

The outer inequalities in (4.2.10) follow from the weak duality of BSS. Moreover there ex-

[3] This is related to the so-called *principle of indifference*; see Part II, Chapter 2 of Ferguson [32].

ist dual-feasible penalties, $z^*(\hat{\mathbf{u}})$ and $z^*(\hat{\mathbf{v}})$ say, for which these outer inequalities become equalities. This is strong duality:

$$\underline{J}_0(x_0; \hat{\mathbf{u}}, z^*(\hat{\mathbf{u}})) = J_0(x_0; \hat{\mathbf{u}}) \leq J_0(x_0) \leq J_0(x_0; \hat{\mathbf{v}}) = \overline{J}_0(x_0; \hat{\mathbf{v}}, z^*(\hat{\mathbf{v}})). \tag{4.2.11}$$

Finally, and as stated earlier if we choose $\hat{\mathbf{u}} = \mathbf{u}^*$ and $\hat{\mathbf{v}} = \mathbf{v}^*$ then the inequalities in (4.2.11) become equalities. In the next subsection we outline the weak and strong duality results of BSS that can be used to justify these statements.

4.2.2 Constructing Dual Bounds for Player A's Dynamic Program

We will focus[4] on the problem of constructing an upper bound for the value function, $J_0(x_0; \hat{\mathbf{v}})$, that is the solution of (4.2.8) and (4.2.9), which we recall is the DP that A must solve when B's policy is fixed at $\hat{\mathbf{v}}$. We can obtain lower bounds on $J_0(x_0; \hat{\mathbf{u}})$ in a similar manner.

We say that a filtration $\mathbb{G} = \{\mathcal{G}_0, ..., \mathcal{G}_T\}$ is a relaxation of the filtration $\mathbb{F}$ if, for each t, $\mathcal{F}_t \subseteq \mathcal{G}_t \subseteq \mathcal{F}$. For example, the perfect information filtration $\mathbb{I} = (\mathcal{I}_0, ..., \mathcal{I}_T)$ is obtained by taking $\mathcal{I}_t = \mathcal{F}$ for all t. Let $\mathcal{A}^{\mathbb{G}}$ represent the set of $\mathcal{G}_t$-adapted policies for A. Clearly we have $\mathcal{A}^{\mathbb{F}} \subseteq \mathcal{A}^{\mathbb{G}}$.

Let $\mathcal{Z}$ denote the set of functions, $z(\mathbf{a}, \hat{\mathbf{v}}, \omega)$, that depend on A's policy, $\mathbf{a}$, B's fixed (and mixed) policy, $\hat{\mathbf{v}}$, and the random noise vector $\omega := (\omega_0, ..., \omega_{T-1})$. BSS defines the set,

[4]BSS can be consulted for all statements in this section regarding dual bounds, weak and strong duality, and the construction of dual feasible penalties.

$\mathcal{Z}_{\mathbb{F}}$, of dual feasible penalties (for A's dynamic program) to be those functions that do not penalize A's $\mathcal{F}_t$-adapted policies. In particular,

$$\mathcal{Z}_{\mathbb{F}} = \{z \in \mathcal{Z} : \mathbb{E}_0[z(\mathbf{a}, \hat{\mathbf{v}}, \omega)] \leq 0 \text{ for all } \mathbf{a} \in \mathcal{A}^{\mathbb{F}}\}. \tag{4.2.12}$$

We then have the following weak duality result which follows immediately from the definition of dual feasibility in (4.2.12).

Proposition 4.2.1. ***(Weak Duality)*** *For any dual feasible penalty, $z \in \mathcal{Z}^{\mathbb{F}}$, we have*

$$J_0(x_0; \hat{\mathbf{v}}) \leq \sup_{\mathbf{a} \in \mathcal{A}^{\mathbb{G}}} \mathbb{E}_0[g(\mathbf{a}, \hat{\mathbf{v}}) - z(\mathbf{a}, \hat{\mathbf{v}}, \omega)] =: \overline{J}_0(x_0; \hat{\mathbf{v}}, z(\hat{\mathbf{v}})). \tag{4.2.13}$$

Therefore any dual feasible penalty and information relaxation provides an upper bound on the optimal value of A's dynamic program $J_0(x_0; \hat{\mathbf{v}})$. For a given information relaxation, we can optimize the upper bound by minimizing it over the set of dual feasible penalties. This yields the dual problem of player A's dynamic program:

$$\inf_{z \in \mathcal{Z}_{\mathbb{F}}} \left\{ \sup_{\mathbf{a} \in \mathcal{A}^{\mathbb{G}}} \mathbb{E}_0[g(\mathbf{a}, \hat{\mathbf{v}}) - z(\mathbf{a}, \hat{\mathbf{v}}, \omega)] \right\}. \tag{4.2.14}$$

We then have the following strong duality result:

Theorem 4.2.2. ***(Strong Duality)*** *If the primal problem (4.2.7) for player* A *is bounded, then the dual problem (4.2.14) has an optimal solution, $z^* \in \mathcal{Z}_F$, and there is no duality gap* *so that*

$$J_0(x_0;\hat{\mathbf{v}}) = \inf_{z\in\mathcal{Z}_{\mathbb{F}}} \left\{ \sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{G}}} \mathbb{E}_0[g(\mathbf{a},\hat{\mathbf{v}}) - z(\mathbf{a},\hat{\mathbf{v}},\omega)] \right\}. \tag{4.2.15}$$

An optimal penalty is therefore a dual feasible penalty, $z^* \in \mathcal{Z}_F$, for which $J_0(x_0;\hat{\mathbf{v}}) = \sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{G}}} \mathbb{E}_0[g(\mathbf{a},\hat{\mathbf{v}}) - z^*(\mathbf{a},\hat{\mathbf{v}},\omega)]$. We can construct such a z^* according to

$$z^*(\mathbf{a},\hat{\mathbf{v}},\omega) = \sum_{t=0}^{T-1} \left(\mathbb{E}\big[J_{t+1}(x_{t+1};\hat{\mathbf{v}})\big|\mathcal{G}_t\big] - \mathbb{E}\big[J_{t+1}(x_{t+1},\hat{\mathbf{v}})\big|\mathcal{F}_t\big] \right). \tag{4.2.16}$$

If we use a dual optimal penalty z^* together with the perfect information relaxation $\mathbb{G} = \mathbb{I}$, then (4.2.15) reduces to

$$\begin{aligned} J_0(x_0;\hat{\mathbf{v}}) &= \sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{I}}} \mathbb{E}_0[g(\mathbf{a},\hat{\mathbf{v}}) - z^*(\mathbf{a},\hat{\mathbf{v}},\omega)] \\ &= \mathbb{E}_0\left[\sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{I}}} \left\{ g(\mathbf{a},\hat{\mathbf{v}}) - z^*(\mathbf{a},\hat{\mathbf{v}},\omega) \right\} \right] \end{aligned} \tag{4.2.17}$$

and in fact the expectation in (4.2.17) is unnecessary since it can be shown that

$$J_0(x_0;\hat{\mathbf{v}}) = \sup_{\mathbf{a}\in\mathcal{A}^{\mathbb{I}}} \left\{ g(\mathbf{a},\hat{\mathbf{v}}) - z^*(\mathbf{a},\hat{\mathbf{v}},\omega) \right\}$$

almost surely. In general, however, the $J_t(x_t;\hat{\mathbf{v}})$'s are hard to compute and so it might be not possible to find an optimal penalty, z^*. Instead we can use an approximation, $\hat{J}_t(x_t;\hat{\mathbf{v}})$, to compute a dual feasible[5] penalty, $\hat{z}(\mathbf{a},\hat{\mathbf{v}},\omega)$, using (4.2.16) but with J_t replaced by $\hat{J}_t$. We can then obtain a valid and hopefully good upper bound as follows. Assuming again that

[5] It can be shown that a penalty, $\hat{z}(\mathbf{a},\hat{\mathbf{v}},\omega)$, constructed from (4.2.16) using $\hat{J}_t(x_t;\hat{\mathbf{v}})$ instead of J_t is indeed a dual feasible penalty.

$\mathbb{G} = \mathbb{I}$, we know from weak duality that

$$J_0(x_0; \hat{\mathbf{v}}) \leq \overline{J}_0(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) = \mathbb{E}_0\left[\sup_{\mathbf{a} \in \mathcal{A}^{\mathbb{I}}} \left\{ g(\mathbf{a}, \hat{\mathbf{v}}) - \hat{z}(\mathbf{a}, \hat{\mathbf{v}}, \omega) \right\}\right]. \tag{4.2.18}$$

The expectation in (4.2.18) therefore yields an upper bound on $J_0(x_0; \hat{\mathbf{v}})$ and it's easily estimated via Monte-Carlo simulation: we simply generate N sample paths $\omega^{(i)} = (\omega_0^{(i)}, ..., \omega_{T-1}^{(i)})$ for $i = 1, ..., N$, and solve the deterministic "inner problem"

$$\overline{J}_0^{(i)}(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) := \sup_{\mathbf{a} \in \mathcal{A}^{\mathbb{I}}} \left\{ g(\mathbf{a}, \hat{\mathbf{v}}) - \hat{z}(\mathbf{a}, \hat{\mathbf{v}}, \omega^{(i)}) \right\}. \tag{4.2.19}$$

along each path. The average, $\frac{1}{N}\sum_{i=1}^{N} \overline{J}_0^{(i)}(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}}))$, is then an unbiased estimate of the upper bound, $\overline{J}_0(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}}))$.

A Simple Example

We first consider a single-period two-person zero-sum game where the payoff is defined by an $m \times n$ matrix, R. In this game players A and B simultaneously select a row, a, and a column, b, respectively, after which B pays A the value $R_{a,b}$. A mixed policy for player A is an $m \times 1$ vector $u = [u^{(1)} \ ... \ u^{(m)}]'$ of probabilities that sum to 1 such that A chooses the a-th row with probability $u^{(a)}$. Similarly, we let $v = [v^{(1)} \ ... \ v^{(n)}]'$ represent player B's mixed policy. The expected payoff to A is then $u'Rv$.

Consider now a 2-period dynamic game defined by the following three matrix games:

$$R^{(1)} = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix}, \; R^{(2)} = \begin{bmatrix} 8 & 15 \\ 10 & 12 \end{bmatrix} \text{ and } R^{(3)} = \begin{bmatrix} -8 & -10 \\ 3 & -11 \end{bmatrix}.$$

The state variable $x_t \in \{i \,|\, i = 1, 2, 3\}$ determines the game $R^{(x_t)}$ that is played at time t, and $\tilde{g}(x_t, a_t, b_t)$ is then the payoff of that game when A and B choose the a_t-th row and b_t-th column, respectively at time t. The game begins at $t = 0$ with A and B playing $R^{(1)}$ so the initial state is $x_0 = 1$. The next matrix game to be played at time $t = 1$ is determined by A and B's actions and a transition probability, $p(x_t, x_{t+1}, a_t, b_t)$. These transition probabilities are determined by the a_t-th row and b_t-th column element in the matrices $P_{x_t, x_{t+1}}$ which are defined as:

$$P_{1,2} = \begin{bmatrix} 0.7 & 0.55 \\ 0.4 & 0.5 \end{bmatrix} \text{ and } P_{1,3} = \begin{bmatrix} 0.3 & 0.45 \\ 0.6 & 0.5 \end{bmatrix}.$$

Note of course that $P_{1,2} + P_{1,3}$ is equal to the matrix with 1 in every entry. For example, if A chooses the second row ($a_0 = 2$) and B chooses the first column ($b_0 = 1$) when playing $R^{(1)}$ at time $t = 0$, B pays A amount of $\tilde{g}(x_0 = 1, a_0 = 2, b_0 = 1) = 6$. At time $t = 1$ A and B will then play game $R^{(2)}$ with probability 0.4, or game $R^{(3)}$ with probability 0.6. We assume the random variable, ω_1, which drives the state transition is uniformly distributed so that the state evolution can be written as

$$x_1 = f_0(x_0, a_0, b_0, \omega_1) = 2 \cdot \mathbf{1}_{\{\omega_1 \leq p(1,2,a_0,b_0)\}} + 3 \cdot (1 - \mathbf{1}_{\{\omega_1 \leq p(1,2,a_0,b_0)\}}) \tag{4.2.20}$$

where $\mathbf{1}_{\{.\}}$ denotes the indicator function. The optimal policies and value functions for this

Table 4.1: Optimal policies and game values

t	x_t	u_t^*	v_t^*	$J_t(x_t)$
0	1	$[0.5 \quad 0.5]'$	$[0.75 \quad 0.25]'$	5
1	2	$[0 \quad 1]'$	$[1 \quad 0]'$	10
1	3	$[1 \quad 0]'$	$[0 \quad 1]'$	-10

2-period zero-sum game are easily calculated using standard techniques and are given in Table 4.1.

We can compute an upper bound for the fair value of the game by fixing B's policy and then solving A's corresponding DP. Towards this end, suppose B fixes his policy at $\hat{\mathbf{v}} = (\hat{v}_0 = [0.6 \quad 0.4]', \hat{v}_1 = v_1^*(x_1))$, the game reduces to A's dynamic program with optimal solutions

$$J_1(x_1; \hat{\mathbf{v}}) = J_1(x_1) \tag{4.2.21}$$

$$J_0(x_0; \hat{\mathbf{v}}) = \max \left\{ R\hat{v}_0 = \begin{bmatrix} 4.4 \\ 5.6 \end{bmatrix} \right\} = 5.6 \tag{4.2.22}$$

$$\text{where } R = R^{(1)} + J_1(2)P_{1,2} + J_1(3)P_{1,3} = \begin{bmatrix} 6 & 2 \\ 4 & 8 \end{bmatrix}.$$

In (4.2.21) because B fixes $\hat{v}_1 = v_1^*(x_1)$ at $t = 1$, $J_1(x_1; \hat{\mathbf{v}})$ is then equal to $J_1(x_1)$ given in Table 4.1. In (4.2.22) at time $t = 0$, R is the expected total payoff matrix, and thus A's optimal game value under B's fixed policy $\hat{\mathbf{v}}$ is $J_0(x_0; \hat{\mathbf{v}}) = 5.6$. We note that $J_0(x_0; \hat{\mathbf{v}})$ is an upper bound on the value of game, $J_0(x_0) = 5$.

Weak Duality

For more complex games we would not be able to compute $J_0(x_0; \hat{\mathbf{v}})$ but we can instead use the information relaxation approach to compute an upper bound on $J_0(x_0; \hat{\mathbf{v}})$. Suppose then we use a perfect information relaxation, and construct the penalty $\hat{z} = \hat{J}_1(x_1; \hat{\mathbf{v}}) - \mathbb{E}\big[\hat{J}_1(x_1; \hat{\mathbf{v}})\big|\mathcal{F}_0\big]$ where $\hat{J}_1(x_1; \hat{\mathbf{v}})$ is an approximate value for player A's DP. For illustrative[6] purposes, we take $\hat{J}_1(x_1; \hat{\mathbf{v}}) = 8 \cdot \mathbf{1}_{\{x_1=2\}} - 8 \cdot \mathbf{1}_{\{x_1=3\}}$.

We estimate the dual bound by simulating 10,000 dual problem instances, i.e. by simulating 10,000 values of ω_1, solving the deterministic dual inner problem (4.2.19) for each instance, and then taking the average of the optimal objective functions. Our simulation yielded an estimated upper bound $\overline{J}_0(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) = 5.82$ with a standard error of 0.02. Note that our numerical results, i.e. $J_0(x_0) < J_0(x_0; \hat{\mathbf{v}}) < \overline{J}_0(x_0; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}}))$, are consistent with weak duality in (4.2.10).

Strong Duality

If instead we constructed the dual optimal penalty $z^* = J_1(x_1; \hat{\mathbf{v}}) - \mathbb{E}\big[J_1(x_1; \hat{\mathbf{v}})\big|\mathcal{F}_0\big]$ according to (4.2.16) using $J_1(x_1; \hat{\mathbf{v}})$ in (4.2.21), each dual inner problem (4.2.19) yields an upper bound $\overline{J}_0(x_0; \hat{\mathbf{v}}, z^*(\hat{\mathbf{v}})) = 5.6 = J_0(x_0; \hat{\mathbf{v}})$. We have therefore demonstrated strong duality (for B's fixed policy, $\hat{\mathbf{v}}$) as given in (4.2.11).

Note also that if we fixed B's policy at $\mathbf{v}^*$ and repeated the numerical calculations above, then we would obtain $J_0(x_0) = J_0(x_0; \mathbf{v}^*) = \overline{J}_0(x_0; \mathbf{v}^*, z^*(\mathbf{v}^*))$. Corresponding lower bounds

[6]This is the game value at time $t = 1$ if A always chooses first row and B always chooses first column when playing $R^{(x_1)}$.

for $J_0(x_0)$ can be obtained by fixing A's policy and solving or lower bounding player B's resulting DP.

4.3 Dual Bounds for Stochastic Shortest Path Games

In this section we extend the information relaxation approach to more general stochastic shortest-path (SSP) zero-sum games. Note that infinite horizon discounted games can be modeled as SSP zero-sum games so the formulation is quite general and indeed general conditions exist to guarantee that these games have an optimal value. Assuming these games (i) have an optimal value and (ii) have optimal policies under which the game terminates finitely almost surely, we can replicate the information relaxation methods of Section 4.2 to obtain valid dual bounds. Our discussion below assumes[7] that both (i) and (ii) hold.

One potential difficulty that arises in constructing dual bounds in this case is in simulating a dual sample path. In particular, how many time periods are there in such a dual path? In general the number of periods (before absorption into a terminal state) will depend on the policies employed by the two players but we don't know these policies in advance so how can we simulate such a path? Very recent work of Brown and Haugh [14] describes several approaches for addressing this problem. One such approach[8] is to use the Rogers [60] approach to finite horizon problems with the perfect information relaxation. Under his approach dual sample paths are simulated under a reference transition probability measure,

[7]We note that (i) and (ii) hold for the example that we consider in this section. In fact the game terminates after a finite number of periods almost surely regardless of what policies the players employ.

[8]There are better approaches; see Brown and Haugh [14] for details.

P^*, that is *action-independent.* In the context of our (generally infinite-horizon) SSP games, we can use such a P^* to generate dual sample paths as long as all such paths terminate after a finite number of periods P^*-almost surely. We also need to choose P^* so that there exists an optimal policies whose induced probability measure (over sample paths) is absolutely continuous with respect to the measure induced by P^*.

More specifically dual bounds are constructed as follows, where for ease of exposition we use the perfect information relaxation, $\mathbb{I}$. We generate the state path $x = (x_0, ..., x_T)$ using the action-independent transition matrix P^* where $p^*(x_t, x_{t+1})$ denotes the transition probability from state x_t to x_{t+1}. Note that while T is random, this does not present a problem in simulating a dual sample path because P^* is action-independent and so the realized value of T does not depend on the policies of the two players.

Because we used P^* to generate dual paths and not the natural transition matrix induced by the players' actions, it is necessary to multiply the various rewards and dual value functions by appropriate action-dependent Radon-Nikodym derivative terms; see Rogers [60] or Brown and Haugh [14]. In particular, suppose we fix B's policy at $\hat{\mathbf{v}}$ and use a dual feasible penalty, $\hat{z}$, that has been constructed from some approximated value function, $\hat{J}_t(x_t; \hat{\mathbf{v}})$. Then the dual inner problem on a P^*-simulated path that player A must solve has the form

$$\sup_{\mathbf{a} \in \mathcal{A}^{\mathbb{I}}} \left\{ \Lambda_T(\mathbf{a}) g_T(x_T) + \sum_{t=0}^{T-1} \Lambda_t(\mathbf{a}) \Big[g_t(x_t, a_t, \hat{v}_t) + \mathbb{E}\big[\hat{J}_{t+1}(x_{t+1}; \hat{\mathbf{v}}) \big| \mathcal{F}_t\big] - \phi(x_t, x_{t+1}, a_t, \hat{v}_t) \hat{J}_{t+1}(x_{t+1}; \hat{\mathbf{v}}) \Big] \right\} \tag{4.3.1}$$

where

$$\Lambda_t(\mathbf{a}) := \prod_{\tau=0}^{t-1} \phi(x_\tau, x_{\tau+1}, a_\tau, \hat{v}_\tau) \tag{4.3.2}$$

$$\phi(x_t, x_{t+1}, a_t, \hat{v}_t) = \frac{\sum_{\hat{b}_t \in \mathcal{B}(x_t)} \hat{v}_t^{(\hat{b}_t)} p(x_t, x_{t+1}, a_t, \hat{b}_t)}{p^*(x_t, x_{t+1})} \tag{4.3.3}$$

and where the numerator in the Radon-Nikodym derivative term (4.3.3) is the probability that the game transitions from x_t to x_{t+1} under A's action a_t and B's mixed policy $\hat{v}_t$. Note also that the expectations in (4.3.1) are taken with respect to the action-dependent transitions $p(\cdot)$. The objective in (4.3.1) is fully deterministic and can be solved recursively according to

$$\begin{aligned}
\overline{J}_T(x_T; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) &= g_T(x_T) \\
\overline{J}_t(x_t; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) &= \sup_{a_t} \Big\{ g_t(x_t, a_t, \hat{v}_t) + \mathbb{E}\big[\hat{J}_{t+1}(x_{t+1}; \hat{\mathbf{v}}) \big| \mathcal{F}_t\big] \\
&+ \phi(x_t, x_{t+1}, a_t, \hat{v}_t) \Big(\overline{J}_{t+1}(x_{t+1}; \hat{\mathbf{v}}, \hat{z}(\hat{\mathbf{v}})) - \hat{J}_{t+1}(x_{t+1}; \hat{\mathbf{v}}) \Big) \Big\}, \ t = 0, ..., T-1.
\end{aligned}$$

A dual bound is then estimated by simulating N sample paths under P^*, solving (4.3.1) along each of these paths, and then taking the average of the optimal objective functions.

Example: A Stochastic Shortest-Path Matrix Game

We construct an SSP game using the two matrix games

$$R^{(1)} = \begin{bmatrix} 4 & 0 \\ 1 & 3 \end{bmatrix} \text{ and } R^{(2)} = \begin{bmatrix} 0 & -5 \\ -4 & 1 \end{bmatrix}. \tag{4.3.4}$$

The state variable is x_t and a value of $x_t \in \{1, 2\}$ implies that $R^{(x_t)}$ is played at time t. The state $x_t = 3$ denotes a cost-free absorbing termination state and the transition probability matrices are

$$P_{1,1} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.5 \end{bmatrix} \quad P_{1,2} = \begin{bmatrix} 0 & 0.4 \\ 0.4 & 0 \end{bmatrix} \quad P_{1,3} = \begin{bmatrix} 0.7 & 0.6 \\ 0.6 & 0.5 \end{bmatrix}$$
$$P_{2,1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix} \quad P_{2,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0.5 \end{bmatrix} \quad P_{2,3} = \begin{bmatrix} 0.5 & 1 \\ 1 & 0.5 \end{bmatrix}.$$

We note $\sum_{j=1}^{3} P_{i,j}$ for $i = 1, 2$ is a matrix with every entry equal to 1. This of course must be the case. Suppose now, for example, that $x_0 = 1$ so that A and B initially play $R^{(1)}$. If A chooses the second row ($a_0 = 2$) and B chooses the first column ($b_0 = 1$), then B pays A 1 unit and at time $t = 1$ they will play $R^{(2)}$ with probability 0.4, and enter the absorbing state with probability 0.6.

Table 4.2: Suboptimal policies for the SSP matrix game

	A's policy			B's policy	
	$u(x_t = 1)$	$u(x_t = 2)$		$v(x_t = 1)$	$v(x_t = 2)$
$\hat{\mathbf{u}}^{(1)}$	[0.7500 0.2500]'	[0.2500 0.7500]'	$\hat{\mathbf{v}}^{(1)}$	[0.7500 0.2500]'	[0.2500 0.7500]'
$\hat{\mathbf{u}}^{(2)}$	[0.4269 0.5731]'	[0.3970 0.6030]'	$\hat{\mathbf{v}}^{(2)}$	[0.5315 0.4685]'	[0.4897 0.5103]'
$\hat{\mathbf{u}}^{(3)}$	[0.4135 0.5865]'	[0.3996 0.6004]'	$\hat{\mathbf{v}}^{(3)}$	[0.5220 0.4780]'	[0.4995 0.5005]'

In order to compute dual bounds we begin with the two sub-optimal policies, $\hat{\mathbf{u}}^{(1)}$ and

$\hat{\mathbf{v}}^{(1)}$, for players A and B, respectively. These policies are defined in the first row of Table 4.2. We can simulate these policies to obtain estimates of the game value, $J_0(x_0, \hat{\mathbf{u}}^{(1)}, \hat{\mathbf{v}}^{(1)})$ as a function of x_0. We use these estimates as our approximate value function, $\hat{J}_t(x_t)$, that we use to construct a dual feasible penalty. Finally, we can then estimate dual lower and upper bounds on the optimal game value by simulating sample paths and solving the corresponding dual problem instances as in (4.3.1). The sample paths (for estimating the dual bounds) were simulated using the action-independent transition matrix

$$P^* = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.3.5}$$

The corresponding game value and dual bounds (mean and standard errors) are displayed in the first two rows of Table 4.3 where $J_0(x_0, \hat{\mathbf{u}}^{(i)}, \hat{\mathbf{v}}^{(i)})$ are estimated using 10^6 simulated paths and $\overline{J}_0(x_0; \hat{\mathbf{v}}^{(i)}, \hat{z}(\hat{\mathbf{v}}^{(i)}))$ and $\underline{J}_0(x_0; \hat{\mathbf{u}}^{(i)}, \hat{z}(\hat{\mathbf{u}}^{(i)}))$ are estimated using 10^3 paths.

The dual lower and upper bounds are quite far apart and we expect this is because $\hat{\mathbf{u}}^{(1)}$ and $\hat{\mathbf{v}}^{(1)}$ are far from optimal. We therefore perform a policy iteration to find new and improved suboptimal policies $(\hat{\mathbf{u}}^{(i+1)}, \hat{\mathbf{v}}^{(i+1)})$ based on $J_0(x_0, \hat{\mathbf{u}}^{(i)}, \hat{\mathbf{v}}^{(i)})$ for $i = 1, 2$. These new policies are also displayed in Table 4.2. We can also estimate their game values and use them to construct dual feasible penalties and dual bounds. These game values and dual bounds are displayed in Table 4.3. As expected, we see that the duality gap, i.e. the difference between the dual upper and lower bounds, decreases with each policy iteration for each value of x_0. In fact, we can use our dual bounds to see that the estimates $J_0(1) = 2.025$ and $J_0(2) = -1.99$

are very accurate[9] with high probability. Note also that we did not actually compute the optimal game value for this problem.

Table 4.3: Game values, dual upper and lower bounds for the SSP matrix game. Means and standard errors (in parentheses) are reported for each of the two possible initial states, $x_0 = 1$ and $x_0 = 2$.

Suboptimal	$x_0 = 1$			$x_0 = 2$		
Polices	$\underline{J}_0(x_0; \hat{\mathbf{u}}^{(i)})$	$J_0(x_0, \hat{\mathbf{u}}^{(i)}, \hat{\mathbf{v}}^{(i)})$	$\overline{J}_0(x_0, \hat{\mathbf{v}}^{(i)})$	$\underline{J}_0(x_0, \hat{\mathbf{u}}^{(i)})$	$J_0(x_0, \hat{\mathbf{u}}^{(i)}, \hat{\mathbf{v}}^{(i)})$	$\overline{J}_0(x_0, \hat{\mathbf{v}}^{(i)})$
$\hat{\mathbf{u}}^{(1)}, \hat{\mathbf{v}}^{(1)}$	-0.2407	3.0123	3.8266	-3.2084	-1.4342	-0.3932
	(0.0246)	(0.0023)	(0.0108)	(0.0201)	(0.0030)	(0.0230)
$\hat{\mathbf{u}}^{(2)}, \hat{\mathbf{v}}^{(2)}$	1.9212	2.0242	2.0953	-2.0366	-1.9969	-1.9445
	(0.0014)	(0.0031)	(0.0005)	(0.0010)	(0.0030)	(0.0007)
$\hat{\mathbf{u}}^{(3)}, \hat{\mathbf{v}}^{(3)}$	2.0221	2.0225	2.0228	-1.9976	-1.9973	-1.9972
	(0.00001)	(0.0029)	(0.00001)	(0.00001)	(0.0030)	(0.00001)

4.4 Conclusions and Further Research

The main contribution of this chapter is to generalize the work of Beveridge and Joshi [12] on optimal stopping games to dynamic zero-sum games. We used the general results of BSS and Rogers [60] on information relaxations for general dynamic programs to do this. While the results were straightforward to prove, there are many interesting applications including pursuit-evasion games, heads-up poker and many two-person computer games. While we considered relatively simple numerical examples in this chapter, the ultimate value of this work will be in applying it successfully to more realistic and interesting applications.

There are several potential problems that arise when we consider these more complex applications. For example, many zero-sum games such as heads-up poker are more naturally

[9]We can of course use standard Monte-Carlo methods to construct conservative confidence intervals for the optimal game values $J_0(1)$ and $J_0(2)$.

studied in so-called[10] *extensive form* whereas our dual formulation requires the games to be modeled in *strategic form*. While it is certainly possible to go back and forth between these two forms, it is not clear how much work is required to do so. Nor is it clear that sub-optimal policies that have been designed for the extensive form of the game are easily handled in strategic form. Computing easy-to-use dual penalties from these sub-optimal strategies may also be problematic.

Another potential difficulty relates to the issue of private information. Our analysis has assumed the players do not have private sources of information but in many games, however, private information does exist. For example, in heads-up poker each player knows his two "hole" cards but does not know his opponent's "hole" cards until the end of the game if ever. From a theoretical perspective, it should be easy to handle this complication but it may well raise practical problems that are difficult to overcome with limited computing resources.

There are many other interesting complications related to learning and bounded rationality that would also be interesting to explore. We hope to address some of these problems in future research.

[10]See, for example, Ferguson [32] for a detailed discussion on zero-sum games as well as the strategic and extensive forms of these games.

Bibliography

[1] C. Alexander. Principal component models for generating large garch covariance matrices. *Economic Notes*, 31(2):337–359, 2002.

[2] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape function. *Quantitative Finance*, 10:143–157, 2010.

[3] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math Programming*, Ser. B 95:3–51, 2003.

[4] R. Almgren. Optimal trading with stochastic liquidity and volatility. *Forthcoming in SIAM Journal on Financial Mathematics*, 2012.

[5] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, pages 5–39, 2001.

[6] R. Almgren and N. Chriss. Value under liquidation. *Risk Magazine*, 3:61–63, Winter 2000 / 2001.

[7] L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional american options. *Management Science*, 50(9):1222–1234, 2004.

[8] C. Bender. Primal and dual pricing of multiple exercise options in continuous time. *SIAM Journal of Financial Mathematics*, 2:562–586, 2011.

[9] C. Bender, J. Schoenmakers, and J. Zhang. Dual representations for general multiple stopping problems. Technical report, 2011.

[10] D. Bertsimas, P. Hummel, and A.W. Lo. Optimal control of execution costs for portfolios. *Computing in Science & Engineering*, pages 40–53, 1999.

[11] D. Bertsimas and A.W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1:1–50, 1998.

[12] C. Beveridge and M. Joshi. Monte carlo bounds for game options including convertible bonds. *Management Science*, 57(5):960–974, 2011.

[13] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.

[14] D.B. Brown and M.B. Haugh. Information relaxation bounds for infinite horizon stochastic control. Technical report, 2014.

[15] D.B. Brown and J.E. Smith. Dynamic portfolio optimization with transaction costs: heuristics and dual bounds. *Management Science*, 57(10):1752–1770, 2011.

[16] D.B. Brown, J.E. Smith, and P. Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4):785–801, 2010.

[17] S. Buti, B. Rindi, and I.M. Werner. Dark pool trading strategies. Technical report, 2011.

[18] S. Chandramouli and M. Haugh. A unified approach to multiple stopping and duality. *Operations Research Letters*, 40:258–264, 2012.

[19] C. O' Cinneide, B. Scherer, and X. Xu. Pooling trades in a quantitative investment process. *The Journal of Portfolio Management*, 32(4):33–43, 2006.

[20] G. Constantinides. Optimal stock trading with personal taxes: Implications for prices and the abnormal january returns. *Journal of Financial Economics*, 13(1):65–89, 1984.

[21] R. Cont and A. Kukanov. Optimal order placement in limit order markets. Technical report, 2013.

[22] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 58:549–563, 2010.

[23] R.M. Dammon, C.S. Spatt, and H.H. Zhang. Optimal consumption and investment with capital gains taxes. *Review of Financial Studies*, 14(3):583–616, 2001.

[24] R.M. Dammon, C.S. Spatt, and H.H. Zhang. Diversification and capital gains taxes with multiple risky assets. Technical report, 2002.

[25] R.M. Dammon, C.S. Spatt, and H.H. Zhang. Optimal asset location and allocation with taxable and tax-deferred investing. *Journal of Finance*, 59:999–1037, 2004.

[26] MHA Davis and I. Karatzas. A deterministic approach to optimal stopping. *Probability, Statistics and Optimisation (ed. FP Kelly). NewYork Chichester: John Wiley & Sons Ltd*, pages 455–466, 1994.

[27] MHA Davis and M. Zervos. A new proof of the discrete-time lqg optimal control theorems. *IEEE Transactions on Automatic Control*, 40:1450–1453, 1995.

[28] V. DeMiguel and R. Uppal. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science*, 51(2):277–290, 2005.

[29] V.V. Desai, V.F. Farias, and C.C. Moallemi. Bounds for markov decision processes. Technical report, 2011.

[30] P.H. Dybvig and H.K. Koo. Investment with taxes. Technical report, 1996.

[31] P. Ehling, M. Gallmeyer, S. Srivastava, S. Tompaidis, and C. Yang. Portfolio choice with capital gain taxation and the limited use of losses. Technical report, 2010.

[32] T.S. Ferguson. Game theory. `http://www.math.ucla.edu/~tom/Game_Theory/mat.pdf`.

[33] M.F. Gallmeyer, , and S. Srivastava. Arbitrage and the tax code. Technical report, 2010.

[34] M.F. Gallmeyer, R. Kaniel, and S. Tompaidis. Tax management strategies with multiple risky assets. *Journal of Financial Economics*, 80:243–291, 2006.

[35] L. Garlappi, V. Naik, and J. Slive. Portfolio selection with multiple assets and capital gains taxes. Technical report, 2001.

[36] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10:749–759, 2010.

[37] J. Gatheral and A. Schied. optimal trade execution under geometric brownian motion in the almgren and chriss framework. *International Journal of Theoretical and Applied Finance*, 14:353–368, 2011.

[38] P. Glasserman. *Monte carlo methods in financial engineering*, volume 53. Springer Verlag, 2004.

[39] O. Guéant. Permanent price impact can be nonlinear. Technical report, 2013.

[40] N.H Hakansson. Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica*, 38:587–607, 1970.

[41] M.B. Haugh and L. Kogan. Pricing american options: a duality approach. *Operations Research*, 52(2):258–270, 2004.

[42] M.B. Haugh and A.E.B. Lim. Linear-quadratic control and information relaxations. *Forthcoming in Operations Research Letters*, 2011.

[43] M.B. Haugh and C. Wang. Dynamic portfolio execution and information relaxations. *SIAM Journal on Financial Mathematics*, Forthcoming, 2013.

[44] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72:1247–1275, 2004.

[45] G. Huberman and W. Stanzl. Optimal liquidity trading. *Review of Finance*, 9:165–200, 2005.

[46] R. Huitema. Optimal portfolio execution using market and limit orders. Technical report, 2012.

[47] F. Jamshidian. The duality of optimal exercise and domineering claims: a doobcmeyer decomposition approach to the snell envelope. *Stochastics*, pages 27–60, 2007.

[48] P. Kratz and T. Schoneborn. Portfolio liquidation in dark pools in continuous time. Technical report, 2012.

[49] G. Lai, F. Margot, and N. Secomandi. An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research*, 58:564–582, 2010.

[50] G. Lai, M. X. Wang, S. Kekre, A. Scheller-Wolf, and N. Secomandi. Valuation of storage at a liquefied natural gas storage terminal. *Operations Research*, 59:602–616, 2011.

[51] A.E.B. Lim and P. Wimonkittiwat. Dynamic portfolio choice with market impact costs. *In Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, 2011.

[52] A. Lynch. Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability. *Journal of Financial Economics*, 62(1):67–130, 2001.

[53] N. Meinshausen and B.M. Hambly. Monte carlo methods for the valuation of multiple-exercise options. *Mathematical Finance*, 14(4):557–583, 2004.

[54] R.C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51:247–257, 1969.

[55] C. Moallemi and M. Sağlam. Dynamic portfolio choice with linear rebalancing rules. Technical report, 2011.

[56] A. Obizhaeva and J. Wang. Optimal trading strategy and supply / demand dynamics. Technical report, 2005.

[57] S.D. Patek. Stochastic shortest path games: theory and algorithms. Technical report, 1997.

[58] S. Predoiu, G. Shaikhet, and S. Shreve. Optimal execution in a general one-sided limit-order book. *SIAM Journal on Financial Mathematics*, 2:183–212, 2010.

[59] L.C.G. Rogers. Monte carlo valuation of american options. *Mathematical Finance*, 12(3):271–286, 2002.

[60] L.C.G. Rogers. Pathwise stochastic optimal control. *SIAM Journal on Control and Optimization*, 46:1116–1132, 2007.

[61] N. V. Sahinidis. *BARON 12.1.0: Global optimization of mixed-integer nonlinear programs,* User's Manual, 2013.

[62] P. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51:239–246, 1969.

[63] A. Schied, T. Schoneborn, and M. Tehranci. Optimal basket liquidation for cara investors is deterministic. *Applied Mathematical Finance*, 17:471–489, 2010.

[64] J. Schoenmakers. A pure martingale dual for multiple stopping. *Finance and Stochastics*, pages 1–16, 2010.

[65] L.S. Shapley. Stochastic games. *Proceedings of the natinal academy of sciences, Mathematics*, 39:1095–1100, 1953.

[66] R. A. Stubbs and D. Vandenbussche. Multi-portfolio optimization and fairness in allocation of trades. Technical report, 2009.

[67] I.B. Tahar, H.M. Soner, and N. Touzi. Merton problem with taxes: characterization, computation, and approximation. *SIAM Journal on Financial Mathematics*, 1(1):366–395, 2010.

[68] M. Tawarmalani and N. V. Sahinidis. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*, 103:225–249, 2005.

[69] B. Tóth, Y. Lempérière, C. Deremble, J. de Lataillade, J. Kockelkoren, and J.P. Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 2011.

[70] G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution with correlated supply / demand curves. Technical report, 2011.

Appendix A

Supplemental Content for Chapter 2

A.1 Generating LDS Scenarios for the RBH and Lookahead Policies

At each time period t, we generate $M = 100,000$ scenarios of time T (RBH) or time $\min(T, t+h)$ (h-period look-ahead) security prices and market state variables. Using the dynamics specified in (2.5.1) and (2.5.2) the distribution of $(\mathbf{p}_T, \mathbf{z}_T)$ or $(\mathbf{p}_{t+h}, \mathbf{z}_{t+h})$ conditional on the information available at time t is available in closed form. We can therefore generate time T or $t+h$ scenarios directly, i.e. in a single step, rather than having to simulate sceanarios at intermediate time points using (2.5.1) and (2.5.2).

We used quasi-Monte-Carlo [38] and low discrepancy sequences (LDS) instead of naive Monte-Carlo since LDS fill the "unit cube" more uniformly. In our numerical experiments we used an $(K+1)$-dimensional Sobol sequence where the first K dimensions were used to generate samples of the stock prices and the last dimension was used to generate samples for the market state. We produced these LDS using MATLAB's LDS functionality. We rejected the first 1000 points, retained every 101-st point thereafter, and also applied the so-called Matousek-Affine-Owen scrambling scheme.

A.2 The Exact-Tax Basis FUL Problem

The exact tax-basis FUL problem is less realistic than the corresponding LUL problem in that net tax losses in any period earn an immediate rebate under the FUL assumption. However, it is an easier problem and for this reason most of the literature on tax-aware portfolio optimization focuses on the FUL model. In this Section, we briefly describe how our LUL problem formulation is easily adapted to the FUL case.

A.2.1 The FUL Model and Problem Formulation

Under the FUL tax rule, the investor receives tax rebates if there are realized capital losses. The taxable capital gains at time t is then

$$g_t = \sum_{j=0}^{t-1} (\mathbf{p}_t - \mathbf{p}_j)'(\mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}) \tag{A.2-1}$$

which is no longer restricted to be non-negative. In particular, when $g_t < 0$ so that there are net capital losses, the agent pays a *negative* tax of τg_t, i.e. the agent receives an tax rebate that is available for immediate investment. Since it is optimal for the agent to realize any capital losses, we do not need l_t to track the accumulated unused losses. By substituting (A.2-1) into the budget constraint (2.2.7) and removing the LUL constraints (2.2.11)–(2.2.13), we can formulate the exact tax-basis FUL problem as follows:

$$\max_{(b_t,\mathbf{n}_{j,t})\in\mathcal{F}_t,\ t=0,\dots,T} \quad \mathbb{E}_0\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \tag{A.2-2}$$

$$\text{s.t.} \quad b_0 + \mathbf{p}_0'\mathbf{n}_{0,0} = w_0 \tag{A.2-3}$$

$$b_t + \sum_{j=0}^{t}\mathbf{p}_t'\mathbf{n}_{j,t} + \tau\sum_{j=0}^{t-1}(\mathbf{p}_t - \mathbf{p}_j)'(\mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}) = b_{t-1}r_0 + \sum_{j=0}^{t-1}\mathbf{p}_t'\mathbf{n}_{j,t-1} \qquad t \geq 1 \tag{A.2-4}$$

$$\mathbf{n}_{t,t} \geq \mathbf{0} \qquad t \geq 0 \tag{A.2-5}$$

$$\mathbf{n}_{j,t-1} \geq \mathbf{n}_{j,t} \geq \mathbf{0} \qquad t \geq 1,\ j < t \tag{A.2-6}$$

$$b_t \geq 0 \qquad t \geq 0 \tag{A.2-7}$$

and security price and market state dynamics.

As before, we let $\mathbf{x}_t := [b_t\ \mathbf{n}_{0,t}'\ \dots\ \mathbf{n}_{t,t}']$ denote the vector of time-t decision variables and let

$$\mathbb{X}_0 := \left\{\mathbf{x}_0 \mid \mathbf{x}_0 \in \mathcal{F}_0 \text{ satisfies constraints (A.2-3), (A.2-5) and (A.2-7) at } t=0\right\}$$

$$\mathbb{X}_t(\mathbf{x}_{t-1}, \mathbf{p}_{0:t}) := \left\{\mathbf{x}_t \mid \mathbf{x}_t \in \mathcal{F}_t \text{ satisfies constraints (A.2-4) to (A.2-7) at time t}\right\}, \quad t \geq 1.$$

denote the feasible set of trades at times 0 and $t > 0$, respectively.

A.2.2 Sub-Optimal Policies

While there are fewer variables and constraints in the exact tax-basis FUL problem when compared to the exact tax-basis LUL problem, it is still path-dependent and very challenging to solve to optimality. All the sub-optimal polices discussed in Section 2.3, however, can be easily adapted for the exact tax-basis FUL problem.

A.2.3 Dual Bound

We can use the information relaxations methods of Section 2.4 to construct unbiased upper bounds for the exact tax-basis FUL problem. After making the variable transformation $\hat{\mathbf{n}}_{j,t} := \mathbf{n}_{j,t-1} - \mathbf{n}_{j,t}$, the dual problem takes the form

$$\begin{aligned}
\max_{b_t,\ \mathbf{n}_{t,t},\ \hat{\mathbf{n}}_{j,t}} \quad & \frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) && && \text{(A.2-8)} \\
\text{s.t.} \quad & b_0 + \mathbf{p}_0' \mathbf{n}_{0,0} = w_0 \\
& b_t + \mathbf{p}_t' \mathbf{n}_{t,t} + \tau \sum_{j=0}^{t-1} (\mathbf{p}_t - \mathbf{p}_j)' \hat{\mathbf{n}}_{j,t} = b_{t-1} r_0 + \sum_{j=0}^{t-1} \mathbf{p}_t' \hat{\mathbf{n}}_{j,t} && t \geq 1, \\
& \mathbf{n}_{t,t} - \sum_{j=t+1}^{T} \hat{\mathbf{n}}_{t,j} = \mathbf{0} && 0 \leq t \leq T-1, \\
& \mathbf{n}_{t,t} \geq \mathbf{0} && t \geq 0 \\
& \hat{\mathbf{n}}_{j,t} \geq \mathbf{0} && t \geq 1,\ j < t \\
& b_t \geq 0 && t \geq 0.
\end{aligned}$$

Each such dual instance has $(T+1)(KT+2K+2)/2$ variables and $TK+T+1$ constraints, again ignoring non-negativity constraints.

We mentioned in Section 2.2 that the optimal value function, V_0^N, of the no-tax problem may *not* be an upper bound for the optimal value function V_0^{FUL} for the FUL problem. This possibly counter-intuitive result is perhaps best understood via a simple example. Consider a multi-period economy with no cash account and where the returns on all securities are symmetrically distributed about zero. As usual, we assume no short-sales. Because there is no cash account the risk-averse agent is forced to invest all her wealth in the risky securities. In the FUL setting, taxes therefore provide a mechanism for smoothing returns in that they offset losses and reduce gains symmetrically. Because the expected security returns are zero we expect $V_0^{FUL} > V_0^N$ in this case. It should also be clear, however, that the set $\mathcal{X}^F$ of feasible adapted policies for the exact tax-basis FUL problem is not a subset of the set, $\mathcal{X}^N$, of feasible adapted policies for the no-tax problem. As a result, the penalty in (2.4.5) may no longer be dual feasible for the FUL problem since the first-order-conditions (2.4.6) may no longer hold for all $\mathbf{x} \in \mathcal{X}^F$.

We can, however, construct a dual penalty using an approximation $\tilde{V}_t$ to the optimal value function V_t^{FUL} of the exact tax-basis FUL problem. In particular, we can set

$$\pi(\mathbf{x}) := \sum_{t=0}^{T-1} \left(\tilde{V}_{t+1}(\mathbf{x}_{0:t}) - \mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{x}_{0:t})] \right) \tag{A.2-9}$$

where $\mathbf{x}_{0:t}$ denotes the sub-vector of $\mathbf{x}$ corresponding to all the decision variables up to and including time t. BSS [16] showed that the penalty defined in (A.2-9) is dual feasible, i.e.

$\mathbb{E}_0[\pi] = 0$ for any $\mathcal{F}_t$-adapted feasible policy $\mathbf{x}$. Moreover, if we choose $\tilde{V}_t = V_t^{FUL}$, the dual bound computed using the corresponding penalty π is equal to V_0^{FUL} almost surely, i.e. strong duality holds.

It is clear that $\pi(\mathbf{x})$ only depends on the *shape* of $\tilde{V}_t$ because any constant terms will cancel out in (A.2-9). In order to preserve concavity of the objective function, we take $\tilde{V}_{t+1}$ to be a first order Taylor's series expansion of V_{t+1}^N about some $\mathcal{F}_t$-adapted policy $\mathbf{x}_t^*$ which may itself be the result of some sub-optimal policy, i.e. we take

$$\tilde{V}_{t+1}(\mathbf{x}_t) = V_{t+1}^N(w_{t+1}(\mathbf{x}_t^*)) + \nabla_{\mathbf{x}_t} V_{t+1}^N(w_{t+1}(\mathbf{x}_t^*))(\mathbf{x}_t - \mathbf{x}_t^*). \tag{A.2-10}$$

where we define the net portfolio wealth, $w_{t+1}(\mathbf{x}_t)$, of an $\mathcal{F}_t$-adapted policy $\mathbf{x}_t$ to be $w_{t+1}(\mathbf{x}_t) = b_t r_0 + \sum_{j=0}^t \mathbf{p}_{t+1}{}'\mathbf{n}_{j,t} - \tau \sum_{j=0}^t (\mathbf{p}_{t+1} - \mathbf{p}_j)'\mathbf{n}_{j,t}$. The quality of the dual bound will clearly depend on the policy $\mathbf{x}_t^*$. In our numerical results, we found that the policy $\mathbf{x}_t^*$ given by the h-step look-ahead policy led to a superior bound.

In order to compute the dual penalty (A.2-9), we need to calculate $\mathbb{E}_t[V_{t+1}^N(w_{t+1}(\mathbf{x}_t^*))]$ and

$\mathbb{E}_t[\nabla_{\mathbf{x}_t} V_{t+1}^N(w_{t+1}(\mathbf{x}_t^*))]$ where the conditional expectations are taken with respect to the uncertainty that is revealed at time $t+1$. We are not able to compute these expectations explicitly and so we use nested[1] Monte-Carlo simulations to estimate them.

[1] BSS [16] showed that the resulting bounds are still valid dual bounds. More recently, Haugh and Wang [43] show that the nested Monte-Carlo can be done very accurately and efficiently using suitably randomized low discrepancy sequences.

A.3 The Average Tax-Basis LUL Problem

Under the average tax basis rule the tax basis for any security in the portfolio is defined as the average price at which the shares in the portfolio were purchased. The advantage of this is that at any time t there is only one tax basis per security. In contrast, under the exact tax basis approach it is possible to have t separate tax bases for each security at each time t. Consequently, the average tax basis formulation results in significantly fewer state variables and is therefore relatively easier to solve. We consider the primal and dual version of the average tax-basis problem in this Appendix.

A.3.1 Problem Formulation

Let $s^+_{t,k}$ (resp. $s^-_{t,k}$) denote the number of shares of security k purchased (resp. sold) at time t. Let $s_{t,k}$ denote the number of shares in security k held after trading at time t, and let $\tilde{p}_{t,k}$ denote the associated average tax basis. Let $\mathbf{s}^+_t := [s^+_{t,1} ... s^+_{t,K}]'$, $\mathbf{s}^-_t := [s^-_{t,1} ... s^-_{t,K}]'$, $\mathbf{s}_t := [s_{t,1} ... s_{t,K}]'$ and $\tilde{\mathbf{p}}_t := [\tilde{p}_{t,1} ... \tilde{p}_{t,K}]'$. Using these variables, we can formulate the average tax-basis LUL problem as follows:

$$\max_{(b_t, \mathbf{s}_t^+, \mathbf{s}_t^-, \mathbf{s}_t, g_t, l_t, \tilde{p}_{t,k}) \in \mathcal{F}_t,\ t=0,\dots,T} \quad \mathbb{E}_0\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \tag{A.3-1}$$

$$\text{s.t} \quad \mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{s}_t^+ - \mathbf{s}_t^- \text{ with } \mathbf{s}_0 = \mathbf{s}_0^+ \qquad t \geq 1 \tag{A.3-2}$$

$$b_0 + \mathbf{p}_0' \mathbf{s}_0 = w_0 \tag{A.3-3}$$

$$b_t + \mathbf{p}_t' \mathbf{s}_t + \tau g_t = b_{t-1} r_0 + \mathbf{p}_t' \mathbf{s}_{t-1} \qquad t \geq 1 \tag{A.3-4}$$

$$\mathbf{s}_t^+ \geq \mathbf{0} \qquad t \geq 0 \tag{A.3-5}$$

$$\mathbf{s}_{t-1} \geq \mathbf{s}_t^- \geq \mathbf{0}, \qquad t \geq 1 \tag{A.3-6}$$

$$b_t \geq 0 \qquad t \geq 0 \tag{A.3-7}$$

$$g_t + l_t = (\mathbf{p}_t - \tilde{\mathbf{p}}_{t-1})' \mathbf{s}_t^- + l_{t-1} \qquad t \geq 1 \tag{A.3-8}$$

$$g_t \geq 0 \qquad t \geq 1 \tag{A.3-9}$$

$$l_t \leq 0,\ l_0 = 0 \qquad t \geq 1 \tag{A.3-10}$$

$$\tilde{\mathbf{p}}_0 = \mathbf{p}_0 \tag{A.3-11}$$

$$\tilde{p}_{t,k} s_{t,k} = \tilde{p}_{t-1,k}(s_{t-1,k} - s_{t,k}^-) + p_{t,k} s_{t,k}^+ \qquad \forall k, t \geq 1 \tag{A.3-12}$$

and security price and market state dynamics

where (A.3-2) updates the security positions after trading at time t, (A.3-3) (resp. (A.3-4)) is the budget constraint for $t = 0$ (resp. $t > 0$), (A.3-5) and (A.3-6) prohibit short-sales, (A.3-7) prohibits borrowing from the cash account, and (A.3-8)–(A.3-10) formulate the LUL tax rule.

The average tax basis for each security is initially set to the time $t = 0$ price in (A.3-11). Suppose the agent buys $s^+_{t,k} > 0$ shares and sells $s^-_{t,k} > 0$ shares of security k at time t. Then the net position is $s_{t,k} = s_{t-1,k} + s^+_{t,k} - s^-_{t,k}$, and the new tax basis is $\tilde{p}_{t,k} = (\tilde{p}_{t-1,k}(s_{t-1,k} - s^-_{t,k}) + p_{t,k}s^+_{t,k})/s_{t,k}$. This rule is implemented in (A.3-12).

As before, we let $\mathbf{x}_t := [b_t \; \mathbf{s}^{+'}_t \; \mathbf{s}^{-'}_t \; \mathbf{s}'_t \; g_t \; l_t \; \tilde{\mathbf{p}}'_t]$ denote the time-t decision variables. Similarly, we let

$$\mathbb{X}_0 := \left\{ \mathbf{x}_0 \mid \mathbf{x}_0 \in \mathcal{F}_0 \text{ satisfies (A.3-3), (A.3-5), (A.3-7) and (A.3-11) at } t = 0 \right\}$$

$$\mathbb{X}_t(\mathbf{x}_{t-1}) := \left\{ \mathbf{x}_t \mid \mathbf{x}_t \in \mathcal{F}_t \text{ satisfies (A.3-2), (A.3-4)–(A.3-10) and (A.3-12) at time t} \right\}, \; t \geq 1,$$

denote the sets of feasible trades at times $t = 0$ and $t \geq 1$, respectively. Note that the tax-basis is not an argument of $\mathbb{X}_t$ as it was in the exact tax basis case; this is because the tax basis is now a decision variable and is included in $\mathbf{x}_t$.

A.3.2 Suboptimal Policies

Although we now have product terms (A.3-8) and (A.3-12), these terms do not cause any difficulty in adapting our sub-optimal policies from Section 2.3 to the average tax-basis problem here. For example, the average-cost RBH problem at time t takes the form:

$$\max_{\mathbf{x}_t, b_T^{(m)}, g_T^{(m)}} \quad \frac{1}{M}\sum_{m=1}^{M}\frac{(b_T^{(m)})^{1-\gamma}}{1-\gamma} \tag{A.3-13}$$

$$\begin{aligned}
\text{s.t} \quad & \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1}), \\
& b_T^{(m)} = b_t r_0^{T-t} + \mathbf{p}_T^{(m)\prime}\mathbf{s}_t - \tau g_T^{(m)} \qquad 1 \le m \le M \\
& g_T^{(m)} \ge \mathbf{p}_T^{(m)\prime}\mathbf{s}_t - \sum_{k=1}^{K}\tilde{p}_{t,k}s_{t,k} + l_t \qquad 1 \le m \le M \qquad \text{(A.3-14)} \\
& g_T^{(m)} \ge 0
\end{aligned}$$

where M denotes the number of time T scenarios used to approximate the time T uncertainty. The average tax basis $\tilde{\mathbf{p}}_{t-1}$ is known before trading at time t, and is therefore not a decision variable here; hence, (A.3-8) is now a linear constraint. The quadratic term $\tilde{p}_{t,k}s_{t,k}$ in (A.3-14) can be linearized by replacing it by the right hand side of (A.3-12). Consequently, all the constraints in (A.3-13) are linear, and (A.3-13) is a convex optimization problem.

After substituting for $\mathbf{s}_t$ using (A.3-2) and then applying the same smoothing technique of Section 2.3.3, we obtain the following optimization problem

$$\max_{\mathbf{x}_t} \quad \frac{1}{M(1-\gamma)}\sum_{m=1}^{M}\left(b_t r_0^{T-t} + \mathbf{p}_T^{(m)\prime}(\mathbf{s}_t^+ - \mathbf{s}_t^-) - \tau\tilde{g}_T^{(m)}(\mathbf{x}_t) + \mathbf{p}_T^{(m)\prime}\mathbf{s}_{t-1}\right)^{1-\gamma} \tag{A.3-15}$$

$$\text{s.t.} \quad \mathbf{x}_t \in \mathbb{X}_t(\mathbf{x}_{t-1})$$

$$\begin{aligned}
\text{where} \quad \tilde{g}_T^{(m)}(\mathbf{x}_t) = \frac{1}{\theta}\ln\Bigg(\exp\Bigg(\theta\Big((\mathbf{p}_T^{(m)} - \mathbf{p}_t)'\mathbf{s}_t^+ - (\mathbf{p}_T^{(m)} - \tilde{\mathbf{p}}_{t-1})'\mathbf{s}_t^- + l_t \\
+ (\mathbf{p}_T^{(m)} - \tilde{\mathbf{p}}_{t-1})'\mathbf{s}_{t-1}\Big)\Bigg) + 1\Bigg).
\end{aligned} \tag{A.3-16}$$

This problem has a total of $2K+3$ variables: b_t, $\mathbf{s}_t^+$, $\mathbf{s}_t^-$, g_t and l_t. Note that the number of decision variables in (A.3-15) does not increase with t and is considerably fewer than that in the exact tax-basis problem (2.3.9). We can use (A.3-16) to substitute for $\tilde{g}_T^{(m)}(\mathbf{x}_t)$ in the objective function and then, in addition to the non-negativity constraints, there will only be two remaining constraints, namely (A.3-4) and (A.3-8). When we solve this problem with the SQP algorithm of Section 2.3.3, it takes approximately one second to solve all 20 problems on a single sample path. In contrast, the corresponding exact tax-basis problem takes approximately three seconds to solve all 20 problems.

We also note that it is similarly straightforward to compute the optimal trades for the h-step lookahead policy.

A.3.3 Dual Bound

It is easy to see that Lemma 2.2.1 holds regardless of how the tax-basis is computed and that the no-tax value function is a valid upper bound for the average tax-basis LUL problem (A.3-1). However, this bound is likely to be too conservative as we saw in the exact tax basis case of Section 2.5. We therefore want to use the information relaxations approach to compute good dual bounds. In order to do this, we need to solve problems of the form

$$\begin{aligned} \max_{\mathbf{x}} \quad & \frac{b_T^{1-\gamma}}{1-\gamma} - \pi(\mathbf{x}) \\ \text{s.t} \quad & \text{Constraints (A.3-2)–(A.3-12)} \end{aligned} \tag{A.3-17}$$

where the the sample path of prices $\{\mathbf{p}_t : t \geq 0\}$ is known, and the average tax basis $\tilde{p}_{t,k}$, trades $s^-_{t,k}$ and $s_{t,k}$, are all decision variables. Thus, (A.3-8) and (A.3-12) are now non-convex quadratic constraints, and the inner optimization problem that we need to solve in order to compute the upper bound is non-convex. Recall that we need to compute the *global* optimum of the inner optimization problem in order to obtain a valid dual bound. Since it is hard in general to solve a non-convex optimization problem, tackling (A.3-17) directly does not lead to an efficient strategy for computing provably valid dual bounds.

We note here, however, that we can still produce a valid dual bound for the average tax-basis LUL problem if we can either solve (A.3-17) exactly or obtain an *upper bound* for (A.3-17) along each dual sample path. So suppose now that we simulate I dual sample paths and let $V_{up}^{(i)}$ be the optimal solution of the dual problem on the i-th path. Let $I^o \subseteq I$ be those dual paths on which we can compute $V_{up}^{(i)}$, and let $I^u := I - I^o$ be the paths on which we can only obtain an upper bound, $\bar{V}_{up}^{(i)}$, for $V_{up}^{(i)}$. We then clearly have

$$\frac{\sum_{i=1}^{I} V_{up}^{(i)}}{I} \leq \frac{\sum_{i \in I^o} V_{up}^{(i)} + \sum_{i \in I^u} \bar{V}_{up}^{(i)}}{I} \tag{A.3-18}$$

so that the right-hand side of (A.3-18) is a still a valid dual bound. Our approach, which we outline below, produces such a bound. Moreover, we shall see that the quality of this bound remains very good because, at least in our numerical experiments, $\bar{V}_{up}^{(i)}$ is very close to $V_{up}^{(i)}$ for those dual paths in I^u.

Solving with `BARON`

Tawarmalani and Sahinidis [68] proposed a polyhedral branch-and-cut approach to perform global optimization for non-convex optimization problems. Their algorithm generated polyhedral cutting planes and relaxations for multivariate non-convex problems, and their algorithm was implemented in the `BARON` solver [61].

Explicitly imposing tight lower and upper bounds for decision variables can significantly improve the efficiency of the `BARON` solver. We can take advantage of this as follows. First recall that $\mathbf{r}_t$ denotes the time t return vector and that $\mathbf{r}_t$, for $t = 1, ..., T$, is known on each dual sample path. If the investor starts with an initial wealth of w_0 then it is clear that

$$\bar{w}_t = w_0 \prod_{j=1}^{t} \max(r_0, \mathbf{r}_j) \tag{A.3-19}$$

provides an upper bound on the time t wealth of the investor. (The max operator in (A.3-19) returns the maximum of r_0 and all elements in the return vector $\mathbf{r}_j$.) We also note that the average tax-basis for each security at time t must lie between the lowest and highest prices of that security along the sample path up until time t. Given these two observations, we can impose the following bounds for the decision variables on each dual sample path:

$$0 \leq \mathbf{s}_t^+ \leq \bar{w}_t/\mathbf{p}_t \tag{A.3-20}$$

$$0 \leq \mathbf{s}_t^- \leq \bar{w}_t/\mathbf{p}_t \tag{A.3-21}$$

$$0 \leq \mathbf{s}_t \leq \bar{w}_t/\mathbf{p}_t \tag{A.3-22}$$

$$0 \leq g_t \leq \bar{w}_t \tag{A.3-23}$$

$$-\bar{w}_{t-1} \leq l_t \leq 0 \tag{A.3-24}$$

$$\min(\mathbf{p}_0, ..., \mathbf{p}_t) \leq \tilde{\mathbf{p}}_t \leq \max(\mathbf{p}_0, ..., \mathbf{p}_t). \tag{A.3-25}$$

In our numerical experiments with `BARON`, we attempted to solve problem (A.3-17) with the additional constraints (A.3-20) to (A.3-25). We set the solving time limit for each dual path to be one minute and we used 5,000 paths in our numerical experiments. The global optimum on approximately 40% of these paths was found with a relative error of 10^{-6} within one minute. For the remaining 60% of the paths we used the best upper bound that was returned by `BARON` when it stopped after reaching the one-minute cutoff.

A.3.4 Numerical Experiments

In our numerical experiments we used the same problem setting as described in Section 2.5. Because of the time required by `BARON` to solve or upper bound the dual problems, we only considered the case where $\tau = 40\%$, $\gamma = 3$ and $\rho = .4$. The results are shown in Table A.1 and they can be compared to the exact tax basis results in the corresponding row of Table

Table A.1: Results for average tax-basis LUL problem with tax rate $\tau = 40\%$

Parameters			Sub-Optimal Policies					Upper Bounds	
γ	ρ	(%)	V^{tb}	V^{mtb}	V^{bh}	V^{rbh}	V^{hl}	V_d^g	V_0^N
3	0.4	CE return	1.80	1.89	1.84	1.96	1.99	2.19	2.61
		95% C.I	(1.80, 1.81)	(1.89, 1.89)	(1.81, 1.86)	(1.95, 1.98)	(1.98, 2.00)	(2.19, 2.20)	

3. We see that the tax-blind, modified tax-blind and the buy-and-hold policies all perform identically to the exact tax basis case. (This must be the case for the buy-and-hold policy since this policy only trades at time $t = 0$). The RBH and h-step look-ahead policies have a CE return that is only 1 basis point lower than the corresponding exact tax basis policies. This is consistent with the observation of DeMiguel and Uppal [28] who noted (in the one- or two stock case) that the CE loss in wealth is small when using the average tax basis instead of the exact tax basis.

The dual bound, V_d^g, was constructed using the same gradient penalty that we used for the the exact tax basis problem and by then using the `BARON` solver as described earlier. We note that V_d^g is only one basis point higher than the corresponding exact tax-basis case and this suggests that the Tawarmalani and Sahinidis [68] approach is a viable approach for obtaining good dual bounds for the average-cost tax basis formulation.

A.4 The Average Tax-Basis FUL Formulation

As discussed in Appendix A.2.1, under the FUL tax rule the taxable capital gains g_t at time t is given by

$$g_t = (\mathbf{p}_t - \tilde{\mathbf{p}}_{t-1})'\mathbf{s}_t^-. \tag{A.4-1}$$

Substituting (A.4-1) into (A.3-4) and removing the LUL tax rule constraints (A.3-8)–(A.3-10), the average tax-basis FUL problem can be formulated as:

$$\max_{(b_t,\mathbf{s}_t^+,\mathbf{s}_t^-,\mathbf{s}_t,\tilde{p}_{t,k})\in\mathcal{F}_t,\ t=0,\dots,T} \quad \mathbb{E}_0\left[\frac{b_T^{1-\gamma}}{1-\gamma}\right] \tag{A.4-2}$$

$$\begin{aligned}
\text{s.t}\quad & \mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{s}_t^+ - \mathbf{s}_t^- \text{ with } \mathbf{s}_0 = \mathbf{s}_0^+ && t \geq 1 \\
& b_0 + \mathbf{p}_0'\mathbf{s}_0 = w_0 && \\
& b_t + \mathbf{p}_t'\mathbf{s}_t + \tau(\mathbf{p}_t - \tilde{\mathbf{p}}_{t-1})'\mathbf{s}_t^- = b_{t-1}r_0 + \mathbf{p}_t'\mathbf{s}_{t-1} && t \geq 1 \\
& \mathbf{s}_t^+ \geq \mathbf{0} && t \geq 0 \\
& \mathbf{s}_{t-1} \geq \mathbf{s}_t^- \geq \mathbf{0} && t \geq 1 \\
& b_t \geq 0 && t \geq 0 \\
& \tilde{\mathbf{p}}_0 = \mathbf{p}_0 && \\
& \tilde{p}_{t,k}s_{t,k} = \tilde{p}_{t-1,k}(s_{t-1,k} - s_{t,k}^-) + p_{t,k}s_{t,k}^+ && \forall k,\ t \geq 1 \\
& \text{and security price and market state dynamics.} &&
\end{aligned}$$

Comparing (A.4-2) with the average tax-basis LUL problem (A.3-1), we note that (A.4-2) has the same structure but fewer decision variables and constraints. The analysis of Appendix A.3 can also be applied here although we do note the no-tax model no longer provides an upper bound because of the FUL tax rule.

Appendix B

Supplemental Content for Chapter 3

B.1 Solving for the Sub-Optimal Policies

Here we consider the approaches we follow for obtaining the various policies outlined in Section 3.2.2. In each of Appendices B.1.1, B.1.2 and B.1.4 we assume that no-short-sales constraints are *not* imposed. When we discuss convexity and existence of solutions we restrict ourselves to positive definite matrices rather than positive semi-definite matrices for ease of exposition.

B.1.1 The Risk-Neutral Policy

Let $V_t(\cdot)$ denote the time t value function for the problem faced by a risk-neutral agent. At time T the agent needs to buy all remaining shares so that $\mathbf{s}_T = \mathbf{w}_T$. We also have

$$
\begin{aligned}
V_T(\tilde{\mathbf{p}}_T, \mathbf{w}_T, \mathbf{A}_T, \mathbf{B}_T) = \mathbf{p}_T'\mathbf{s}_T &= (\tilde{\mathbf{p}}_T + \mathbf{A}_T\mathbf{w}_T + \mathbf{B}_T\mathbf{w}_T)'\mathbf{w}_T \\
&= \frac{1}{2}\mathbf{w}_T'(\mathbf{A}_T + \mathbf{A}_T' + \mathbf{B}_T + \mathbf{B}_T')\mathbf{w}_T + \tilde{\mathbf{p}}_T'\mathbf{w}_T \\
&= \frac{1}{2}\mathbf{w}_T'\mathbf{G}_T\mathbf{w}_T + \tilde{\mathbf{p}}_T'\mathbf{w}_T \qquad \text{(B.1-1)}
\end{aligned}
$$

where $\mathbf{G}_T := \mathbf{A}_T + \mathbf{A}_T' + \mathbf{B}_T + \mathbf{B}_T'$. At time $T-1$ we need to solve

$$
\begin{aligned}
V_{T-1}(\tilde{\mathbf{p}}_{T-1}, \mathbf{w}_{T-1}, \mathbf{A}_{T-1}, \mathbf{B}_{T-1}) &= \min_{\mathbf{s}_{T-1}} \mathbb{E}_{T-1}\left[\mathbf{p}_{T-1}'\mathbf{s}_{T-1} + V_T(\tilde{\mathbf{p}}_T, \mathbf{w}_T, \mathbf{A}_T, \mathbf{B}_T)\right] \qquad \text{(B.1-2)} \\
&= \min_{\mathbf{s}_{T-1}} \frac{1}{2}\mathbf{s}_{T-1}'\mathbf{N}_{ss,T-1}\mathbf{s}_{T-1} - (\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1})'\mathbf{s}_{T-1} \\
&\quad + \frac{1}{2}\mathbf{w}_{T-1}'\mathbb{E}_{T-1}[\mathbf{G}_T]\mathbf{w}_{T-1} + \tilde{\mathbf{p}}_{T-1}'\mathbf{w}_{T-1} \qquad \text{(B.1-3)}
\end{aligned}
$$

where we have used (B.1-1) to substitute for V_T in (B.1-2) and then used the price dynamics of Section 3.2.1 and $\mathbf{w}_T = \mathbf{w}_{T-1} - \mathbf{s}_{T-1}$ in obtaining (B.1-3) and where we have defined

$$
\begin{aligned}
\mathbf{N}_{ss,T-1} &:= \mathbb{E}_{T-1}[\mathbf{G}_T] + \mathbf{B}_{T-1} + \mathbf{B}_{T-1}' \qquad \text{(B.1-4)} \\
\mathbf{N}_{ws,T-1} &:= \mathbb{E}_{T-1}[\mathbf{G}_T] - \mathbf{A}_{T-1}'. \qquad \text{(B.1-5)}
\end{aligned}
$$

Assuming $\mathbf{N}_{ss,T-1}$ is positive definite, then the objective function in (B.1-3) is convex and the optimal solution to this problem is given by

$$\mathbf{s}^*_{T-1} = \mathbf{N}^{-1}_{ss,T-1}\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1} \tag{B.1-6}$$

with the optimal value function satisfying

$$V_{T-1}(\tilde{\mathbf{p}}_{T-1}, \mathbf{w}_{T-1}, \mathbf{A}_{T-1}, \mathbf{B}_{T-1}) = \frac{1}{2}\mathbf{w}'_{T-1}\mathbf{G}_{T-1}\mathbf{w}_{T-1} + \tilde{\mathbf{p}}'_{T-1}\mathbf{w}_{T-1}$$

where

$$\mathbf{G}_{T-1} \;:=\; \mathbb{E}_{T-1}[\mathbf{G}_T] - \mathbf{N}'_{ws,T-1}\mathbf{N}^{-1}_{ss,T-1}\mathbf{N}_{ws,T-1}. \tag{B.1-7}$$

Continuing in this manner we obtain, for $t = T-2, \ldots, 0$,

$$\begin{aligned} V_t(\tilde{\mathbf{p}}_t, \mathbf{w}_t, \mathbf{A}_t, \mathbf{B}_t) &= \min_{\mathbf{s}_t} \mathbb{E}_t\left[\mathbf{p}'_t\mathbf{s}_t + V_{t+1}(\tilde{\mathbf{p}}_{t+1}, \mathbf{w}_{t+1}, \mathbf{A}_{t+1}, \mathbf{B}_{t+1})\right] \\ &= \min_{\mathbf{s}_t} \frac{1}{2}\mathbf{s}'_t\mathbf{N}_{ss,t}\mathbf{s}_t - (\mathbf{N}_{ws,t}\mathbf{w}_t)'\mathbf{s}_t + \frac{1}{2}\mathbf{w}'_t\mathbb{E}_t[\mathbf{G}_{t+1}]\mathbf{w}_t + \tilde{\mathbf{p}}'_t\mathbf{w}_t \end{aligned}$$

where

$$\mathbf{N}_{ss,t} \;:=\; \mathbb{E}_t[\mathbf{G}_{t+1}] + \mathbf{B}_t + \mathbf{B}'_t \tag{B.1-8}$$

$$\mathbf{N}_{ws,t} \;:=\; \mathbb{E}_t[\mathbf{G}_{t+1}] - \mathbf{A}'_t. \tag{B.1-9}$$

Again assuming $\mathbf{N}_{ss,t}$ is positive definite, the optimal solution satisfies

$$\mathbf{s}_t^* = \mathbf{N}_{ss,t}^{-1}\mathbf{N}_{ws,t}\mathbf{w}_t \tag{B.1-10}$$

and the optimal value function satisfies

$$V_t(\tilde{\mathbf{p}}_t, \mathbf{w}_t, \mathbf{A}_t, \mathbf{B}_t) = \frac{1}{2}\mathbf{w}_t'\mathbf{G}_t\mathbf{w}_t + \tilde{\mathbf{p}}_t'\mathbf{w}_t \tag{B.1-11}$$

where

$$\mathbf{G}_t \;:=\; \mathbb{E}_t[\mathbf{G}_{t+1}] - \mathbf{N}_{ws,t}'\mathbf{N}_{ss,t}^{-1}\mathbf{N}_{ws,t}. \tag{B.1-12}$$

We note that the optimal trading quantities, $\mathbf{s}_t^*$, depend on $\mathbf{w}_t$ only, and not on the price vector $\tilde{\mathbf{p}}_t$ or the stochastic variance-covariance process. It is not path-independent, however, due to its dependence on $\mathbf{A}_t$ and $\mathbf{B}_t$ which are in general stochastic. Note that if any $\mathbf{N}_{ss,t}$ fails to be positive definite then the agent's time t objective function will be unbounded from below (in the absence of constraints) and economic considerations alone would imply that this possibility should be ruled out. One way to do this would be to first impose sufficient structure on the dynamics of $\mathbf{A}_t$ and $\mathbf{B}_t$ so that $\mathbb{E}_t[\mathbf{G}_{t+1}]$ can be computed analytically. We could then look to impose additional conditions that guarantee the positive-definiteness of the $\mathbf{N}_{ss,t}$'s.

Regardless, we could only implement the policy given by (B.1-10) if we can solve for the $\mathbf{G}_t$'s analytically. One situation where it is straightforward to determine the optimality of (B.1-10) and actually implement the policy is when the the $\mathbf{A}_t$'s and $\mathbf{B}_t$'s are deterministic.

Indeed if we assume the $\mathbf{A}_t$'s and $\mathbf{B}_t$'s are constant across time then it is possible to determine explicit expressions for $\mathbf{s}_t^*$. In Appendix B.1.2 we also identify a particular case where the risk-neutral and simple policies coincide.

In the numerical results of Section 3.4 we assumed that the $\mathbf{B}_t$'s were stochastic and so we were not able to implement the risk-neutral policy. Instead we implemented the risk-neutral OLFC policy as discussed in Section 3.2.2 and at the end of Appendix B.1.3.

B.1.2 When the Simple and Risk-Neutral Policies Coincide

Suppose the following two conditions both hold:

(i) $\mathbf{A}_t$ and $\mathbf{B}_t$ are martingales: $\mathbb{E}_t[\mathbf{A}_{t+1}] = \mathbf{A}_t$ and $\mathbb{E}_t[\mathbf{B}_{t+1}] = \mathbf{B}_t$.

(ii) $\mathbf{A}_t$ and $\mathbf{B}_t$ are symmetric: $\mathbf{A}_t = \mathbf{A}_t'$ and $\mathbf{B}_t = \mathbf{B}_t'$.

Then the simple and risk-neutral policies coincide and we can show this by induction as follows. At time T, we have $\mathbf{G}_T = 2\mathbf{A}_T + 2\mathbf{B}_T$ and $\mathbf{s}_T^* = \mathbf{w}_T$. Under conditions (i) and (ii), equations (B.1-4) to (B.1-7) reduce to

$$\begin{aligned}
\mathbf{N}_{ss,T-1} &= \mathbb{E}_{T-1}[\mathbf{G}_T] + \mathbf{B}_{T-1} + \mathbf{B}_{T-1}' = 2\mathbf{A}_{T-1} + 4\mathbf{B}_{T-1} \\
\mathbf{N}_{ws,T-1} &= \mathbb{E}_{T-1}[\mathbf{G}_T] - \mathbf{A}_{T-1}' = \mathbf{A}_{T-1} + 2\mathbf{B}_{T-1} \\
\mathbf{s}_{T-1}^* &= \mathbf{N}_{ss,T-1}^{-1}\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1} = \frac{1}{2}\mathbf{w}_{T-1} \\
\mathbf{G}_{T-1} &= \mathbb{E}_{T-1}[\mathbf{G}_T] - \mathbf{N}_{ws,T-1}'\mathbf{N}_{ss,T-1}^{-1}\mathbf{N}_{ws,T-1} = \left(1 + \frac{1}{2}\right)\mathbf{A}_{T-1} + \frac{1}{2}2\mathbf{B}_{T-1}
\end{aligned}$$

and in this case the simple and risk-neutral policies do indeed coincide. Suppose now that at time $t+1$ we have

$$\begin{aligned}\mathbf{G}_{t+1} &= \left(1+\frac{1}{T-(t+1)+1}\right)\mathbf{A}_{t+1}+\frac{1}{T-(t+1)+1}2\mathbf{B}_{t+1}\\ \text{and }\ \mathbf{s}^*_{t+1} &= \frac{1}{T-(t+1)+1}\mathbf{w}_{t+1}.\end{aligned}$$

Then (B.1-8), (B.1-9), (B.1-12) and (B.1-10) yield

$$\begin{aligned}\mathbf{N}_{ss,t} &= \mathbb{E}_t[\mathbf{G}_{t+1}]+\mathbf{B}_t+\mathbf{B}'_t = \left(1+\frac{1}{T-t}\right)\mathbf{A}_t+\left(1+\frac{1}{T-t}\right)2\mathbf{B}_t\\ \mathbf{N}_{ws,t} &= \mathbb{E}_t[\mathbf{G}_{t+1}]-\mathbf{A}'_t = \frac{1}{T-t}\mathbf{A}_t+\frac{1}{T-t}2\mathbf{B}_t\\ \mathbf{G}_t &= \mathbb{E}_t[\mathbf{G}_{t+1}]-\mathbf{N}'_{ws,t}\mathbf{N}^{-1}_{ss,t}\mathbf{N}_{ws,t} = \left(1+\frac{1}{T-t+1}\right)\mathbf{A}_t+\frac{1}{T-t+1}2\mathbf{B}_t\\ \mathbf{s}^*_t &= \mathbf{N}^{-1}_{ss,t}\mathbf{N}_{ws,t}\mathbf{w}_t = \frac{1}{T-t+1}\mathbf{w}_t \qquad \text{(B.1-13)}\end{aligned}$$

and by (B.1-13) we see the inductive step is complete. $\tilde{\mathbf{s}}$, $\tilde{\mathbf{s}}$

B.1.3 Computing the OLFC Policy When $\mathbf{s}_j \geq \mathbf{0}$ Is Imposed

We now consider calculation of the OLFC when non-negativity constraints on the $\mathbf{s}_j$'s are imposed. It should be clear that imposing additional convex constraints on the $\mathbf{s}_j$'s would also be straightforward. We first note that equations (3.2.5) and (3.2.6) imply

$$\mathbf{p}_j = \tilde{\mathbf{p}}_t+\sum_{i=t}^{j-1}\mathbf{A}_i\mathbf{s}_i+\sum_{i=t+1}^{j}\mathbf{r}_i+\mathbf{A}_j\mathbf{s}_j+\mathbf{B}_j\mathbf{s}_j \quad \text{for } j=t,\ldots,T. \qquad \text{(B.1-14)}$$

The OLFC policy assumes the price impact matrices evolve deterministically by taking their time-t conditional expectations. We therefore substitute (B.1-14) into the objective function in (3.2.10) but with $\mathbf{A}_j$ and $\mathbf{B}_j$ replaced by $\mathbb{E}_t[\mathbf{A}_j]$ and $\mathbb{E}_t[\mathbf{B}_j]$, respectively. We then see the OLFC problem can be formulated as

$$\min_{\mathbf{s}_{t:T}\geq 0} \mathbb{E}_t\left[\exp\left(\gamma\tilde{\mathbf{p}}_t'\mathbf{w}_t + \gamma\sum_{j=t}^{T}\left(\sum_{i=t}^{j-1}\mathbb{E}_t[\mathbf{A}_i]\mathbf{s}_i + \mathbb{E}_t[\mathbf{A}_j]\mathbf{s}_j + \mathbb{E}_t[\mathbf{B}_j]\mathbf{s}_j\right)'\mathbf{s}_j + \gamma\sum_{j=t+1}^{T}\mathbf{r}_j'\mathbf{w}_j\right)\right] \quad \text{(B.1-15)}$$

$$\equiv \min_{\mathbf{s}_{t:T}\geq 0} \exp\left(\gamma\sum_{j=t}^{T}\left(\sum_{i=t}^{j-1}\mathbb{E}_t[\mathbf{A}_i]\mathbf{s}_i + \mathbb{E}_t[\mathbf{A}_j]\mathbf{s}_j + \mathbb{E}_t[\mathbf{B}_j]\mathbf{s}_j\right)'\mathbf{s}_j\right)\mathbb{E}_t\left[\exp\left(\gamma\sum_{j=t+1}^{T}\mathbf{r}_j'\mathbf{w}_j\right)\right] \quad \text{(B.1-16)}$$

where we have used the fact that $\mathbf{w}_j = \sum_{i=j}^{T}\mathbf{s}_i$ for $j = t, \ldots, T$ and switched the order of summation to obtain the last term in (B.1-15). The OLFC policy also assumes the $\mathbf{r}_j$'s are IID normal with mean vector $\mathbf{0}$ and conditional covariance matrix, $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_t$, for $j = t+1, \ldots, T$. Under this assumption we have

$$\mathbb{E}_t\left[\exp\left(\gamma\sum_{j=t+1}^{T}\mathbf{r}_j\mathbf{w}_j\right)\right] = \exp\left(\sum_{j=t+1}^{T}\frac{1}{2}\gamma^2\mathbf{w}_j'\boldsymbol{\Sigma}_t\mathbf{w}_j\right)$$

so that (B.1-16) becomes

$$\min_{\mathbf{s}_{t:T}\geq 0} \exp\left(\gamma\sum_{j=t}^{T}\left(\sum_{i=t}^{j-1}\mathbb{E}_t[\mathbf{A}_i]\mathbf{s}_i + \mathbb{E}_t[\mathbf{A}_j]\mathbf{s}_j + \mathbb{E}_t[\mathbf{B}_j]\mathbf{s}_j\right)'\mathbf{s}_j + \frac{1}{2}\gamma^2\sum_{j=t+1}^{T}\mathbf{w}_j'\boldsymbol{\Sigma}_t\mathbf{w}_j\right). \quad \text{(B.1-17)}$$

Because the exponential function is monotonic increasing, we can ignore it so that solving (B.1-17) reduces to the following constrained quadratic programming problem:

$$\begin{aligned} \min_{\mathbf{s}_{t:T}} \quad & \frac{1}{2}[\mathbf{s}_t' \, \dots \, \mathbf{s}_T'](\mathbf{Q}_{OLFC,t} + \mathbf{Q}_{\mathbf{\Sigma},t})[\mathbf{s}_t' \, \dots \, \mathbf{s}_T']' \\ \text{s.t.} \quad & \sum_{j=t}^{T} \mathbf{s}_j = \mathbf{w}_t \\ & \mathbf{s}_j \geq \mathbf{0} \quad \text{for } j = t, \dots, T \end{aligned} \tag{B.1-18}$$

where

$$\mathbf{Q}_{OLFC,t} := \begin{bmatrix} \mathbb{E}_t[\mathbf{A}_t + \mathbf{A}_t' + \mathbf{B}_t + \mathbf{B}_t'] & \mathbb{E}_t[\mathbf{A}_t'] & \dots & \mathbb{E}_t[\mathbf{A}_t'] \\ \mathbb{E}_t[\mathbf{A}_t] & \mathbb{E}_t[\mathbf{A}_{t+1} + \mathbf{A}_{t+1}' + \mathbf{B}_{t+1} + \mathbf{B}_{t+1}'] & \dots & \mathbb{E}_t[\mathbf{A}_{t+1}'] \\ \dots & \dots & \dots & \dots \\ \mathbb{E}_t[\mathbf{A}_t] & \mathbb{E}_t[\mathbf{A}_{t+1}] & \dots & \mathbb{E}_t[\mathbf{A}_T + \mathbf{A}_T' + \mathbf{B}_T + \mathbf{B}_T'] \end{bmatrix}, \tag{B.1-19}$$

and

$$\mathbf{Q}_{\mathbf{\Sigma},t} := \gamma \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma}_t & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \\ \mathbf{O} & \mathbf{\Sigma}_t & 2\mathbf{\Sigma}_t & \dots & 2\mathbf{\Sigma}_t \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{O} & \mathbf{\Sigma}_t & 2\mathbf{\Sigma}_t & \dots & (T-t)\mathbf{\Sigma}_t \end{bmatrix},$$

where $\mathbf{O}$ denotes the $n \times n$ zero matrix. We first note that $\mathbf{Q}_{\mathbf{\Sigma},t}$ may be expressed as a positively weighted sum of positive semi-definite matrices according to

$$\mathbf{Q}_{\mathbf{\Sigma},t} = \gamma \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma}_t & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \\ \mathbf{O} & \mathbf{\Sigma}_t & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{O} & \mathbf{\Sigma}_t & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \end{bmatrix} + \gamma \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{O} & \mathbf{O} & \mathbf{\Sigma}_t & \dots & \mathbf{\Sigma}_t \end{bmatrix} + \cdots + \gamma \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \dots & \mathbf{\Sigma}_t \end{bmatrix}.$$

and so it follows that $\mathbf{Q}_{\mathbf{\Sigma},t}$ is itself positive semi-definite. As stated in Section 3.3.1 it seems appropriate to insist on the positive-definiteness of $\mathbf{Q}_{OLFC,t}$ so that the OLFC decision-maker does not perceive that there are arbitrage opportunities in the market. In Appendix B.2.1 we will provide sufficient conditions to guarantee that $\mathbf{Q}_{OLFC,t}$ is indeed positive definite. If these conditions are satisfied (as they are in the numerical experiments of Section 3.4) then $\mathbf{Q}_{OLFC,t} + \mathbf{Q}_{\mathbf{\Sigma},t}$ will be positive definite so that the OLFC policy can be found by solving a constrained convex quadratic optimization problem.

We also note that the **risk-neutral OLFC** policy can be obtained by taking $\gamma = 0$ and then solving the optimization problem in (B.1-18). This results in an objective function of the form

$$\min_{\mathbf{s}_{t:T}} \quad \frac{1}{2}[\mathbf{s}_t' \dots \mathbf{s}_T']\mathbf{Q}_{OLFC,t}[\mathbf{s}_t' \dots \mathbf{s}_T']'.$$

B.1.4 Computing the OLFC Policy When $\mathbf{s}_j \geq \mathbf{0}$ Is Not Imposed

If we drop the non-negativity constraints on the $\mathbf{s}_j$'s then at each time t the OLFC policy can be computed analytically and very quickly using dynamic programming. Recall the agent assumes that $\mathbf{\Sigma}_j = \mathbf{\Sigma}_t$ for all $j \geq t$ so that he assumes

$$\mathbb{E}_j[\exp(\gamma \mathbf{r}'_{j+1}\mathbf{w})] = \exp\left(\frac{1}{2}\gamma^2 \mathbf{w}'\mathbf{\Sigma}_t \mathbf{w}\right) \tag{B.1-20}$$

for all $j \geq t$ and all constant vectors, $\mathbf{w}$. We use $V_j^{ol}(\cdot)$ for $j \geq t$ to denote the expected utility of trading from time j onwards as perceived by the agent at time t. The dynamics of Section 3.2.1, but now taking $\mathbf{\Sigma}_j = \mathbf{\Sigma}_t$ and replacing the $\mathbf{A}_j$'s and $\mathbf{B}_j$'s by their time-t conditional expectations for all $j \geq t$, imply that

$$\begin{aligned} V_T^{ol}(\tilde{\mathbf{p}}_T, \mathbf{w}_T) &= \exp\left(\gamma(\tilde{\mathbf{p}}_T + \mathbb{E}_t[\mathbf{A}_T]\mathbf{w}_T + \mathbb{E}_t[\mathbf{B}_T]\mathbf{w}_T)'\mathbf{w}_T\right) \\ &= \exp\left(\gamma\left(\frac{1}{2}\mathbf{w}'_T\mathbf{G}_T\mathbf{w}_T + \tilde{\mathbf{p}}'_T\mathbf{w}_T\right)\right) \end{aligned}$$

where $\mathbf{G}_T := \mathbb{E}_t[\mathbf{A}_T + \mathbf{A}'_T + \mathbf{B}_T + \mathbf{B}'_T]$. At time $T-1$ the agent needs to solve

$$V_{T-1}^{ol}(\tilde{\mathbf{p}}_{T-1}, \mathbf{w}_{T-1}) = \min_{\mathbf{s}_{T-1}} \ \mathbb{E}_{T-1}[\exp(\gamma \mathbf{p}'_{T-1}\mathbf{s}_{T-1})V_T^{ol}(\tilde{\mathbf{p}}_T,\ \mathbf{w}_T)]$$

and using (B.1-20) this problem reduces to

$$\begin{aligned} & \min_{\mathbf{s}_{T-1}} \quad \exp\Bigg(\gamma\Bigg(\frac{1}{2}\mathbf{s}'_{T-1}\mathbf{N}_{ss,T-1}\mathbf{s}_{T-1} - (\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1})'\mathbf{s}_{T-1} \\ & \qquad + \frac{1}{2}\mathbf{w}'_{T-1}(\mathbf{G}_T + \gamma\Sigma_t)\mathbf{w}_{T-1} + \tilde{\mathbf{p}}'_{T-1}\mathbf{w}_{T-1}\Bigg)\Bigg) \\ \equiv & \min_{\mathbf{s}_{T-1}} \quad \frac{1}{2}\mathbf{s}'_{T-1}\mathbf{N}_{ss,T-1}\mathbf{s}_{T-1} - (\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1})'\mathbf{s}_{T-1} \\ & \qquad + \frac{1}{2}\mathbf{w}'_{T-1}(\mathbf{G}_T + \gamma\Sigma_t)\mathbf{w}_{T-1} + \tilde{\mathbf{p}}'_{T-1}\mathbf{w}_{T-1} \end{aligned} \tag{B.1-21}$$

where

$$\begin{aligned} \mathbf{N}_{ss,T-1} &:= \mathbf{G}_T + \gamma\mathbf{\Sigma}_t + \mathbb{E}_t[\mathbf{B}_{T-1} + \mathbf{B}'_{T-1}] \\ \mathbf{N}_{ws,T-1} &:= \mathbf{G}_T + \gamma\mathbf{\Sigma}_t - \mathbb{E}_t[\mathbf{A}'_{T-1}]. \end{aligned}$$

Assuming $\mathbf{N}_{ss,T-1}$ is positive definite, then the optimal solution to (B.1-21) is

$$\mathbf{s}^*_{T-1} = \mathbf{N}^{-1}_{ss,T-1}\mathbf{N}_{ws,T-1}\mathbf{w}_{T-1}$$

with corresponding value function

$$V^{ol}_{T-1}(\tilde{\mathbf{p}}_{T-1}, \mathbf{w}_{T-1}) = \exp\left(\gamma\left(\frac{1}{2}\mathbf{w}'_{T-1}\mathbf{G}_{T-1}\mathbf{w}_{T-1} + \tilde{\mathbf{p}}'_{T-1}\mathbf{w}_{T-1}\right)\right)$$

where

$$\mathbf{G}_{T-1} := \mathbf{G}_T + \gamma\mathbf{\Sigma}_t - \mathbf{N}'_{ws,T-1}\mathbf{N}^{-1}_{ss,T-1}\mathbf{N}_{ws,T-1}.$$

Continuing backwards in this manner we find for $j = T-2, ..., t$

$$\begin{aligned}
V_j^{ol}(\tilde{\mathbf{p}}_j, \mathbf{w}_j) &= \min_{\mathbf{s}_j} \ \mathbb{E}_j[\exp(\gamma \mathbf{p}_j' \mathbf{s}_j) V_{j+1}(\tilde{\mathbf{p}}_{j+1}, \mathbf{w}_{j+1})] \\
&= \min_{\mathbf{s}_j} \ \exp\left(\gamma\left(\frac{1}{2}\mathbf{s}_j'\mathbf{N}_{ss,j}\mathbf{s}_j - (\mathbf{N}_{ws,j}\mathbf{w}_j)'\mathbf{s}_j + \frac{1}{2}\mathbf{w}_j'(\mathbf{G}_{j+1} + \gamma\Sigma_t)\mathbf{w}_j + \tilde{\mathbf{p}}_j'\mathbf{w}_j\right)\right) \\
&\equiv \min_{\mathbf{s}_j} \ \frac{1}{2}\mathbf{s}_j'\mathbf{N}_{ss,j}\mathbf{s}_j - (\mathbf{N}_{ws,j}\mathbf{w}_j)'\mathbf{s}_j + \frac{1}{2}\mathbf{w}_j'(\mathbf{G}_{j+1} + \gamma\Sigma_t)\mathbf{w}_j + \tilde{\mathbf{p}}_j'\mathbf{w}_j \qquad \text{(B.1-22)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{N}_{ss,j} &:= \mathbf{G}_{j+1} + \gamma\mathbf{\Sigma}_t + \mathbb{E}_t[\mathbf{B}_j + \mathbf{B}_j'] \\
\mathbf{N}_{ws,j} &:= \mathbf{G}_{j+1} + \gamma\mathbf{\Sigma}_t - \mathbb{E}_t[\mathbf{A}_j']
\end{aligned}$$

Again assuming $\mathbf{N}_{ss,j}$ is positive definite, the optimal value of $\mathbf{s}_j$ in (B.1-22) is given by

$$\mathbf{s}_j^* = \mathbf{N}_{ss,j}^{-1}\mathbf{N}_{ws,j}\mathbf{w}_j$$

with corresponding value function

$$V_j^{ol}(\tilde{\mathbf{p}}_j, \mathbf{w}_j) = \exp\left(\gamma\left(\frac{1}{2}\mathbf{w}_j'\mathbf{G}_j\mathbf{w}_j + \tilde{\mathbf{p}}_j'\mathbf{w}_j\right)\right). \qquad \text{(B.1-23)}$$

where

$$\mathbf{G}_j := \mathbf{G}_{j+1} + \gamma\mathbf{\Sigma}_t - \mathbf{N}_{ws,j}'\mathbf{N}_{ss,j}^{-1}\mathbf{N}_{ws,j}.$$

The OLFC policy implements $\mathbf{s}_t^*$ at time t, and we note that $\mathbf{s}_t^*$ depends on $\mathbf{w}_t$ only, and not on the price, $\tilde{\mathbf{p}}_t$. The OLFC policy is path dependent, however, because of the dependence of $\mathbf{N}_{ss,t}$ and $\mathbf{N}_{ws,t}$ on $\boldsymbol{\Sigma}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$ which in general are stochastic. As mentioned in Section 3.5, it is also possible to model return predictability via linear state variable dynamics and still compute the OLFC value function and policy analytically. This would induce an explicit path dependence of $\mathbf{s}_t^*$ on these state variables.

Note that the DP formulation here is equivalent to the static problem formulation in Appendix B.1.3 if we drop the non-negativity constraints on the $\mathbf{s}_t^*$'s there. It follows then that we can check positive definiteness of the $\mathbf{N}_{ss,j}$'s for $j \geq t$ by confirming that $\mathbf{Q}_{OLFC,t} + \mathbf{Q}_{\boldsymbol{\Sigma},t}$ is positive definite. We can do the latter using, for example, the results of Appendix B.2.1.

B.1.5 Using Control Variates to Estimate the Primal Bounds

In order to reduce the number of Monte-Carlo paths that we used for estimating the primal bound, V_{ub}, we considered two possible control variates. First note that if a trading sequence $\mathbf{s} := [\mathbf{s}_0' \ldots \mathbf{s}_T']'$ is deterministic, then the expected execution cost can be computed analytically under the price impact model of Section 3.2. In particular, it is easy to check that

$$\begin{aligned}
\mathbb{E}_0\left[\sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right] &= \mathbb{E}_0\left[\sum_{t=0}^{T}\left(\tilde{\mathbf{p}}_0 + \sum_{i=0}^{t-1}\mathbf{A}_i\mathbf{s}_i + \sum_{i=1}^{t}\mathbf{r}_i + \mathbf{A}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{s}_t\right)'\mathbf{s}_t\right] \\
&= \frac{1}{2}\mathbf{s}'\mathbb{E}_0[\mathbf{Q}]\mathbf{s} + \tilde{\mathbf{p}}_0'\mathbf{w}_0 \qquad \text{(B.1-24)}
\end{aligned}$$

where $\mathbf{Q}$ is as defined in Appendix B.2.1. If we can evaluate $\mathbb{E}_0[\mathbf{Q}]$ then by (B.1-24) we can use $\sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t$ as a control variate. However as the level of risk aversion (as measured by γ) increases, the variance reduction that it achieves will not be as effective.

The expected utility of a deterministic trading sequence cannot be computed analytically when there are stochastic variance-covariance dynamics and stochastic linear price-impact dynamics. However, if we assume the volatility, $\mathbf{\Sigma}$, is constant, and $\mathbf{A}_t$ and $\mathbf{B}_t$ evolve deterministically according to their time $t = 0$ conditional expectations, $\mathbb{E}_0[\mathbf{A}_t]$ and $\mathbb{E}_0[\mathbf{B}_t]$, then it is easy to check that

$$\begin{aligned}
& \mathbb{E}_0\left[\exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right)\right] \\
= & \ \mathbb{E}_0\left[\exp\left(\gamma \sum_{t=0}^{T} \left(\tilde{\mathbf{p}}_0 + \sum_{i=0}^{t-1} \mathbb{E}_0[\mathbf{A}_i]\mathbf{s}_i + \sum_{i=1}^{t} \mathbf{r}_i + \mathbb{E}_0[\mathbf{A}_t]\mathbf{s}_t + \mathbb{E}_0[\mathbf{B}_t]\mathbf{s}_t\right)' \mathbf{s}_t\right)\right] \\
= & \ \exp(\gamma \tilde{\mathbf{p}}_0' \mathbf{w}_0) \exp\left(\frac{\gamma}{2} \mathbf{s}'(\mathbf{Q}_{OLFC,0} + \mathbf{Q}_{\mathbf{\Sigma},0})\mathbf{s}\right). \qquad \text{(B.1-25)}
\end{aligned}$$

We can therefore use $\exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right)$ and (B.1-25) as a control variate. Note that we can do this even when each of $\mathbf{\Sigma}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$ evolve stochastically in the "true" model by simulating in parallel the model with deterministic price impacts and variance-covariance dynamics that yields (B.1-25). In our numerical experiments we took $\mathbf{\Sigma} = \mathbf{\Sigma}_0$ and took $\mathbf{s}$ to be the optimal (deterministic) policy as calculated in Appendix B.1.4 with $t = 0$. We generally find (B.1-25) to be much more effective than (B.1-24) and it resulted in a variance reduction for the primal bounds on the order of 75% to 95% depending on the value of γ.

It is worth mentioning that these control variates provided little benefit when estimating the dual bounds. Since the calculation of the primal bound was the computational bottleneck, however, there was no need to construct good control variates for the dual problem.

B.2 The Dual Problem for the Portfolio Execution Problem

We define $f(\mathbf{s}) := \exp\left(\gamma \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t\right)$, from Section 3.3.1. We can substitute for $\mathbf{p}_t$ using (B.1-14) to obtain

$$f(\mathbf{s}) = \exp\left(\gamma \sum_{t=0}^{T} \left(\tilde{\mathbf{p}}_0 + \sum_{i=0}^{t-1} \mathbf{A}_i \mathbf{s}_i + \sum_{i=1}^{t} \mathbf{r}_i + \mathbf{A}_t \mathbf{s}_t + \mathbf{B}_t \mathbf{s}_t\right)' \mathbf{s}_t\right) \quad \text{(B.2-1)}$$

$$= \exp(\gamma \tilde{\mathbf{p}}_0' \mathbf{w}_0) \, \exp\left(\gamma \left(\frac{1}{2} \mathbf{s}' \mathbf{Q} \mathbf{s} + \mathbf{c}_{p,0}' \mathbf{s}\right)\right) \quad \text{(B.2-2)}$$

where $\mathbf{Q}$ is an $n(T+1) \times n(T+1)$ symmetric matrix given by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_0 + \mathbf{A}_0' + \mathbf{B}_0 + \mathbf{B}_0' & \mathbf{A}_0' & \dots & \mathbf{A}_0' \\ \mathbf{A}_0 & \mathbf{A}_1 + \mathbf{A}_1' + \mathbf{B}_1 + \mathbf{B}_1' & \dots & \mathbf{A}_1' \\ \dots & \dots & \dots & \dots \\ \mathbf{A}_0 & \mathbf{A}_1 & \dots & \mathbf{A}_T + \mathbf{A}_T' + \mathbf{B}_T + \mathbf{B}_T' \end{bmatrix} \quad \text{(B.2-3)}$$

and $\mathbf{c}_{p,0}$ is a $n(T+1) \times 1$ vector given by

$$\mathbf{c}_{p,0} = \begin{bmatrix} \mathbf{0} \\ \mathbf{r}_1 \\ \mathbf{r}_1 + \mathbf{r}_2 \\ \ldots \\ \mathbf{r}_1 + \mathbf{r}_2 + \ldots + \mathbf{r}_t \\ \ldots \\ \mathbf{r}_1 + \mathbf{r}_2 + \ldots + \mathbf{r}_t + \ldots + \mathbf{r}_T \end{bmatrix}.$$

Using (B.2-2) we see that the gradient vector and Hessian matrix of f satisfy

$$\begin{aligned} \nabla f(\mathbf{s}) &= \gamma f(\mathbf{s})(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0}) \\ \nabla^2 f(\mathbf{s}) &= \gamma f(\mathbf{s})\Big(\mathbf{Q} + \gamma(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0})(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0})'\Big). \end{aligned}$$

If $\mathbf{Q}$ is positive definite, then the Hessian matrix $\nabla^2 f(\mathbf{s})$ is also positive definite and $f(\mathbf{s})$ is therefore convex. This follows because $\gamma > 0$, $f(\mathbf{s}) > 0$ for all $\mathbf{s}$ and because $(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0})(\mathbf{Q}\mathbf{s} + \mathbf{c}_{p,0})'$, as the outer-product of a column vector, is positive semi-definite for all $\mathbf{s}$. It therefore follows that if $\mathbf{Q}$ is positive definite then *all* dual problem instances will be convex. In Appendix B.2.1, provide some sufficient conditions that ensure $\mathbf{Q}$ will be positive definite.

B.2.1 Conditions that Guarantee the Positive Definiteness of Q and $\mathbf{Q}_{OLFC,t}$

The positive definiteness of $\mathbf{Q}$ depends on the dynamics of $\mathbf{A}_t$ and $\mathbf{B}_t$. Here we provide two sufficient conditions that guarantee this.

(i) The permanent price impact is constant over time $\mathbf{A}_t = \mathbf{A}$, and $\mathbf{A} + \mathbf{A}'$ is positive definite.

(ii) $\mathbf{B}_t + \mathbf{B}_t'$ are positive semi-definite.

Assuming that conditions (i) and (ii) hold we can then write (B.2-3) as

$$\mathbf{Q} = \begin{bmatrix} \frac{\mathbf{A}+\mathbf{A}'}{2} & \mathbf{A}' & \dots & \mathbf{A}' \\ \mathbf{A} & \frac{\mathbf{A}+\mathbf{A}'}{2} & \dots & \mathbf{A}' \\ \dots & \dots & \dots & \dots \\ \mathbf{A} & \mathbf{A} & \dots & \frac{\mathbf{A}+\mathbf{A}'}{2} \end{bmatrix} + \begin{bmatrix} \frac{\mathbf{A}+\mathbf{A}'}{2} + \mathbf{B}_0 + \mathbf{B}_0' & & & \\ & \frac{\mathbf{A}+\mathbf{A}'}{2} + \mathbf{B}_1 + \mathbf{B}_1' & & \\ & & \dots & \\ & & & \frac{\mathbf{A}+\mathbf{A}'}{2} + \mathbf{B}_T + \mathbf{B}_T' \end{bmatrix}. \tag{B.2-4}$$

The second matrix on the right-hand-side of (B.2-4) is positive definite since it is block-diagonal and each $\frac{1}{2}(\mathbf{A} + \mathbf{A}') + \mathbf{B}_t + \mathbf{B}_t'$ is positive definite. The first matrix is positive semi-definite since for any given vector $\mathbf{z} = [\mathbf{z}_1 \ \dots \ \mathbf{z}_T]'$ where each $\mathbf{z}_i$ is $n \times 1$ vector, we have

$$[\mathbf{z}_1 \ \dots \ \mathbf{z}_T]' \begin{bmatrix} \frac{\mathbf{A}+\mathbf{A}'}{2} & \mathbf{A}' & \dots & \mathbf{A}' \\ \mathbf{A} & \frac{\mathbf{A}+\mathbf{A}'}{2} & \dots & \mathbf{A}' \\ \dots & \dots & \dots & \dots \\ \mathbf{A} & \mathbf{A} & \dots & \frac{\mathbf{A}+\mathbf{A}'}{2} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_T \end{bmatrix} = (\mathbf{z}_1 + \mathbf{z}_2 + \dots + \mathbf{z}_T)' \frac{\mathbf{A} + \mathbf{A}'}{2} (\mathbf{z}_1 + \mathbf{z}_2 + \dots + \mathbf{z}_T) \tag{B.2-5}$$

which is non-negative since $\frac{\mathbf{A}+\mathbf{A}'}{2}$ is positive definite. This implies $\mathbf{Q}$ is positive definite.

$\mathbf{Q}_{OLFC,t}$ in (B.1-19) has a similar structure to $\mathbf{Q}$. Given the $\mathbf{A}_t$'s are constant we have $\mathbb{E}_t[\mathbf{A}_j] = \mathbf{A}$, and then

$$\mathbf{Q}_{OLFC,t} = \begin{bmatrix} \mathbf{A}+\mathbf{A}'+\mathbb{E}_t[\mathbf{B}_t+\mathbf{B}_t'] & \mathbf{A}' & \dots & \mathbf{A}' \\ \mathbf{A} & \mathbf{A}+\mathbf{A}'+\mathbb{E}_t[\mathbf{B}_{t+1}+\mathbf{B}_{t+1}'] & \dots & \mathbf{A}' \\ \dots & \dots & \dots & \dots \\ \mathbf{A} & \mathbf{A} & \dots & \mathbf{A}+\mathbf{A}'+\mathbb{E}_t[\mathbf{B}_T+\mathbf{B}_T'] \end{bmatrix}$$

If the dynamics of $\mathbf{B}_t$ ensures that $\mathbb{E}_t[\mathbf{B}_j + \mathbf{B}_j']$ are positive semi-definite, $\mathbf{Q}_{OLFC,t}$ is guaranteed to be positive definite by the same argument we used for (B.2-4).

B.2.2 Computing Dual Penalties

We saw in Section 3.3.1 that we would like to take $\tilde{V}_t$ (required for (3.3.2)) to be a linearized version of an approximate value function, $\hat{V}_t$, as given by (3.3.5). However, if we construct $\hat{V}_t$ using the OLFC value function, V_t^{ol}, and take $\hat{V}_{t+1}(\mathbf{s}_{0:t}) = \exp(\gamma \sum_{j=0}^{t} \mathbf{p}_j' \mathbf{s}_j) V_{t+1}^{ol}$, then we cannot compute $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})]$ analytically. This is because V_{t+1}^{ol} is calculated under the assumption that the conditional covariance matrix remains constant at $\mathbf{\Sigma}_{t+1}$ and that the price impact matrices for all $j \geq t+1$ are $\mathbb{E}_{t+1}[\mathbf{A}_j]$'s and $\mathbb{E}_{t+1}[\mathbf{B}_j]$'s. None of these terms are adapted to $\mathcal{F}_t$ and so computing $\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})]$ analytically is not possible in general. We could in theory use the Monte-Carlo approach of Section 3.5 to overcome this problem, but in this particular case it would be prohibitively expensive to do so. This is because we would need to solve an OLFC optimization problem at each simulated point at time $t+1$.

Instead we simply modify the assumptions of the OLFC policy and assume that at time $t+1$ the conditional covariance matrix remains constant at $\mathbf{\Sigma}_t$ (rather than $\mathbf{\Sigma}_{t+1}$) and that the price impact matrices are given by $\mathbb{E}_t[\mathbf{A}_j]$ and $\mathbb{E}_t[\mathbf{B}_j]$ (rather than $\mathbb{E}_{t+1}[\mathbf{A}_j]$ and $\mathbb{E}_{t+1}[\mathbf{B}_j]$) for all $j \geq t+1$. Using precisely the same DP approach of Appendix B.1.4 we obtain the *modified* OLFC value function

$$V_{t+1}^{mol} = \exp\left(\gamma\left(\frac{1}{2}\mathbf{w}_{t+1}'\tilde{\mathbf{G}}_{t+1}\mathbf{w}_{t+1} + \tilde{\mathbf{p}}_{t+1}'\mathbf{w}_{t+1}\right)\right) \tag{B.2-6}$$

where the $\tilde{\mathbf{G}}_t$'s are the analog of the $\mathbf{G}_t$'s in Appendix B.1.4. The important feature of (B.2-6) is that $\tilde{\mathbf{G}}_{t+1} \in \mathcal{F}_t$ so that $\mathbb{E}_t[V_{t+1}^{mol}]$ can be computed in closed form.

B.2.3 Solving Dual Problem Instances

Recall that we define $f(\mathbf{s}) := \exp(\gamma \sum_{t=0}^{T} \mathbf{p}_t'\mathbf{s}_t)$. From (3.3.2) and (3.3.5) it follows that each dual problem instance that we need to solve has an objective function equal to $f(\mathbf{s})$ plus a linear function of $\mathbf{s}$. While all of the problem instances in our numerical experiments will be convex, the exponential operator combined with the linear term can be a source of difficulty. In addition, the permanent price impact implies that $\mathbf{p}_t$ is a function of $\mathbf{s}_{0:t-1}$ so that we cannot reduce the problem to a separable convex problem which is easily solved. Instead we solve each dual problem instance by using the following SQP approach which is also applied to the tax-aware portfolio allocation problem in Section 2.3.3:

1. Choose a starting point, $\hat{\mathbf{s}}$.

2. Approximate $f(\mathbf{s})$ with a second order Taylor expansion about $\hat{\mathbf{s}}$ to obtain

$$\hat{f}(\mathbf{s}) \;:=\; f(\hat{\mathbf{s}}) + \nabla f(\hat{\mathbf{s}})'(\mathbf{s}-\hat{\mathbf{s}}) + \frac{1}{2}(\mathbf{s}-\hat{\mathbf{s}})'\nabla^2 f(\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}}) \qquad \text{(B.2-7)}$$

   where $\nabla f(\hat{\mathbf{s}})$ and $\mathrm{H}f(\hat{\mathbf{s}})$ are, respectively, the gradient vector and Hessian matrix of f evaluated at $\hat{\mathbf{s}}$.

3. Solve

$$\min_{\mathbf{s}\in\mathbb{S}} \quad \hat{f}(\mathbf{s}) + \sum_{t=0}^{T-1}\Big(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t})\Big) \qquad \text{(B.2-8)}$$

   which is a constrained convex quadratic programming problem and therefore easy to solve. Let $\mathbf{s}^{opt}$ be the optimal solution.

4. Evaluate the objective value in (3.3.2) at $\mathbf{s}^{opt}$, and stop if we have converged to within a given error tolerance. Otherwise set $\hat{\mathbf{s}} = \mathbf{s}^{opt}$ and return to step 2.

In our numerical experiments we use an absolute error tolerance of 10^{-5}. Depending on the level of risk aversion, γ, this corresponds to a relative error tolerance between 10^{-4} and 10^{-5}. Typically we find that convergence occurs after just two or three iterations.

B.3 The Model of Bertsimas, Hummel and Lo

Bertsimas, Hummel and Lo [10] (BHL) was one of the earliest papers to consider the portfolio execution problem. We use the calibrated model parameters of BHL to: (i) investigate their conjecture that an OLFC policy should be close to optimal and (ii) provide a simple demonstration of how duality can be used to determine *in advance* whether or not a particular market feature, in this case cross-price impacts, are worth accounting for in a portfolio execution policy. In particular, by considering how small the calibrated cross-price impact parameters are in BHL, it is reasonable to conjecture that simply following the optimal single stock execution policies should be close to optimal. We confirm this by using the single stock value functions to construct a dual penalty which we then use to bound how far the single-stock policy is from optimality. This example therefore serves as an illustration of how a policy and the dual penalty from a given model can be used to determine *in advance* whether a more complicated model even needs to be considered. Note that we only claim to show that cross-price impact parameters are insignificant *within* the model of Bertsimas et al [10] and do not make this claim more generally.

B.3.1 Model Description

The linear percentage price-impact model of BHL has dynamics

$$\mathbf{p}_t = \tilde{\mathbf{p}}_t + \boldsymbol{\delta}_t \tag{B.3-1}$$

$$\boldsymbol{\delta}_t = \tilde{\mathbf{P}}_t(\mathbf{A}\tilde{\mathbf{P}}_t\mathbf{s}_t + \mathbf{B}_t\mathbf{x}_t) \tag{B.3-2}$$

$$\tilde{\mathbf{p}}_{t+1} = \exp(\mathrm{Diag}(\boldsymbol{\epsilon}_{t+1}))\tilde{\mathbf{p}}_t \tag{B.3-3}$$

$$\mathbf{x}_{t+1} = \mathbf{C}\mathbf{x}_t + \boldsymbol{\eta}_{t+1} \tag{B.3-4}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \mathbf{s}_t \tag{B.3-5}$$

where $\tilde{\mathbf{P}}_t := \text{Diag}[\tilde{\mathbf{p}}_t]$, a diagonal matrix with $\tilde{\mathbf{p}}_t$ on the diagonal. The execution price $\mathbf{p}_t$ is the sum of the no-impact price, $\tilde{\mathbf{p}}_t$, and the price impact, $\boldsymbol{\delta}_t$. $\tilde{\mathbf{p}}_t$ follows a vector-geometric Brownian motion, where the $\boldsymbol{\epsilon}_t$'s are IID normal with mean vector, $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}$, and covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. The price impact, $\boldsymbol{\delta}_t$, is temporary and assumes the percentage impact is a linear function of the dollar values, $\tilde{\mathbf{P}}_t\mathbf{s}_t$, and an information vector, $\mathbf{x}_t$, that represents market or private information and is assumed to follow a vector autoregressive process with a one period time-lag. The $\boldsymbol{\eta}_t$'s are IID multivariate normal with mean $\mathbf{0}$ and covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$.

BHL's objective was to minimize the expected execution cost which leads to the following problem formulation:

$$
\begin{aligned}
\min_{\mathbf{s}_t \in \mathcal{F}_t, t=1,\ldots,T} \quad & \mathbb{E}_0\left[\sum_{t=0}^{T} \mathbf{p}_t'\mathbf{s}_t\right] \\
\text{s.t.} \quad & \sum_{t=0}^{T} \mathbf{s}_t = \mathbf{w}_0
\end{aligned}
$$

and dynamics (B.3-1) to (B.3-5). This problem can be solved using dynamic programming and explicit solutions for the optimal value function as well as the optimal trading quantities can be found in BHL. When non-negativity constraints $\mathbf{s}_t \geq 0$ are imposed, however, then it is not possible to solve this problem explicitly. Moreover, because of the return predictability induced by $\mathbf{x}_t$, we expect the non-negativity constraints to be binding in general.

B.3.2 Sub-Optimal Policies

We consider several different sub-optimal policies that can be employed when non-negativity constraints are imposed.

The Simple Policy: The agent buys the same quantity of shares in each of the $T+1$ time periods so that $\mathbf{s}_t = \mathbf{w}_0/(T+1)$.

A One-Step Look-Ahead Policy: At each time t the agent solves

$$\begin{aligned} \min_{\mathbf{s}_t \in \mathcal{F}_t} \quad & \mathbb{E}_t[\mathbf{p}_t' \mathbf{s}_t + V_{t+1}(\tilde{\mathbf{p}}_{t+1}, \mathbf{x}_{t+1}, \mathbf{w}_{t+1})] \\ \text{s.t.} \quad & \mathbf{0} \leq \mathbf{s}_t \leq \mathbf{w}_t \end{aligned} \tag{B.3-6}$$

and implements the optimal solution, $\mathbf{s}_t^{os}$, say. We take V_{t+1} to be the value function for the unconstrained problem which can be computed analytically via a matrix recursion. While the calculations are somewhat tedious the expectation of V_{t+1} conditional on $\mathcal{F}_t$ can be computed in closed form.

A Rolling Open-Loop Feedback Control (OLFC) Policy: At each time t the agent first solves

$$\begin{aligned} \min_{\mathbf{s}_{t:T} \in \mathcal{F}_t} \quad & \mathbb{E}_t \left[\sum_{j=t}^{T} \mathbf{p}_j' \mathbf{s}_j \right] \\ \text{s.t.} \quad & \sum_{j=t}^{T} \mathbf{s}_j = \mathbf{w}_t \\ & \mathbf{s}_j \geq \mathbf{0} \text{ for } j = t, \ldots, T. \end{aligned} \tag{B.3-7}$$

Let V_t^{ol} denote the optimal value of (B.3-7) and let $\mathbf{s}_{t:T}^{ol,t} := [\mathbf{s}_t^{ol,t\,\prime} \ \ldots \ \mathbf{s}_T^{ol,t\,\prime}]'$ denote the corresponding optimal solution. The OLFC policy implements $\mathbf{s}_t^{ol,t}$ at time t and ignores $\mathbf{s}_{t+1}^{ol,t}, \ldots, \mathbf{s}_T^{ol,t}$.

The Single-Stock Execution Policy: Here the agent simply assumes the cross-price impacts are zero in which case the problem decouples into n separate single-stock problems. The agent solves each of these problems using a simple approximate dynamic programming algorithm and implements the resulting policy. This policy was suggested by the fact that the diagonal elements of the matrix $\mathbf{A}$, as calibrated by BHL, were typically an order of magnitude larger than the off-diagonal elements. This can be seen from Table B.2 in Appendix B.3.4 where we display a 10×10 sub-matrix of the price impact matrix, $\mathbf{A}$. We therefore expected this policy to perform well. Moreover, this case provides a clear example of where the duality technology could be used to determine whether or not a given policy, i.e. the

n single stock policies, should be adapted to account for a new feature, i.e. the cross-price impact costs, that are known to exist in the market-place.

B.3.3 The Dual Problem

A dual problem instance is obtained by simulating paths of $\tilde{\mathbf{p}}_t$ and $\mathbf{x}_t$ and then solving the resulting deterministic optimization problem with a dual-feasible penalty that is obtained from some approximation to the value function. A dual problem instance therefore takes the form

$$\begin{aligned} \min_{\mathbf{s}_t \in \mathbb{S}} \quad & \sum_{t=0}^{T} \mathbf{p}_t' \mathbf{s}_t + \sum_{t=0}^{T-1} \Big(\mathbb{E}_t[\tilde{V}_{t+1}(\mathbf{s}_{0:t})] - \tilde{V}_{t+1}(\mathbf{s}_{0:t}) \Big) \\ \text{s.t.} \quad & \mathbf{p}_t = \tilde{\mathbf{p}}_t + \tilde{\mathbf{P}}_t(\mathbf{A}_t \tilde{\mathbf{P}}_t \mathbf{s}_t + \mathbf{B}_t \mathbf{x}_t) \end{aligned} \tag{B.3-8}$$

where the $\tilde{\mathbf{p}}_t$'s and $\mathbf{x}_t$'s have been simulated according to the true model dynamics and are known to the decision-maker at time $t = 0$. We only consider $\tilde{V}_{t+1}(\mathbf{s}_{0:t})$'s that are linear in the actions, $\mathbf{s}_{0:t}$, and this can be achieved in the manner described in Section 3.3.1. In the numerical results below, we compute dual bounds using penalties constructed from: (i) the unconstrained value function, V_t and (ii) the value function, V_t^{ss}, obtained by summing together the n single stock value functions.

B.3.4 Numerical Results

We use the same calibration for $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}\}$ as given by BHL. The agent therefore needs to purchase $100,000$ shares in each of 25 securities. There are $T+1=20$ time periods. Primal bounds are computed on the basis of $10,000$ Monte-Carlo paths while only 100 paths were required to estimate the dual bounds. Monte-Carlo results are displayed in Table B.1. We use V^s, V^{os}, V^{ol} and V^{ss} to denote the estimated average execution cost of the simple, one-step look-ahead, OLFC and single-stock policies, respectively. The mean and standard errors of the reported execution costs are cents per share. We also report run-times (in seconds) for each of the policies with the exception of the simply policy whose value can be determined analytically. Note that these run times are for the entire 10,000 paths. We see that the OLFC policy yields the lowest execution cost but that all policies provide very similar performance. We note, however, that computing the OLFC policy is particularly time-consuming relative to the other policies.

Table B.1 also reports two dual bounds, V_d^{unc} and V_d^{ss}, that were obtained by using the unconstrained value function and the single-stock ADP value function, respectively, to construct dual penalties, $\tilde{V}_t$, in (B.3-8). We also report the unconstrained optimal value function, V_0, which is itself a lower bound on the optimal execution cost for the constrained problem. We see the two dual bounds generated are very similar and are very close indeed to the primal bounds.

Confirming the Conjecture of BHL

These results allow us to confirm the conjecture of BHL, namely that the OLFC policy

Table B.1: Simulation results for the portfolio execution problem with non-negativity constraints.

	sub-optimal policy				dual bound		
	V^s	V^{os}	V^{ol}	V^{ss}	V_0	V_d^{unc}	V_d^{ss}
mean	5.364	4.123	3.979	4.194	2.023	3.976	3.975
std error	0	0.0523	0.0501	0.0542	0	0.00797	0.00404
run times	N/A	324	7465	13.81	N/A	14.51	14.54

should be close to optimal. Indeed the average cost-per-share of the OLFC policy, 3.979 cents, is only .003 cents greater than the best dual bound.

Using the Dual Methodology to Determine Whether Cross-Price Impacts Need to be Included

We also note that if this problem were ever encountered in practice, then it would be possible to "solve" it without constructing any of the more elaborate policies that explicitly account for the cross-price impacts. To see this note that the average execution cost of the single-stock policy is 4.194 cents-per-share and that the dual bound constructed by using the sum of the single-stock value functions to construct dual penalties, is 3.975 cents-per-share. This implies that building a more elaborate execution policy that incorporates the cross-price impacts will be worth at most $4.194 - 3.975 = .219$ cents-per-share. As such, we may decide that it is not worth explicitly accounting for the cross-price impacts, at least in this model. This is one of the principal benefits of the dual technology.

We also mention here that we performed additional numerical experiments that included sector balance constraints in the problem formulation. We again found that the various policies performed very well, at least using the calibrated parameters of BHL.

Table B.2: Price impact coefficients for a 10×10 subset of the 25 stocks in BHL. All values have been multiplied by 10^{10}.

	AHP	AN	BLS	CHV	DD	DIS	DOW	F	FNM	GE
AHP	12.40	-1.69	-1.99	1.04	-1.07	1.09	1.26	1.37	-1.97	-0.17
AN	-1.32	10.10	-1.96	1.09	-1.63	-1.24	2.68	-2.03	-0.80	-0.30
BLS	3.49	0.83	14.40	2.26	-2.45	4.25	-1.02	-5.80	1.68	1.18
CHV	2.09	-0.17	2.34	21.20	2.84	1.62	-0.03	3.60	-3.35	0.91
DD	-0.93	0.66	6.18	1.55	11.70	-1.17	1.00	0.67	1.92	2.24
DIS	1.98	2.73	-0.30	-2.59	0.92	19.30	2.22	0.13	8.27	2.37
DOW	-0.79	0.18	-3.88	-0.23	-1.59	-0.93	7.21	-3.18	1.66	0.05
F	-0.23	0.66	3.29	-0.65	4.24	1.74	3.15	21.90	2.12	0.65
FNM	0.01	-0.92	-0.43	3.27	0.39	-3.57	3.40	2.78	13.70	-0.88
GE	0.77	-0.45	0.53	0.03	-0.41	-0.26	-1.10	0.12	1.60	5.06

B.4 Additional Calibration Details for Section 3.4

B.4.1 Variance-Covariance Dynamics

As stated in Section 3.4.2 we assume that $\mathbf{\Sigma}_t$ follows an O-GARCH model as in Alexander [1] so that

$$\mathbf{\Sigma}_t = \mathbf{F}\mathbf{\Omega}_t\mathbf{F}' + \mathbf{\Upsilon} \tag{B.4-1}$$

where $\mathbf{\Omega}_t$ is a diagonal matrix, $\mathbf{F}$ is a matrix of factor loadings and $\mathbf{\Upsilon}$ is a diagonal matrix of idiosyncratic variances. The diagonal elements in $\mathbf{\Omega}_t$ are assumed to follow independent GARCH(1,1) processes.

Our calibration of the variance-covariance dynamics is very simple and is also based on Alexander [1]. We first compute the variance-covariance matrix, $\mathbf{\Sigma}$, of standardized 5-minute returns, treating the observations for each security as IID. We then perform a principal components analysis on $\mathbf{\Sigma}$, and then assume the first k principal components follow independent GARCH$(1,1)$ models. We emphasize here that we do not claim that this is a particulary good choice of model. Indeed we expect the proprietary models that are used in practice to be much better than our model which is only intended to help demonstrate how the duality techniques can be used.

We use 5-minute return data on the 50 stocks between October 12 and October 25 2011 which corresponds to a total of $d = 10$ trading days. Ignoring time-of-day and other effects we obtain a series of $10 \times 78 = 780$ observations that we initially treat as IID. After normalizing each time series by subtracting the mean return, we compute the covariance matrix, $\mathbf{\Sigma}$, of the $n = 50$ normalized observation series. We then perform a principal components analysis on $\mathbf{\Sigma}$ and obtain

$$\mathbf{\Sigma} \ = \ \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}' \qquad \text{(B.4-2)}$$

where $\mathbf{\Gamma}$ is the matrix of eigenvectors, $\mathbf{c}_1, \ldots, \mathbf{c}_n$ say, and $\mathbf{\Lambda}$ is the diagonal matrix of corresponding eigenvalues, $\lambda_1, \ldots, \lambda_n$, arranged in decreasing order. We let $\mathbf{F}$ be the $n \times k$ matrix containing the k eigen vectors, $\mathbf{c}_1, \ldots, \mathbf{c}_k$, corresponding to the k largest eigen values. We now approximate the covariance matrix with

$$\mathbf{\Sigma} \ \approx \ \mathbf{F}\mathbf{\Omega}\mathbf{F}' + \mathbf{\Upsilon} \qquad \text{(B.4-3)}$$

where $\mathbf{\Omega} = diag(\lambda_1, \ldots, \lambda_k)$ and $\mathbf{\Upsilon}$ is a diagonal matrix chosen to ensure that the diagonal terms on both sides of (B.4-3) agree. Note that λ_i is the variance of the i-th principal component, $\mathbf{c}_i$.The eigenvalue analysis is shown in Table B.3 where we see that the first six

Table B.3: Eigenvalue analysis

Component	Eigenvalue	Cumulative Explained Variance
$\mathbf{c}_1$	1.2879	0.5776
$\mathbf{c}_2$	0.3261	0.7239
$\mathbf{c}_3$	0.1801	0.8047
$\mathbf{c}_4$	0.1399	0.8674
$\mathbf{c}_5$	0.0491	0.8894
$\mathbf{c}_6$	0.0348	0.9051

principal components explain more than 90% of the total variance in the 50 return series. We therefore chose $k = 6$. The problem with (B.4-3) is that it provides a static description for the covariance-matrix, $\mathbf{\Sigma}$. We can use it to create a dynamic model, however, by setting

$$\mathbf{\Sigma}_t \approx \mathbf{F}\mathbf{\Omega}_t\mathbf{F}' + \mathbf{\Upsilon} \tag{B.4-4}$$

where the diagonal elements of $\mathbf{\Omega}_t$ are assumed to follow independent GARCH$(1,1)$ processes. In particular, let $\sigma_{i,t}$ denote the value of the i-th diagonal element of $\mathbf{\Omega}_t$. Then the GARCH$(1,1)$ model for $\sigma_{i,t}$ assumes

$$\sigma_{i,t+1}^2 = \omega_i + \alpha_i c_{i,t}^2 + \beta_i \sigma_{i,t}^2 \tag{B.4-5}$$

where $\omega_i > 0$ and $\alpha_i, \beta_i \geq 0$ are fixed parameters and $c_{i,t}$ is the value of the i-th principal component at time t. The parameters for the six GARCH models were fitted using standard MLE techniques and are given in Table B.4. We note that all t-statistics are significant.

Table B.4: GARCH(1,1) parameter estimates for the top 6 principal components

	ω		α		β	
	coefficient	t-stat	coefficient	t-stat	coefficient	t-stat
$\mathbf{c}_1$	0.1889	2.381	0.1563	3.988	0.7001	8.165
$\mathbf{c}_2$	0.1360	2.423	0.2519	4.198	0.3284	1.548
$\mathbf{c}_3$	0.0231	3.386	0.0885	3.822	0.7801	14.882
$\mathbf{c}_4$	0.0114	3.357	0.1671	5.02	0.7575	17.895
$\mathbf{c}_5$	0.0046	4.532	0.1845	4.971	0.7322	18.403
$\mathbf{c}_6$	0.0033	3.273	0.2052	5.471	0.7131	15.341

B.4.2 Parameter Values for Top 50 Stocks in the S&P

Table B.5: Parameter values for top 50 stocks in the S&P.

Security Ticker	Initial price	Average Daily Volume (Million)	Annual Volatility	Permanent Linear Price Impact Coefficient $\times 10^9$	Temporary Linear Price Impact Coefficient $\times 10^6$
AAPL	407.33	22.85	21.53%	178.3009	8.9150
ABT	52.47	24.65	21.55%	21.2826	1.0641
AIG	22.63	7.44	47.10%	30.4169	1.5208
AMGN	57.39	4.99	20.66%	114.9911	5.7496
AMZN	236.84	6.15	33.77%	385.3714	19.2686
AXP	45.93	10.36	31.70%	44.3491	2.2175
BAC	6.51	263.47	51.04%	0.2469	0.0123
BRK/B	74.03	6.44	24.55%	115.0245	5.7512
C	28.40	68.25	52.89%	4.1609	0.2080
CAT	81.80	10.63	37.09%	76.9853	3.8493
CMCSA	23.16	15.95	29.09%	14.5161	0.7258
COP	67.69	11.22	24.62%	60.3056	3.0153
CSCO	17.16	44.23	26.29%	3.8795	0.1940
CVS	34.44	8.04	19.53%	42.8311	2.1416
CVX	98.08	8.61	26.07%	113.9225	5.6961
DIS	32.95	10.10	25.96%	32.6170	1.6309
EMC	23.26	24.72	31.23%	9.4084	0.4704
GE	16.25	63.55	31.44%	2.5571	0.1279
GOOG	548.10	4.09	24.39%	1340.9765	67.0488
GS	98.05	7.83	41.66%	125.2213	6.2611
HD	34.92	10.67	24.01%	32.7156	1.6358
IBM	185.80	7.32	18.14%	253.9455	12.6973
INTC	23.00	84.04	24.81%	2.7367	0.1368
JNJ	64.10	11.39	18.39%	56.2684	2.8134
JPM	32.71	51.60	42.89%	6.3386	0.3169
KFT	34.67	8.23	16.83%	42.1164	2.1058
KO	67.22	16.75	16.81%	40.1212	2.0061
MCD	89.35	6.36	17.66%	140.5327	7.0266
MMM	76.77	5.19	28.50%	147.8743	7.3937
MO	27.89	11.77	19.24%	23.7031	1.1852
MRK	32.24	13.86	19.48%	23.2673	1.1634
MSFT	27.18	54.55	22.81%	4.9830	0.2491
ORCL	31.35	27.79	26.15%	11.2801	0.5640
OXY	81.25	4.72	35.25%	171.9937	8.5997
PEP	62.45	8.12	16.41%	76.9138	3.8457
PFE	18.85	36.61	23.23%	5.1491	0.2575
PG	64.71	9.26	13.87%	69.9185	3.4959
PM	65.82	8.40	19.55%	78.3178	3.9159
QCOM	52.29	13.83	28.01%	37.8144	1.8907
SLB	67.46	12.53	41.87%	53.8395	2.6920
T	28.83	24.74	16.46%	11.6527	0.5826
UNH	47.21	9.23	35.99%	51.1342	2.5567
UPS	68.32	4.67	23.28%	146.3456	7.3173
USB	24.15	18.35	37.35%	13.1607	0.6580
UTX	74.31	4.79	27.04%	155.0823	7.7541
V	90.99	3.97	25.88%	229.3355	11.4668
VZ	36.64	13.44	17.03%	27.2555	1.3628
WFC	26.38	48.78	39.73%	5.4085	0.2704
WMT	55.15	12.59	16.54%	43.8191	2.1910
XOM	76.87	21.67	20.25%	35.4807	1.7740

Appendix C

Review of Duality Based on Information Relaxations

We begin with a general finite-horizon discrete-time dynamic program with a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. Time is indexed by the set $\mathcal{T} := \{0, \ldots, T\}$ and the evolution of information is described by the filtration $\mathbb{F} = \{\mathcal{F}_0, \ldots, \mathcal{F}_T\}$ with $\mathcal{F} = \mathcal{F}_T$. We make the usual assumption that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so that the decision maker starts out with no information regarding the outcome of uncertainty. There is a state vector, $x_t \in \mathbb{R}^n$, whose dynamics satisfy

$$x_{t+1} \;=\; f_t(x_t, u_t, w_{t+1}), \qquad t = 0, \ldots, T-1 \tag{C.0-1}$$

where $u_t \in U_t(x_t) \subseteq \mathbb{R}^m$ is the control taken at time t, w_{t+1} is an $\mathcal{F}_{t+1}$-measurable random disturbance and $U_t(x_t)$ is the feasible control set at time t. A *feasible* policy, $u := (u_0, \ldots, u_T)$ is one where each individual action satisfies $u_t \in U_t(x_t)$ for all t. We let $\mathcal{U}$ denote the set of such policies. A feasible *adapted* policy is a feasible policy that is $\mathcal{F}_t$-adapted. We let $\mathcal{U}_{\mathbb{F}}$ denote the set of all such $\mathcal{F}_t$-adapted policies. For example, in the context of the portfolio execution problem of Chapter 3, a feasible but not $\mathcal{F}_t$-adapted strategy would be an execution schedule that satisfies the non-negativity constraints in all time periods but where the number of shares purchased in a given period is allowed to depend on prices in later periods. The objective is to select a feasible adapted policy, u, to minimize the total loss,

$$g(u) \;:=\; \sum_{t=0}^{T} g_t(x_t, u_t)$$

where we assume without loss of generality that each $g_t(x_t, u_t)$ is $\mathcal{F}_t$-measurable. In particular, the decision maker's problem is then given by

$$V_0^*(x_0) \;\;:=\;\; \inf_{u \in \mathcal{U}_{\mathbb{F}}} \mathbb{E}_0 \left[\sum_{t=0}^{T} g_t(x_t, u_t) \right] \tag{C.0-2}$$

where the expectation in (C.0-2) is taken over the set of possible outcomes, $w = (w_1, \ldots, w_T) \in \Omega$. Letting V_t^* denote the time t value function for the problem (C.0-2), the associated dynamic programming recursion is given by

$$V_t^*(x_t) \;\;:=\;\; \inf_{u_t \in U_t(x_t)} \left\{ g_t(x_t, u_t) \;+\; \mathbb{E}_t \left[V_{t+1}^* \left(f_t(x_t, u_t, w_{t+1}) \right) \right] \right\} \qquad t = 0, \ldots, T \tag{C.0-3}$$

with the understanding that $V_{T+1}^* \equiv 0$ (and therefore does not depend on w_{T+1} which is undefined). In practice of course it is often too difficult or time-consuming to perform the iteration in (C.0-3). This can occur, for example, if the state vector, x_t, is high-dimensional or if the constraints imposed on the controls are too complex or difficult to handle. In such circumstances, we must be satisfied with sub-optimal policies. These policies are generally easy to simulate and can therefore be used to construct unbiased upper bounds on the optimal value function, $V_0^*(x_0)$. Duality methods based on information relaxations can be used to construct lower bounds on $V_0^*(x_0)$.

We will now briefly describe the theory behind these duality methods. BSS [16] should be consulted for further details. We omit most of the technical details and we will only consider *perfect information* relaxations as these relaxations are usually most useful in practice and are all that we will require in this disertation. Let $\mathcal{S}$ denote the space of real-valued measurable functions that are defined on the state space $\mathbb{R}^n$. We now define an operator Δ that maps $\mathcal{S}$ to the space of real-valued measurable functions on $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ according to

$$(\Delta V_t)(x_t, x_{t-1}, u_{t-1}) \;\; := \;\; V_t(x_t) \; - \; \mathbb{E}\left[V_t(x_t) \,|\, x_{t-1}, u_{t-1}\right]. \tag{C.0-4}$$

Loosely speaking, Δ is an operator on (approximate) value functions. Note in particular that $\mathbb{E}_0\left[(\Delta V_t)(x_t, x_{t-1}, u_{t-1})\right] = 0$ for all integrable V_t. Let $\mathcal{D}$ be the space of real-valued functions on $\mathbb{R}^n \times \mathcal{T}$ such that if $V \in \mathcal{D}$ then $V_t := V(\cdot, t)$ is measurable and $\mathbb{E}_0\left[|V_t(x_t)|\right] < \infty$ for all $t \in \mathcal{T}$ and all feasible policies $u \in \mathcal{U}$. We now define an operator $F : \mathcal{D} \to \mathcal{S}$ according to

$$FV(x) \;\; := \;\; \mathbb{E}_0\left[\inf_{u \in \mathcal{U}} \left\{ g_0(x_0, u_0) \; + \; \sum_{t=1}^{T} \left(g_t(x_t, u_t) \; - \; (\Delta V_t)(x_t, x_{t-1}, u_{t-1})\right) \right\} \;\middle|\; x_0 = x \right]. \tag{C.0-5}$$

Note that the infimum in (C.0-5) is over the space $\mathcal{U}$ of feasible policies and not the space $\mathcal{U}_{\mathbb{F}}$ of feasible adapted policies. Our first result is weak duality.

Theorem C.0.1. (Weak Duality) $V_0^*(x_0) \;\geq\; FV(x_0)$ *for all* $V \in \mathcal{D}$.

Proof: Using the definition of $V_0^*(x_0)$ in (C.0-2) and the fact that the (ΔV_t)'s have zero mean we have

$$\begin{aligned} V_0^*(x) &= \inf_{u \in \mathcal{U}_{\mathbb{F}}} \mathbb{E}_0 \left[\sum_{t=0}^{T} g_t(x_t, u_t) \;-\; \sum_{t=1}^{T} (\Delta V_t)(x_t, x_{t-1}, u_{t-1}) \;\middle|\; x_0 = x \right] && \text{(C.0-6)} \\ &= \inf_{u \in \mathcal{U}_{\mathbb{F}}} \mathbb{E}_0 \left[g_0(x_0, u_0) \;+\; \sum_{t=1}^{T} \left(g_t(x_t, u_t) \;-\; (\Delta V_t)(x_t, x_{t-1}, u_{t-1}) \right) \;\middle|\; x_0 = x \right] \\ &\geq \mathbb{E}_0 \left[\inf_{u \in \mathcal{U}} \left\{ g_0(x_0, u_0) \;+\; \sum_{t=1}^{T} \left(g_t(x_t, u_t) \;-\; (\Delta V_t)(x_t, x_{t-1}, u_{t-1}) \right) \right\} \;\middle|\; x_0 = x \right] \\ &= FV(x_0). \;\blacksquare && \text{(C.0-7)} \end{aligned}$$

Theorem C.0.1 suggests that we can compute a lower bound on $V_0^*(x_0)$ by evaluating $FV(x_0)$ for any $V \in \mathcal{D}$. It gives no guidance, however, on how to choose V so that the lower bound is as large as possible. In fact we can formulate the following dual problem

$$\textbf{Dual Problem:} \qquad \sup_{V \in \mathcal{D}} \; FV(x_0). \tag{C.0-8}$$

Theorem C.0.2 below is a strong duality result which states that the dual problem is solved by taking $V = V^*$ and that there is no duality gap between the primal DP and the dual problem.

Theorem C.0.2. (Strong Duality) *For all* x, $V_0^*(x) = FV^*(x)$.

Proof: By weak duality we need only show that $FV^*(x) \geq V_0^*(x)$. We have

$$\begin{aligned}
FV^*(x) &= \mathbb{E}_0\left[\inf_{u\in\mathcal{U}}\left\{g_0(x_0,u_0) \;+\; \sum_{t=1}^{T}\left(g_t(x_t,u_t) \;-\; (\Delta V_t^*)(x_t,x_{t-1},u_{t-1})\right)\right\} \;\middle|\; x_0 = x\right] \\
&= \mathbb{E}_0\left[\inf_{u\in\mathcal{U}}\left\{V_0^*(x_0) \;+\; \sum_{t=1}^{T-1}\left[g_t(x_t,u_t) \;+\; \mathbb{E}\left[V_{t+1}^*(x_{t+1}) \,|\, x_t,u_t\right] - V_t^*(x_t)\right]\right.\right. \\
&\qquad\qquad\qquad\qquad \left.\left. +\; g_T(x_T,u_T) - V_T^*(x_T)\right\} \;\middle|\; x_0 = x\right] \\
&\geq V_0^*(x) \qquad\qquad (\text{C.0-9})
\end{aligned}$$

where the inequality in (C.0-9) follows because of (C.0-3) and since $V_T^*(x_T) = g_T(x_T,u_T)$. ■

Strong duality suggests that we might be able to obtain good dual bounds on the optimal value function, $V_0^*(x)$, if we can find $V \approx V^*$ and then compute $FV(x)$. The numerical experiments for the portfolio execution problem of Chapter 3 and in the literature cited in Section 3.1 support this claim.

Remark C.0.1. *The portfolio execution problem that we consider in Chapter 3 has an objective function of the form* $\mathbb{E}_0\left[\exp\left(\gamma\sum_{t=0}^{T}\mathbf{p}_t'\mathbf{s}_t\right)\right]$. *This is not in the form* $\mathbb{E}_0\left[\sum_{t=0}^{T} g_t(x_t,u_t)\right]$ *as we assumed in (C.0-2) but this does not present any problems because we can introduce an additional state variable,* z_t *say, with dynamics* $z_{t+1} = z_t + \mathbf{p}_{t+1}'\mathbf{s}_{t+1}$ *for* $t = 0,\ldots,T-1$ *with* $z_0 := \mathbf{p}_0'\mathbf{s}_0$. *We can then take* $g_0 \equiv g_1 \equiv \cdots \equiv g_{T-1} \equiv 0$ *and* $g_T := \exp(\gamma z_T)$. *The portfolio execution problem now has an objective function of the appropriate form. We also note that given any policy,* θ *say, the corresponding time* $t+1$ *value function,* V_{t+1}^{θ}, *can be written as* $V_{t+1}^{\theta} = \exp\left(\gamma\sum_{j=0}^{t}\mathbf{p}_j'\mathbf{s}_j\right)\,\mathbb{E}_{t+1}^{\theta}\left[\exp\left(\gamma\sum_{j=t+1}^{T}\mathbf{p}_j'\mathbf{s}_j\right)\right]$ *where the expectation* $\mathbb{E}_{t+1}^{\theta}[\cdot]$ *is with* *respect to the probability measure induced by the policy* θ. *This also explains why we want to include the first term on the right-hand side of (3.3.3).*