OPTIMIZING DYNAMIC PORTFOLIO SELECTION

By

HALEH VALIAN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Industrial and Systems Engineering

written under the direction of

Professor Mohsen A. Jafari

and approved by

____

____

____

____

____

New Brunswick, New Jersey

October 2009

ABSTRACT OF THE DISSERTATION

OPTIMIZING DYNAMIC PORTFOLIO SELECTION

by HALEH VALIAN

Dissertation Director: Professor Mohsen A. Jafari

In this dissertation, a control-theoretic decision model is proposed for an agent to "optimally" allocate and deploy its financial resources over time among a dynamically changing list of opportunities (e.g., financial assets), in an uncertain market environment. This control-theoretic approach is unique in the sense that it solves the problem at distinct time epochs over a finite time horizon. The solution is a sequence of actions with the objective of optimizing a reward function over that time horizon.

While the above problem is quite general, we will focus solely on the problem of dynamic financial portfolio management. The dynamic portfolio model looks at the portfolio as a moving object to achieve a maximal expected utility for a given risk level and time horizon. We tackle this problem using Semi-Markov Decision Processes and develop an efficient solution methodology based on the Q-learning algorithm. The performance of the model is analyzed, and results from the model are compared to a known market index.

The "optimal" portfolio management policy is then extended to configurations whereby only incomplete information is available. Furthermore, quality of information and its impact on the decision making process is assessed. Here the market environment is characterized by its volatility and price dynamics. The existence of other agents in the market place, who can act adversarial or collaborative, further complicates the underlying price dynamics. The complexity of interactions among different agents is an important challenge for the dynamic portfolio management problem. We fully examine this challenge using a game-theoretic approach to determine the optimal actions of non-price-taking agents with and without a debt constraint.

Acknowledgements

I would like to thank Professor Mohsen Jafari, chairman of my committee, for his outstanding guidance and support during the process of researching and writing this dissertation. It has been such a precious opportunity for me to work with him over this period. He has been a great advisor and is a person of integrity. Without his support, I could not have completed my dissertation.

I would also like to thank Professor Mocanu, Professor Elsayed, Professor Longo and Professor Boucher for their time, support and advice.

And I want to thank Ms. Ellen Molly from JP Morgan Chase for her encouragement and support over this period

I am deeply indebted to my husband, Davood Golmohammadi, who provided me with the love, time, support, inspiration, and motivation needed to finish this dissertation. And my overriding debt is to my parents, for their love, encouragement and support. I am fortunate to have a wonderful family.

Table of Contents

ABSTRACT .... **ii**

Acknowledgements .... **iv**

List of Tables .... **viii**

List of Illustrations .... **viii**

INTRODUCTION AND PROBLEM DEFINITION **1**

1.1. Problem Definition 1

1.2. Major Contributions 2

1.3. General Overview 2

1.4. Objectives 3

1.5. Motivation 3

1.6. Scientific and Technical Merits of This Research 5

1.7. Related Literature Review 6

1.7.1. Agent 6

1.7.2. Machine Learning 7

1.7.3. Markov Decision Process 8

1.7.4. Portfolio Management or Asset Management 9

PROBLEM FORMULATION **17**

2.1. Problem Definition 17

2.2. Problem Formulation Contribution 17

2.3. Defining the Model 17

2.4. Asset Price 19

2.5. Asset Management 21

2.5.1. States 21

2.5.2. Actions 22

2.5.3. Transitions 22

2.5.4. Rewards 22

2.5.5. Decision Epochs 23

2.5.6. Utility Function 24
2.5.7. Objective Function 27
2.6. Solving the Model 28
2.7. Validation of the Model 30
2.8. Numerical Example 30

ANALYZING THE RESULTS 33
3.1. Major Contributions 33
3.2. General Overview 33
3.3. Dynamic Portfolio Management 34
3.3.1. Asset Price and Scenario Generation 35
3.3.1.1. Simple Asset Price Model 37
3.3.1.2. Mean Reverting Model 37
3.3.1.3. Ornstein-Uhlenbeck Model 38
3.3.1.4. Geometric Brownian Motion Model 38
3.3.1.5. Square Root Brownian Motion Model 40
3.4. Asset Allocation 40
3.4.1. Dynamic Programming and Q-learning 41
3.4.2. Learning 41
3.5. Experimental Result 42
3.5.1. Static Portfolio Management for Two Assets 42
3.5.2. Dynamic Portfolio Management for Two Assets 46
3.5.3. Dynamic Portfolio Management for 47 Assets 47
3.6. Conclusion 50

IMPACT OF INCOMPLETE INFORMATION 52
4.1. Problem Definition 52
4.2. Major Contributions 52
4.3. General Overview 52
4.4. Incomplete Information 54
4.5. Information Structure 55
4.6. Assumptions 55
4.7. The Methodology 56

4.7.1. Determining the Model for Asset Pricing 57
4.7.2. Determining the Model for Drift of Asset Prices 58
4.7.3. Applying Filtering Theory and Minimizing Noises 59
4.7.4. Learning Process 61
4.7.5. Optimizing the Portfolio 64
4.8. Examining, the Impact of Uncertainty on Agent's Action 65
4.9. Summary and Conclusions 68

IMPACT OF AGENT ACTIONS 70
5.1. Problem Definition 70
5.2. Major Contributions 70
5.3. General Overview 70
5.4. Game Theory 71
5.5. Market 73
5.6. Asset Price Model 74
5.7. The Game in the Market 75
5.7.1. Definition: Nash Equilibrium 76
5.8. Game without any Constraint 76
5.8.1. The Model 77
5.8.2. Static Portfolio 78
5.8.3. Dynamic Portfolio 81
5.9. Game with a Debt Constraint 82
5.9.1. Model 84
5.9.2. Structure and Notation 84
5.10. Conclusion 86

CONCLUSION 87

REFERENCES 89

CURRICULUM VITA 97

List of Tables

Table 1.1: Definitions of Agents ... 6
Table 3.1: Models Used in Scenario Generation ... 36
Table 3.2: Curve Fitting Results ... 45

List of Illustrations

Figure 1.1: Methods for Solving a Dynamic Portfolio Management ... 4
Figure 1.2 : The Standard Reinforcement Learning Model ... 8
Figure 1.3: Explanatory Grid for Markov Models ... 9
Figure 2.1: The Basic Setup for Multi-Period Investment Model ... 24
Figure 2.2: Some Popular Utility Functions ... 25
Figure 2.3: Observed Price of Kinder Management Stock ... 31
Figure 2.4: Histogram of Logarithm Changes of Kinder Morgan Management Stock ... 31
Figure 3.1: Model of Dynamic Portfolio Management ... 34
Figure 3.2: Wal-Mart Stock Prices ... 43
Figure 3.3: Best Fitted Distributions on Wal-Mart Stock Price ... 44
Figure 3.4: Fitted Geometric Brownian Motion with Jump on Wal-Mart Stock Price ... 44
Figure 3.5: Q-value of One Stock and One Bond for One Period ... 46
Figure 3.6: Q-value of 250 Optimal Portfolios ... 47
Figure 3.7: A Snapshot of Q-learning Program in MATLAB ... 48
Figure 3.8: Realized Return of Portfolio Scenarios ... 49
Figure 3.9: Mean-Variance Performance and Comparison with DJIA Benchmark ... 49
Figure 3.10: Drawdowns of Portfolio Scenarios ... 50
Figure 4.1 : The Process of Selecting Optimal Actions with Incomplete Information ... 57
Figure 4.2 : The Process of Predicting Asset Prices ... 60
Figure 4.3: Conceptual Overview of Filtering Theory ... 61
Figure 5.1: The Timeline of the Game ... 75
Figure 5.2: The Timeline of Game with a Debt Constraint ... 83

CHAPTER 1

INTRODUCTION AND PROBLEM DEFINITION

1.1. Problem Definition

A fundamental decision faced by an individual (or an agent in a virtual sense) is how to optimally allocate and deploy its available resources at each epoch during a time horizon in an uncertain environment. From a control-theoretic point of view, one can think of an agent as a controller that must optimize its reward function at each decision epoch by selecting appropriate actions from its action space. The controller must make decisions in lieu of the following challenges:

1. The environment is only minimally controllable at its best, with a stochastic return function. Thus, the impact of the agent's control actions may not necessarily be positively rewarding.
2. There are other agents in the environment, which may act adversarially or collaboratively with respect to this agent, depending on their own reward systems.
3. The agent solution space may be too large for the amount of time that it has to respond appropriately. Thus, it must seek simple and fast solutions.
4. The agents are not fully aware of the underlying processes that generate rewards.

While the above problem is quite general, we will focus solely on the problem of dynamic financial portfolio management. Here the environment is characterized by market volatility and price dynamics which are often controlled by factors outside of the agent's action space; and other asset owners and managers which could dynamically act adversarially or collaboratively.

1.2. Major Contributions

In this research, we tackle the dynamic portfolio selection problem using Semi-Markov Decision Processes, and develop an efficient solution methodology based on the Q-learning algorithm. Our control-theoretic approach is unique in the sense that the problem is solved at distinct time epochs over a finite time horizon. The solution is an "optimal" sequence of actions with the objective of optimizing a reward function over that time horizon. Chapters 2 and 3 provide details of our methodology and give illustrative examples. In Chapter 4, we compare the optimal sequence of investors' actions under complete and incomplete information configurations. We show how the quality of information influences the agents' sequence of actions. Chapter 5 focuses on the development of game-theoretic and collaborative approaches among several agents.

1.3. General Overview

Markowitz (1952), in his pioneering work, formulated a single period portfolio management problem where decisions were made according to the mean and variance of the portfolio. In his model, all agents chose the same risky assets regardless of their attitudes toward risk. Furthermore, he assumed that the optimal portfolio was independent of the agent's investment time horizon. In reality, different agents hold different portfolios, and a rational agent's optimal portfolio is dependent on the investment time horizon. Many extensions of the Markowitz model exist in the literature.

In the dynamic portfolio management problem, decisions are made more than once over a planning horizon. The decision epochs are not necessarily equally spaced over time. The original works on this topic were pioneered by Merton (1969) and Samuelson (1969) in continuous-time and by Fama (1970) in discrete-time. These works led to substantial insights into the properties of optimal portfolio policies. Prompted by these pioneering works, many researchers have been studying different aspects of the dynamic portfolio management problem. It turns out that in most cases, the optimal portfolio weights over time cannot be derived.

1.4. Objectives

In this dissertation, a control-theoretic decision model is proposed for an agent to allocate and deploy its financial resources over time among a dynamically changing list of opportunities (financial assets) in an uncertain market environment. The problem is formulated as a Semi-Markov Decision Process (SMDP), and Q-learning is applied as a solution methodology. We use real world data to experiment with the models and validate the reasonableness of our solutions. The impact of incomplete information and other agents' actions on the optimal policies of our agent are also investigated, and solutions are presented.

1.5. Motivation

In recent years, there has been a growing interest in the development of resource allocation models, which enable agents to handle the uncertainty of future returns. This type of problem has been categorized as dynamic portfolio management in finance, which can also be referred to as asset management, investment management and money management, and can be approached by different techniques.

Different methods have been applied to asset management problems. While there have been many theoretical papers in recent years, the solutions often have complicated forms that are hard to interpret. The usual methods for solving dynamic portfolio management are summarized in Figure 1.1. Dynamic programming, Martingale and Stochastic programming methods are not applicable to real world problems, since the computational time required for the generation of an optimal solution grows exponentially with the number of variables. Also, the increased data requirement limits the application of these methods to only a relatively small number of assets. In real life, however, portfolio managers have to choose profitable investments from among a large number of assets.

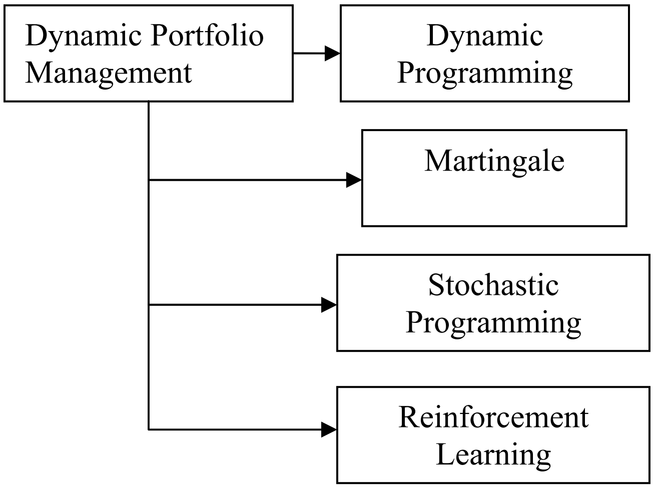


Figure 1.1: Methods for Solving Dynamic Portfolio Management

Reinforcement learning method (known as Neuro-Dynamic Programming) solves problems with a large amount of data [XufreCasqueiroa 2006]. A number of reinforcement learning models have been applied to the dynamic asset management problem. These models include restrictive assumptions and simplifications of the market characteristics. Their objectives are the acquisition of an optimal action under which the agent achieves the maximal average reward from the environment [Lee 2002]. In our work, we tackle dynamic asset management in a reinforcement learning framework to determine the optimal sequential trading actions.

1.6. Scientific and Technical Merits of This Research

A particularly challenging class of portfolio management problems involves dynamic cases where decisions are made at multiple time epochs during a planning horizon, with information on uncertain parameters revealed only incrementally but on a progressive basis over that horizon. Under uncertainty, the construction of models requires that we distinguish factual from non-factual (e.g., merely speculative) information, and find appropriate mechanisms to reconcile our knowledge [Tapiero 1998].

The methodology we are about to present is general, but in this research we focus our attention on its application to the dynamic portfolio management problem. We consider the problem of optimal portfolio choice in a financial market with one bond and *n* assets. This problem is formulated as a Semi-Markov Decision Process, and Q-learning is used to solve it. Time is assumed to be continuous, so that system states change continually between discrete decision epochs. This is unlike regular MDPs, where the states change only due to actions assumed in the model. The dynamics of the model include the following steps:

Step 1: Observe the historical data of asset prices at time 0.

Step 2: Forecast the price of assets till time *T*.

Step 3: Compute the agent's wealth for this period.

Step 4: Compute the agent's expected utility function of wealth.

Step 5: Optimize the expected utility function and find the agent's optimal actions.

Step 6: Perform the first optimal action.

Step 7: Go back to Step 2.

The formulation of this methodology is explained in Chapter 2.

1.7. Related Literature Review

1.7.1. Agent

A variety of definitions of "Agent" are found in the literature. Russell (1995) defined an agent as anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors. The various definitions discussed in the literature involve a host of properties of an agent. Some of these definitions are listed in **Error! Not a valid bookmark self-reference.**.1.

Table 1.1: Definitions of Agents

Name	**Definition**
Reactive	Responds in a timely fashion to changes in the environment based on local information [Sycara 2003]
Proactive	Has ability to take the initiative; is not driven solely by events, but is capable of generating goals and acting RATIONALLY to achieve them [Wooldridge 1995]
Goal-Oriented	Does not simply act in response to the environment; plans to achieve goals with domain knowledge [Smith 1994]
Autonomous	Senses the environment and acts on it, over time, in pursuit of its own agenda so as to effect what it senses in the future [Franklin 1996]
Learning	Changes its behavior based on its previous experience [Franklin 1996]
Communicative	Communicates with other agents and solves problems by collaboration and synergy [Wooldridge 1995]
Mobile	Has ability to transport itself from one machine to others [Petrie 1996]
Intelligent	Attempts to make the best decisions based on a given performance measure [Vlassis 2003]

For the purpose of this research, *an agent is defined as an intelligent being that has one or more goals and is capable of communicating with other agents in its environment. Furthermore, agents are goal-oriented and react to changes in the environment.*

1.7.2. Machine Learning

Machine Learning addresses the question of how to program agents to automatically learn and improve with experience. Machine Learning algorithms fall into three groups with respect to the kind of feedbacks that the system/learner has access to [Duba 2000]:

- *Supervised Learning* is based on a given sample of input-output pairs (also called the training sample), and the task is to find a deterministic function that maps any input to an output such that disagreement with future input-output observations is minimized.
- *Unsupervised Learning* is based on the similarities and differences among the input patterns without any feedback from the environment.
- *Reinforcement Learning* is a sub-area of machine learning concerned with how an agent ought to take actions in an environment in order to maximize its long-term reward. Reinforcement learning models attempt to find a policy that maps states of the world to the actions the agent should take in those states [Sutton 1998].

Learning by trial and error and optimal control came together in the late 1980s to produce the modern field of reinforcement learning. Minsky (1954) in his Ph.D. dissertation discussed computational models of reinforcement learning and described his construction of Stochastic Neural-Analog Reinforcement Calculators. In the 1960s, the terms "reinforcement" and "reinforcement learning" were used in the engineering literature for the first time (e.g., [Waltz 1965], [Mendel 1966] and [Fu 1970]).

The standard reinforcement learning model is depicted in Figure 1.2**Error! Reference source not found.**. In each step of the interaction of the agent with the environment, the agent computes the state *S(t)* of the environment based on the information it receives; the agent then chooses an action, *a(t)*, to generate an output. The action changes the state of the environment to *S´(t),* and the value of this state transition is communicated to the agent through a reward value, *R(S(t),S´(t), a(t))*. The agent should choose actions that tend to increase the long-run sum of its reward values [Kaelbling 1996].

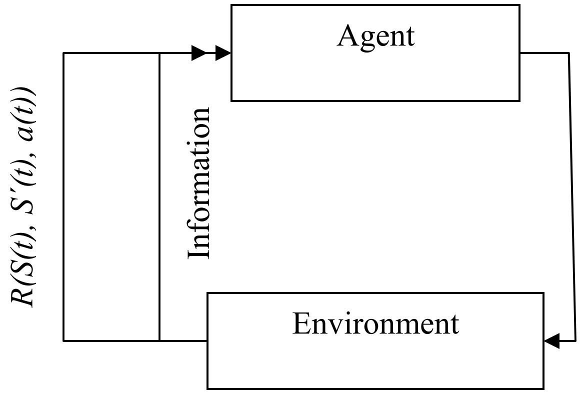


Figure 1.2 : The Standard Reinforcement Learning Model

1.7.3. Markov Decision Process

Markov Decision Processes [Bellman 1957] provide an elegant mathematical framework for modeling and solving sequential decision problems in the presence of uncertainty. Formally, a finite-state Markov Decision Process (MDP) is expressed as *M=(SS,A,P,RR),* where *SS={S(1)... S(n)}* is a set of states, *A* = *{a(1)... a(m)}* is a set of actions, $P:SS\times A\times SS\rightarrow[0,\ 1]$ is a stochastic transition function of state dynamics conditioned on the preceding state and action, RR is a set of reward functions, and R is immediate payoff function of state-action configurations [Puterman 1994]. An MDP represents a controlled stochastic process with described dynamics [Kveton 2006]. In the simplest form, the states and actions of an MDP are discrete and unstructured. The discrete-time MDP can be solved efficiently by standard dynamic programming methods [Bellman 1957], [Puterman 1994] and [Bertsekas 1996]. The explanatory grid for Markov models is shown in Figure 1.3.

		Do we have control over state transitions?	
		No	Yes
Can states be continuous?	No	Markov Chain	MDP Markov Decision Process
	Yes	Semi-Markov Chain	SMDP Semi-Markov Decision Process

Figure 1.3: Explanatory Grid for Markov Models

1.7.4. Portfolio Management or Asset Management

Traditionally, investment is defined as the current commitment of resources in order to achieve later benefits [Luenberger 1997]. Portfolio management is a decision process of dividing the total investment fund among some major asset classes such as equities, bonds, cash, options, etc [Zhao 2000]. To have a good understanding of portfolio management, first we should look at the definition of “portfolio” in the literature. Morgan Stanley's Dictionary of Financial Terms offers the following explanation: “If you own more than one security, you have an investment portfolio. You build the portfolio by buying additional stocks, bonds, mutual funds, or other investments. Your goal is to increase the portfolio's value by selecting investments that you believe will go up in price.”

For some investments, such as bonds, the amount of money to be gained in the future is known. However, in most situations we do not know the amount of money to be gained later, and its determination is complex because stock returns are very often volatile as stated by Grossman et. al. (1981). Over the period of 1871 to 1996, the standard deviation of real annual continuously compounded stock returns in the U.S. was 17.4% [Brennan 2001]. Also the prediction of future price of an asset is complex since all of the available information is already reflected in the history of past prices.

Essential to portfolio management theory is the quantification of the relationship between *risk* and *return,* and the assumption that investors must be compensated for assuming risks [Downes 2006]. Markowitz formulated the portfolio problem as a choice between the mean and variance of a portfolio of assets [Markowitz 1952]. This approach is widely utilized because of its simplicity. Although this model has led to some important results, such as capital asset pricing, its simplicity is its major shortcoming [Zhao 2000]. Considering just the mean return and variance of return of a portfolio is an oversimplification of the problem at hand. Higher order moments, if included, could provide a more complete description of the distribution of portfolio returns [Elton 1997]. In addition, the Markowitz model was developed to find the optimum portfolio when an investor is concerned with return distributions over a single period. A major theoretical problem here is how to modify and extend the single-period model to multi-period problems. Also, the existing solutions to this problem commonly make the assumption of perfect information; i.e., they assume that investors are fully aware of the underlying processes that generate dividends or returns. Yet it is clear that investors are uncertain about the stochastic process generating stock returns and consumption [Brennan 2001].

The multi-period portfolio problem is generally hard to solve; nevertheless, in the literature various approaches have been developed for its solution. Examples include Dynamic Programming (DP), Martingale, Stochastic Programming and Reinforcement Learning. The traditional approach for solving dynamic portfolio optimization problems is dynamic programming (DP). However, this approach suffers from of dimensionality: it cannot handle high dimension problems and requires an exact model of the environment, (see [Samuelson 1969], [Richard 1975], [Kim 1996], [Lioui 2001], [Wachter 2002] and [Liu 2007]).

As a result of the difficulties encountered in DP, there has been considerable interest in the application of the Martingale approach. This interest was generated by Harrison in 1979, and was applied by [Karatzas 1987], [Pliska 1986], [Cox 1989], [Basak 1995], [Grossman 1996], [Detemple 2003] and [Cvitanic 2003]. This approach is implemented if markets are sufficiently "complete", that is, if individuals have full information about the pricing of future states. The degree of market completeness is a major challenge for the practical validity of this approach. It works by transforming a dynamic problem into an optimizing invested wealth problem. It computes the optimal amount for investment and consumption, but not the optimal trading actions [Han 2005].

Another class of approaches uses Stochastic Programming. These approaches provide a useful tool for discrete space approximation with many constraints on investment strategies [Zhao 2000]. They require additional assumptions about the stochastic processes that the agent cannot control.

For all the above-mentioned approaches, the transaction cost should be considered since in a more realistic market model, brokerage fees have to be paid for each transaction. Also, the optimal strategy without transaction costs entails an infinite number of transactions [Merton 1969]. To account for this, some papers considered a market model with transaction costs (such as: [Magil 1976], [Taksar 1988], [Akian 2001], [Davis 1990], [Korn 1998] and [Øksendal 2002]). In this research, we consider the transaction cost in the model of the dynamic portfolio management problem.

1.7.4.1. Agents in Portfolio Management and Financial Markets

The use of agents is certainly not new to financial markets and portfolio management [Grossman 1976]. LeBaron (2006) mentioned several reasons why financial markets are particularly appealing applications for agent-based methods. First, financial time series contain many curious puzzles that are not well understood. Second, financial markets provide a wealth of pricing and volume data that can be analyzed.

In a portfolio management, an agent is a decision maker that solves an optimization problem. For example, buyers and sellers are two commonly encountered types of agents. Snarska et al. (2006) introduced an automatic decision-making system, which allows a single agent to use complex methods of Modern Portfolio Theory.

Gode and Sunder (1993) were interested in just how much “intelligence” was necessary to generate the behavior of many real trading experiments. They ran a computer experiment with agents who will not bid more than what the asset is worth in redemption value. Their results show that in most cases the agents allocate the assets at over 97% efficiency which is very close to that for human [LeBaron 2000]. This means that the agent does not need to be highly intelligent if it follows the structure of the market. However, agents in financial markets may change over time in response to past performance. Lettau (1997) implemented many of the ideas of evolution and learning in a population of traders in a very simple setting that provided a useful benchmark.

1.7.4.2. Markov Decision Process in Portfolio Management

Norman et al. (1965) applied the Markov Decision Process to the portfolio of an insurance company, since such a company has to make decisions each day about how much of its effective bank balance should be invested, in light of random claims, expenses, and call-offs by stockbrokers. The state of the system on any given day is defined by its effective bank balances. The possible states at the next decision epoch depend upon the above events and on the decisions made previously. Norman's objective function was to maximize the growth rate of the company. This is similar to the problem of Bartmann (1980), who applied the Markov Decision Process to credit institutions. Credit institutions have to make daily decisions to figure out how much, and in what form, they should hold cash, in light of possible robberies, and when they need to ask for shipment of cash when shortages appear imminent [Bartmann 1980]. Wessels (1980) also formulated the problem of how much money the bank should hold. The states of the system on any day are its cash levels. The possible states at the next decision epoch depend upon current deposits and withdrawals, and on the decisions made.

The Markov Decision Process was applied by Krawczyk (2000) to optimize the portfolio management problem. Also, the weak Euler scheme was used to make the time evolution of a portfolio discrete, and an inverse distance method was used to describe the transition probabilities. The approximating Markov decision problem was solved by value iteration, and numerical solutions to a few specific portfolio problems were obtained, with varying degrees of accuracy.

Labbi et al. (2007) described a new tool, the IBM Customer Equity Lifetime Management Solution (CELM) which maximizes the return on investment and minimizes the risks (uncertainty). Labbi et al. (2007) applied the Markov Decision Process in order to find the optimal allocation of marketing resources.

Derman et al. (1984) considered an investment problem as a finite-horizon Semi-Markov stochastic dynamic program to maximize expected profit over a finite horizon. He computed the amount of available resources that should be invested as soon as an opportunity presents itself. These opportunities occur with certain probabilities at any given time. At any decision epoch, the state of the system is defined by the level of holdings available for investment. The only random factor is the time interval to the next investment opportunity [White 1993].

1.7.4.3. Machine Learning in Portfolio Management

There are two basic steps involved in portfolio management. The first step is the prediction of asset prices and the second step is the allocation and management of assets. Papers on machine learning in portfolio management can be classified into two groups: predicting future asset prices and managing the assets.

Many machine learning methods have been applied to predicting future asset prices. The prediction of asset prices is an aim of supervised learning, which models the relationship between the input and output [Duda 2000]. Since Artificial Neural Networks

(ANN) is not sensitive to unusual data patterns, there has been a lot of interest in applying it to the prediction of asset prices. ANN was used to predict asset prices by Beltratti et al. (1992), Armano et al. (2005), Lee (2004) and Saad et al. (1998). Kuo et al. (1998) applied Fuzzy Delphi and Fuzzy Neural Networks and constructed the informative macro-economic indicators by a fuzzy equation. Saad et al. (1998) compared time delay, recurrent, and probabilistic Neural Networks for the stock trend prediction.

There is a wide range of clustering techniques for stock selection, such as Random Matrix Theory [Pafka 2004], Chaotic Map Synchronization [Basalto 2005], Potts Magnetization Model [Kullmann 2004], Transfer Entropy [Baek 2005] and Support Vector Machine [Fan 2001]. For example, Fan et al. (2001) used the Support Vector Machine as a classifier of stocks in the Australian stock market, by formulating the problem as a binary classification.

On the other hand, there has not been much interest in applying supervised learning methods to asset management since these models cannot cover goals of asset management. These methods basically try to minimize errors between the input and the predicted output of actions, without considering the effect of action. The major alternative method is reinforcement learning, which tries to achieve the maximal reward from the environment by considering the effect of the original decision.

Asset management has been intensively studied in terms of reinforcement learning. Gao et al. (2000) presented a solution technique for portfolio management scheduling, namely, how to switch between two price series within a Q-learning framework. Moody et al. (2001) formulated portfolio management using direct reinforcement but they focused only on how to switch between a few two-price series. O

(2006) and Neuneier (1998) made assumptions about the market to make the problem manageable in a reinforcement learning framework. Neuneier (1998) formulated the financial market as MDP under some assumptions and simplifications about the market's characteristics, and one year later he modified Q-learning by adding preference to the risk avoiding tendency and to make decisions with less risk [Neuneier 1999]. He focused on how to change one's position to either Dollar or Deutsche Mark. Xiu et al. (2000) proposed a static portfolio management system using Q-learning, and used two performance functions, absolute profit and relative risk-adjusted profit [Xiu 2000].

Most of these papers have treated the problem of asset allocation in a single period. Some others have formulated asset allocation between multiple markets or by considering multiple agents. In this work, in contrast, we aim at solving the dynamic portfolio management problem in a single market with one agent.

CHAPTER 2

PROBLEM FORMULATION

2.1. Problem Definition

Dynamic Portfolio Management or Dynamic Asset Management is the investment of liquid capital in various trading opportunities during a given time horizon. The investor's ultimate goal is to optimize some relevant measure of the trading system performance, such as return, risk, or expected utility function value. In this research, the agent tries to find actions which maximize expected the utility function value for a given risk level and time horizon.

2.2. Problem Formulation Contribution

In this chapter, we give a rigorous mathematical formulation of the problem. We construct a policy (sequence of actions) that is optimal in the sense that, starting from any state, it yields the maximum possible objective function that can be achieved from that state. The dynamic portfolio management problem is formulated as the following optimization problem:

- Maximizing the expected discounted utility between time zero and T
- *Subject to:* everything is reinvested and there is no consumption or labor income during the investment horizon.

2.3. Defining the Model

The asset allocation or portfolio management problem consists of determining how to allocate the available capital to different assets. As Gennotte (1986) showed in a similar setting, the investor's decision problem can be divided into two separate problems: an inference problem, in which the agent updates its estimate of the future value of the asset's price, and an optimization problem, in which it uses its current estimate to choose an optimal portfolio.

Here, optimizing the asset allocation is further modeled in two steps. In the first step, the agent estimates the price of all assets. It is assumed that there are only two types of assets in the market: risk-free assets and risky assets. A risk-free asset, such as a bank account or a bond, offers a known return if held over some period of time. This asset has a deterministic future value and return. A risky asset has a stochastic future value and return. We assume that the rate of return of risky assets is governed by Geometric Brownian Motion. The use of Geometric Brownian Motion to model asset price fluctuations was proposed by Samuelson in 1969 and became widely accepted through the work of Merton in 1990.

In the second step, the agent assigns a value to all possible choices using the expected utility function of wealth. Then it tries to find a policy that maximizes the expected value of utility. The optimal policy determines fractions of wealth invested in each asset over a finite time horizon. The agent invests its wealth in the market among available assets and rebalances its portfolio at any time by incurring some transaction costs. This means that the agent can transfer funds from one asset to other assets at any time. However; there is a penalty for this transaction.

Our model is an abstraction of the real world and, as such, is based on some assumptions. The following simplifications do not necessarily restrict the generalization of the proposed methods, but make our models more tractable from a mathematical standpoint. The assumptions are:

- The agent is small and does not influence the market by its trading.
- The agent always invests the total amount of its wealth.
- There are no taxes.
- There are no restrictions requiring the agent to buy or sell assets at their market price.

The first assumption is relaxed in Chapter 5, where we discuss the optimal portfolio management while considering the impact of agent action.

2.4. Asset Price

As mentioned earlier, the first step of portfolio management is forecasting the future price of all available assets in the market. The possible future asset prices can be generated by one of the scenario-generation models which are discussed in Chapter 3.

In this research, scenarios are generated for the key parameter of a portfolio investor (future assets price) by Geometric Brownian Motion model. Osborne's paper showed that the rate of return on asset prices in the market varies in a similar fashion to molecules in Geometric Brownian Motion (GBM) [Osborne 1972]. The stochastic process of the asset price satisfies the following stochastic differential equation:

$$\frac{ds(t)}{s(t)} = \mu\, dt + \sigma\, dB_t \qquad (2.1)$$

$$\text{or } ds(t) = \mu s(t)\, dt + \sigma s(t)\, dB_t, \qquad (2.2)$$

where B_t is a Brownian motion, *s(t)* is the asset price, and *ds(t)* represents an infinitesimal change in the asset price. In Equation (2.1), proportional changes in the asset price are assumed to have drift μ and volatility σ, which is the standard deviation of asset price change. dB_t is the instantaneous change of B_t and has a drift rate of zero and a variance rate of one. Therefore, the expected value of B_t is equal to its current value.

In Equation (2.2), the first term on the right hand side implies that *ds(t)* has an expected drift rate of $\mu s(t)$ per unit time which is the predicted movement during interval *dt*. $\sigma s(t)\, dB_t$ is the random shock term and represents the noise or variability in the unpredictable path followed by *s(t).* This term is the underlying uncertainty in the model during *dt*. By using Ito's lemma [Luenberger 1997], the following process is obtained:

$$d \ln s(t) = (\mu - \frac{\sigma^2}{2})dt + \sigma\, dB_t .$$

The equation has an analytic solution:

$$s(t) = s(0)\, e^{(\mu - \frac{\sigma^2}{2})t + \sigma\, B_t} ,$$

where *s(0)* is the initial value. *s(t)* is a log-normally distributed random variable with expected value $E(s(t)) = s(0)\, \exp((\mu - \frac{\sigma^2}{2})t + \sigma B_t)$ and variance $Var(s(t)) = e^{2\mu t} s_0^2 (e^{\sigma^2 t} - 1)$.

At this point it is natural to ask how well this model fits actual asset price behavior. First, notice that the price remains positive in this model, which is also the case in the real world. Second, the asset prices in the above model follow a lognormal distribution. Based on an analysis of past asset price records, the price distributions of most assets are actually quite close to lognormal [Luenberger 1997]. In Chapter 3, other asset price models will be explained and appropriate models for different situations will be discussed.

2.5. Asset Management

The asset management is modeled here by a Semi-Markov Decision Process (SMDP). SMDP provides solutions for stochastic decision-making problems where outcomes are only partially under the control of the decision maker. SMDP can be described by a finite state set, $S \in SS$; a finite set of admissible control actions, $a \in A$, for every state; a set of transition probabilities, $p_{S,S'}(a)$, which describes the dynamics of the system; a return function, $R \in RR$, and a function giving probability distribution of transition time for each state-action, *F*. More precisely, a Semi-Markov Decision Process is a controlled stochastic process characterized by a set of states where in each state there are several actions from which the decision maker must choose. For a state *S* and an action *a*, a state transition function $p_{S,S'}(a)$ determines the transition probabilities to the next state. The agent wants to optimize its total rewards by taking appropriate actions.

2.5.1. States

The state of SMDP at period *t* is represented by *S(t),* which includes prices and positions of holding assets. $s^i(t)$ denotes the price of asset *i* at time *t*, and $x^i(t)$ defines the agent's holding number or its holding positions of asset *i* at time *t.* These variables describe the state of the system completely for the purpose of portfolio management. If this information is available for the current time, then it is not necessary to know anything more about the history of the process. Explicitly, we write *S(t)* as:

$$S(t) = \left(\left(s^1(t),\ldots,s^{n+1}(t)\right),\left(x^1(t),\ldots,x^{n+1}(t)\right)\right).$$

The state space is the set of possible situations that the agent may face. The state space *SS* of the SMDP consists of all possible combinations of prices, and weights for all possible assets.

2.5.2. Actions

The action space *A* consists of all possible actions that an agent can possibly take. An action means the trading volume of each asset. If there are *n* risky assets and one risk-free asset, the action *a(t)* has an *n+1* element vector defined as: $a(t) = \left(a^1(t),...,a^{n+1}(t)\right)$, where $a^i(t)$ is the change in the holding of asset *i* at time *t*. Each agent has a given initial endowment, $x(0)$, of the asset, and it can continuously trade the asset by choosing its action, $a(\tau)$; hence, at time *t* its position, $x(t)$, is:

$$x(t) = x(0) + \int_0^t a(\tau)\, d\tau \,.$$

2.5.3. Transitions

At each decision epoch, the SMDP either stays in the same state or makes a transition to a new state, depending not only on the action taken, but also on the stochastic nature of the environment. If the SMDP is in state *S* and action *a* is chosen, the probability that it will make a transition to state *S′* is expressed as: $p_{S,S'}(a)$.

2.5.4. Rewards

Suppose that the system is originally observed in state *S* and the action *a* is applied. The reward is then given by:

$$R(S,S',a) = \int_0^T \int_0^t e^{-\alpha k}\ u(w_k)\ dk\, dF_{S,S'}(t|a),$$

where $F_{S,S'}(t|a)$ is the probability distribution (given the action *a*) that the transition from *S* to *S′* occurs within time *t*, and *u(w$_k$)* is the utility function of wealth at time *k*. Since time is continuous, the discount factor for an interval of length *k* will be given by the exponential function $e^{-\alpha k}$, where α is the interest rate.

2.5.5. Decision Epochs

The manner in which information flows in the financial market makes the "time interval" different on different trading days. On some days an agent may have more volatile markets than on others. Changing volatility may require changing the basic observation period. The agent may choose finer time intervals depending on the level of volatility. Thus, asset price as a random variable should be defined over a continuous time.

One factor to take into account in making decisions is the length of the time horizon. But by using an appropriate utility function, the dependency between the time horizon and the optimal investment strategy can be reduced. Samuelson (1969) found that the optimal investment strategy is mostly independent of the time horizon by using a logarithmic function for the utility. Decisions are made at the beginning of each time period over a predetermined planning horizon, here assumed to be *T*. It is assumed that time is continuous: $t \in [0,T]$. The infinitesimal intervals are considered and symbolized by dt. The set of decision epochs is defined as: $t = 0, dt,\ 2\,dt \ldots\ T.$ For each dt we observe the state, *S*, choose an action, *a*, and receive a reward, *R*, which is a function of the state and action taken at that decision epoch. Uncertainties are assumed to occur within each time period. The basic setup is shown in Figure 2.1 [Mulvey 2006].

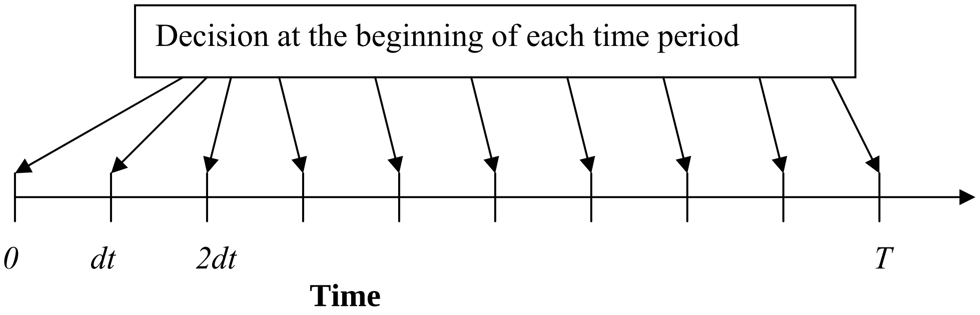


Figure 2.1: The Basic Setup for a Multi-Period Investment Model

2.5.6. Utility Function

In life, there are many situations where agents face two or more choices. The economic "theory of choice" uses the concept of a utility function to describe the way agents make decisions when faced with a set of options. A utility function assigns a value to all possible choices faced by the agent. The higher the value of a particular choice, the greater the utility derived from that choice.

Here, the utility function (u) is defined on the wealth, and its output is a real value. The agent's wealth can be defined as: $w_t = \sum_{i=1}^{n} \left(x^i(t) s^i(t) - c(a^i(t)) \right)$, where $x^i(t)$ is the fraction of wealth invested in the risky asset i in period t; $c(a^i(t))$ is the transaction cost of action $a^i(t)$, and $s^i(t)$ is the price of asset i at period t.

The specific utility function used varies among individuals, depending on their individual risk tolerance and their individual financial environment. However, most agents are more sensitive to losses than to gains. Utility functions of wealth can capture various kinds of risk preferences. Broadly speaking, the aim of our agent is to maximize its expected utilities. The agent does not merely wish to maximize the immediate utility in the current state, but wishes to maximize the utilities it will receive over a period of time in the future. Some popular utility functions are shown in Figure 2.2 [Luenberger 1997]. The exponential utility function is defined as: $u(w_t) = -e^{-aw_t}$ for *a>0*, and the logarithmic utility function is defined as: $u(w_t) = \ln(w_t)$ for $w_t > 0$. If there is any positive probability of obtaining an outcome of 0, the expected logarithmic utility function will be -∞. The power utility function is defined as $u(w_t) = bw_t^b$ for *b≤1* and *b≠0*. The Quadratic utility function is defined as $u(w_t) = w_t - bw_t^2$ for some parameter *b>0*.

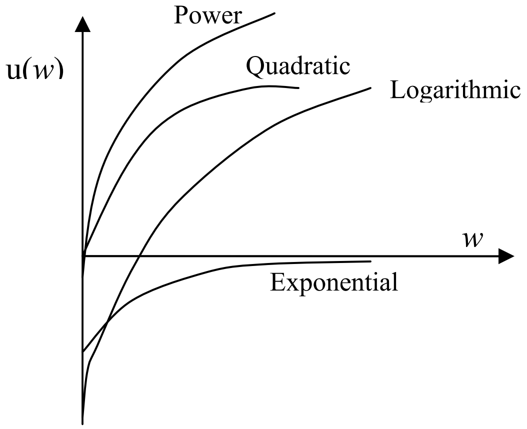


Figure 2.2: Some Popular Utility Functions

The main purpose of a utility function is to provide a systematic way to rank alternatives that captures the principle of risk aversion [Luenberger 1997]. Risk aversion is related to the behavior of an agent under uncertainty. It is measured as the additional marginal reward that an agent requires in order to accept additional risk. In other words, it measures how much utility the agent wants to gain for its investment in risky assets.

The risk aversion coefficient is related to the magnitude of the bend in the utility function. Arrow and Pratt proposed the following formula for a risk aversion coefficient:

Arrow- Pratt coefficient of relative risk aversion $= -\frac{w_t\ u''(w_t)}{u'(w_t)}$ [Damodaran 2008],

with w_t as the wealth accumulated at the end of period t. $u''(w_t)$ is the second derivative of the utility of wealth, and it measures the magnitude of the bend in the utility function. $u'(w_t)$ is the first derivative of the utility of wealth, and it measures how the utility changes as wealth changes. Pratt and Arrow proposed this formula, since the second derivative of the utility function measures the change in the utility itself changes as a function of the wealth level. $u'(w_t)$ appears in the denominator to arrive at a normalized Arrow-Pratt coefficient.

Decreasing this coefficient indicates that the proportion of wealth that agents are willing to put at risk increases. For many utility functions, the Arrow- Pratt coefficients decrease as the wealth increases. This makes sense, since the risk-seeking of an agent depends on the agent's wealth. Many agents are willing to take more risk when they are financially secure.

The risk aversion of an agent depends on the agent's feelings about risk and its current financial situation. The appropriate risk factor and utility function for wealth increments should be determined by an agent's internal feelings toward risk and by an agent's financial environment.

Since power utility functions capture varying degrees of risk sensitivity, in this research, the power utility function is selected: $u(w_t) = \frac{w_t^{1-\beta}}{1-\beta}$. The Arrow-Pratt coefficient of relative risk aversion for this utility function is defined as:

$$-\frac{w_t\ u''(w_t)}{u'(w_t)} = \beta .$$

This coefficient of risk aversion, *β,* is constant. If *β* has a value close to one, the utility function will be logarithmic. The case *β=0* is risk-neutral, and the utility function of wealth corresponds to absolute wealth or profit, while the larger values of *β* correspond to greater sensitivity to loss. In this research, *β=1* has been considered; this means that agents will invest the same percentage of their wealth in risky assets as they get wealthier.

2.5.7. Objective Function

The agent's problem is how to choose portfolio strategies in order to maximize the objective function. The objective function is defined in one of the following ways based on the agent's situation [Han 2005]:

- When the agent lives forever, the objective function is:

  $$F = Max\ E\left[\int_0^\infty e^{-\alpha t}\ u(w_t)\ dt\right],$$

  where E indicates the expected value and $e^{-\alpha t}$ is the discount factor.

- When the agent makes a decision at time *T*, the objective function is:

  $$F = Max\ E\left[u(w_T)\right].$$

- When the agent makes its decision at random time $\widetilde{T}$, where $\widetilde{T}$ is an exponentially distributed random variable, the objective function is:

$$F = Max\, E\left[u(w_{\widetilde{T}})\right].$$

- When the agent makes its decision during the time horizon 0 to *T*, the objective function is:

$$F = Max\, E\left[\int_{0}^{T} e^{-\alpha\, t}\; u(\, w_t)\; dt\right].$$

In this research, the agent makes a decision at each point in time during a time horizon. Therefore, the objective function is considered as maximizing the expected utility or reward function between time zero and *T.* The dynamic portfolio management is formulated as:

$$\begin{aligned} &Max\, E\left[\int_{0}^{T} e^{-\alpha t}\; u(\, w_t)\; dt\right] \\ &St : \sum_{j} a^{j} s^{j}(0) = 0. \end{aligned}$$

The constraint shows that everything is reinvested and there is no consumption or labor income during the investment horizon.

2.6. Solving the Model

The future evolution of the process depends on the current state and on the policy that will be followed. If the process is in state *S* and the policy π is followed, the expected return will be written as $V_{\pi}(S)$. That is,

$$V_{\pi}(S) = \sum_{S'\in SS} p_{S,S'}(a) R(S,S',a) + \sum_{S'\in SS} p_{S,S'}(a) \int_{0}^{\infty} e^{-\alpha t} V_{\pi}(S') dF_{S,S'}(t|a),$$

and $R(S,S',a)=\int_0^T\int_0^t e^{-\alpha k}\;u(w_k)\;dk\,dF_{S,S'}(t|a)$.

The optimal value function for an SMDP satisfies the following Bellman optimality equation:

$$V^*(S)=\underset{a\in A}{Max}\left\{\sum_{S'\in SS}p_{S,S'}(a)R(S,S',a)+\sum_{S'\in SS}p_{S,S'}(a)\int_0^\infty e^{-\alpha t}V^*(S')dF_{S,S'}(t|a)\right\}.$$

This model can be solved by Q-learning:

$$Q^\pi(S,a)=\sum_{S'\in SS}p_{S,S'}(a)R(S,S',a)+\sum_{S'\in SS}p_{S,S'}(a)\int_0^\infty e^{-\alpha t}Q^\pi(S',\pi(S'))dF_{S,S'}(t|a).$$

$Q^\pi(S,a)$ represents the discounted cumulative reward of doing action a in state S and then subsequently following policy π. The optimal Q-function corresponding to state S and action a is:

$$Q^*(S,a)=\sum_{S'\in SS}p_{S,S'}(a)R(S,S',a)+\sum_{S'\in SS}p_{S,S'}(a)\int_0^\infty e^{-\alpha t}\max_{a'\in A}Q^*(S',a')dF_{S,S'}(t|a).$$

This leads to the following Q-learning:

$$Q^{(k+1)}(S,a)=Q^{(k)}(S,a)+\alpha_k\left[\frac{1-e^{-\alpha\tau}}{\alpha}R(S,S',a)+e^{-\alpha\tau}\max_{a'}Q^{(k)}(S',a')-Q^{(k)}(S,a)\right],$$

where α_k is the learning rate and $\frac{1-e^{-\alpha\tau}}{\alpha}R(S,S',a)$ is the sample reward received at τ time units, and $e^{-\alpha\tau}$ is the sample discount on the value of the next state given a transition time of τ time units.

2.7. Validation of the Model

For the validation of the model, the Sharpe Ratio is used. Two strategies will be considered: always investing in the risk-free asset, and the buy and hold strategy. The Sharpe Ratio is a measure of performance of a portfolio over a given period of time. The important aspect of the Sharpe Ratio is that it takes into consideration the portfolio risk. In order to use the Sharpe Ratio, three factors must be known: the portfolio return, the risk-free rate of return, and the variance of the portfolio. The portfolio return is equal to the sum of all individual assets' weights in the portfolio, times their returns. The variance of the portfolio is the sum of the squared weighted variances of the individual assets, plus two times the sum of the weighted pair-wise covariance of the assets. For the risk-free rate of return, the average return (over a period of time) of some government bonds or notes may be used. The Sharpe Ratio has the following formula:

$$\textit{Sharpe Ratio} = (\textit{Portfolio Return - Risk-Free Return}) / \textit{Portfolio Variance}^{\,0.5}\,.$$

2.8. Numerical Example

Here, a very simple numerical example is considered and in Chapter 3 another example is discussed. Suppose that there are two choices for investment: one risk-free and one risky asset. The annual interest rate of the risk-free asset has been set to 2%. The Kinder Management stock is considered as a risky asset. The historical value of the closing price of this stock is shown in Figure 2.3. For the purpose of testing, the observations of stock price from 5/16/2006 to 5/16/2007 are set apart. The remaining data are used for estimating of parameters of asset price. Figure 2.4 shows the histogram of changes of logarithm of Kinder Morgan Management stock price in time period 5/16/2001 to 5/16/2006.

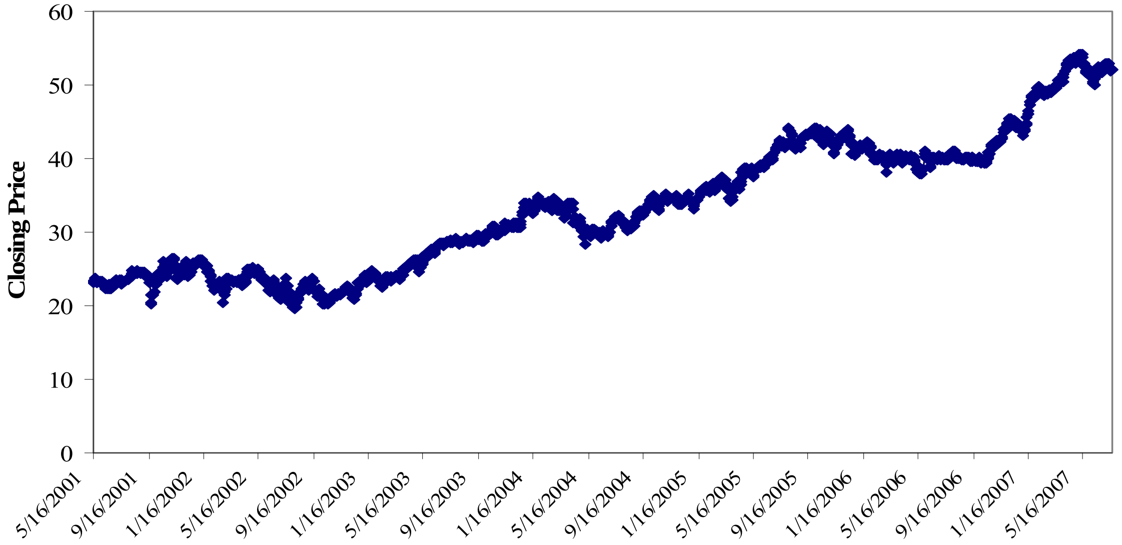


Figure 2.3: Observed Price of Kinder Management Stock

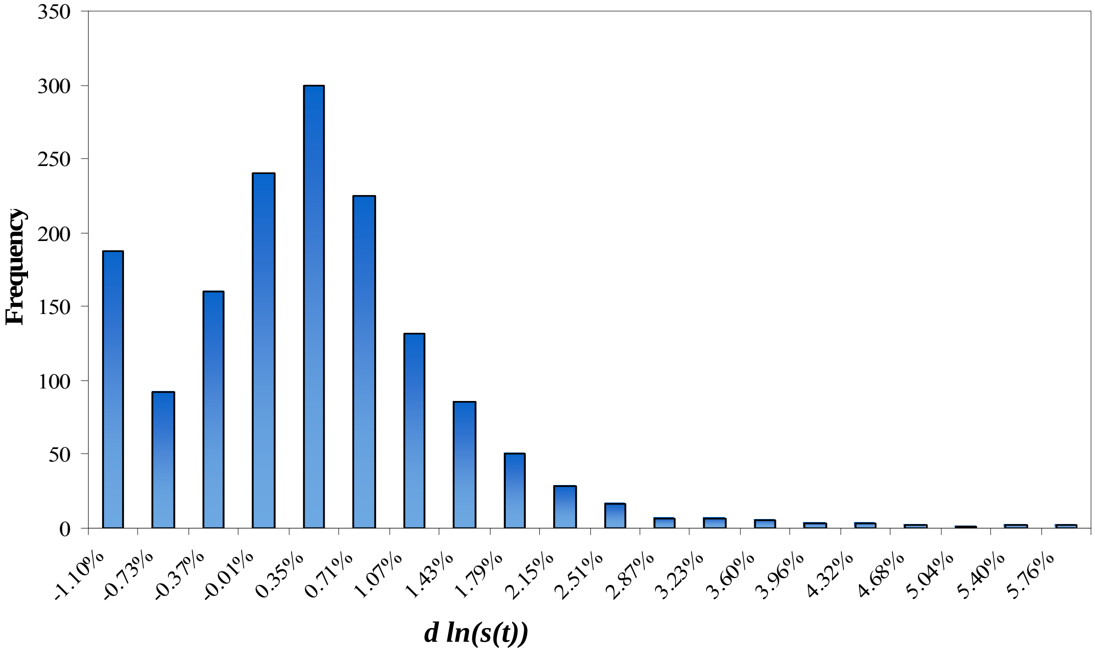


Figure 2.4: Histogram of Logarithm Changes of Kinder Morgan Management Stock

The graph in Figure 2.4 is quite close to log-normal, and on the righthand side tail the observed distribution is larger than a normal distribution. Since the period considered in this example is too short, there are not any significant jumps in the historical data. The future stock price is generated by the Geometric Brownian Motion Method using $s(t) = s(0)\, e^{(\mu - \frac{\sigma^2}{2})t + \sigma\, B_t}$ with parameters estimated at μ=.0638, and σ^2=.005513. These parameters were estimated by the Curvefit Toolbox of MATLAB.

The utility function is considered as $\dfrac{(x(t)s(t) - c(a(t)))^{1-\beta}}{1-\beta}$. Let us consider that β=1 and $c(a(t))$ =*0.* The utility function reduces to $Ln(x(t)s(t))$. If the discount factor is 0.1 and the objective function for trading between time zero and 12 is considered the objective function is: $Max\ E\left[\int_0^{12} e^{-.1\,t}\ Ln(x(t)s(t))\ dt\right]$.

At the end of each month, the agent must decide about the amount of risky asset. We test different approaches as: buy and hold strategy, investing always on the risk-free asset and Q-learning. To compare these approaches, we consider several performance measures, computed within 5/16/2006 and 5/16/2007. These performance measures are profit, and Sharpe Ratio. For this data set, the main results obtained with Q-learning are: profit=1.459 and Sharpe Ratio=2.257. The results obtained from buy and hold strategy are: profit=1.21 and Sharpe Ratio=1.363. In the case of investing in risk-free asset, the profit is 1.019. This shows Q-learning solution gives us a portfolio with a higher profit and a higher Sharpe Ratio.

CHAPTER 3

ANALYZING THE RESULTS

3.1. Major Contributions

In this chapter, we review how a dynamic portfolio management problem may be effectively solved by using Q-learning. The more general case of the methodology is discussed, and an illustrative example is given. The performance of the model is analyzed, and results from the model are compared to a known market index.

3.2. General Overview

Dynamic programming is a common approach for solving optimal control problems. However, for problems with large state or solution space, it is not effective. On the other hand, Q-learning can compute the optimal course of actions to be learned directly, without any requirement for modeling the environment or remembering previous actions for more than a short period of time. We find that the Q-learning framework enables a simpler optimization problem representation, avoiding Bellman's curse of dimensionality.

Dynamic Portfolio Management requires making sequential decisions in stochastic environments to maximize the expected utility function of wealth for a given finite horizon. The optimization problem requires taking into account the current state of the environment. In this model, state $S(t)=(s(t),x(t))$ consists of element $s(t)$, which characterizes the assets price, and element $x(t)$, which characterizes the allocation of the wealth at time t. Note that prices of assets are the only variables which are independent of portfolio decisions. Therefore, asset allocation is a control problem and may not be reduced to a pure prediction problem.

Figure 3.1 demonstrates our modeling paradigm of optimizing the expected utility function of a dynamic portfolio. The model iterates by the agent interacting with the environment through state sequence *S(t)*, selecting actions *a(t)*, and evolving to the next state *S'(t+dt).* There is a cost *c(a(t))* for the agent's actions. This problem is complicated because the investor revises the decision *(a(t))* at every time step.

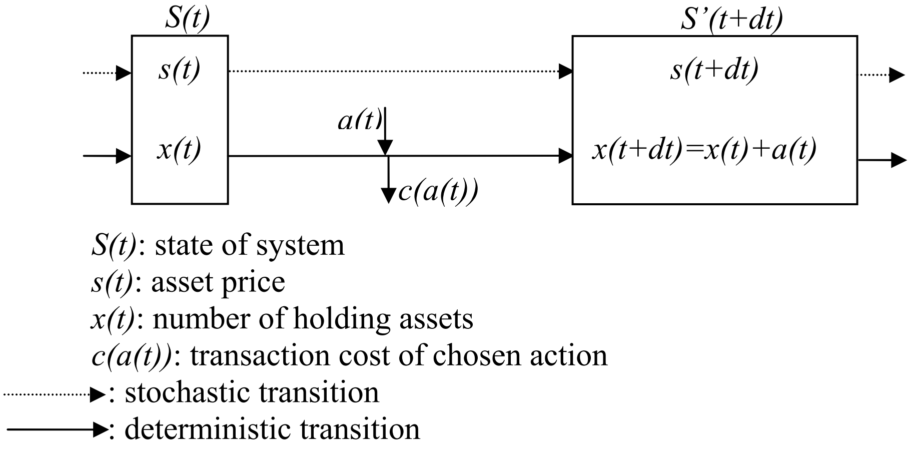


Figure 3.1: Model of Dynamic Portfolio Management

3.3. Dynamic Portfolio Management

To obtain an optimal portfolio, the investor agent has to solve an optimization problem consisting of two steps: (1) estimating the future state (*S(t+dt))*, and (2) maximizing the expected utility function. In the first step, we need only to estimate the asset price, since the number of holding assets has a deterministic transition. In the second step, based on the estimated asset price, a portfolio is formed that takes into account the risk level that the investor is willing to choose.

In the following sections, the asset pricing model (Section 3.3.1) and the search for an optimal portfolio (Section 3.4) are explained. With real financial data, we find that our approach based on Q-learning produces more profit than an approach based on the market index.

3.3.1. *Asset Price and Scenario Generation*

The first step of portfolio management is to forecast the future prices of all available assets in the market, *s(t+dt)*. These prices can be forecasted by one of the scenario generation models. In general, a scenario generation model includes some or all the following steps [Domenica 2007]:

(a) Model assumptions, which explain the behavior of the random parameters (for instance, econometric models for interest rates, etc.).

(b) Estimation/calibration of parameters for the chosen model, using historical data and expert opinions.

(c) Generation of state trajectories according to the chosen model, or discretization of the distributions using approximation of statistical properties.

Some of the models used in scenario generation [Domenica 2007] are given in Table 3.1.

Table 3.1: Models Used in Scenario Generation

Purpose	**Methods**
Generation of Data Trajectories	Econometric Models: Auto Regressive: (AR) Moving Average: (MA) Auto Regressive Moving Average: (ARMA) Generalized Auto Regressive Conditional Heteroscedasticity: (GARCH) Vector Auto Regressive: (VAR) Bayesian VAR Reduced Rank Regression
	Diffusion Processes: Simple Asset Price Models Mean Reverting Models Ornstein-Uhlenbeck Model Geometric Brownian Motion Model Square Root Brownian Motion Model
	Other Method: Neural Networks
Discretization	Statistical Approximation: Property Matching Moment Matching Non-parametric Methods
	Sampling: Random Sampling Stratified Sampling

In recent years, different diffusion process models have been successfully used. Typically, these models deliver the future values of asset prices based on past data. Each model is appropriate in different situations, as follows:

3.3.1.1. *Simple Asset Price Model*

The Simple Asset Price Model is appropriate in practice if, over time, the behavior of the asset prices is stable, trends and variations do not change, and there are no jumps in asset prices. In this case, the drift and diffusion coefficients are independent of the information received over time. This model uses:

$$ds(t) = \mu\, dt + \sigma\, dB_t ,$$

where B_t is a Brownian motion, and $ds(t)$ represents an infinitesimal change in the asset price. The instantaneous change in B_t (dB_t) has a drift rate of zero and a variance rate of one. In this formula, the coefficients μ and σ are constants in time and do not depend on the information set I_t. The behavior of $ds(t)$ seems to fluctuate around a straight line with slope μ. The size of the diffusion determines the extent of the fluctuations around this line.

3.3.1.2. *Mean Reverting Model*

The Mean Reverting Model is appropriate when the asset price has an equilibrium level. There is a major difference between the Mean Reverting Model and the previous model: the deviations around the trend for the Mean Reverting Model are not completely random. If the prices of assets get plucked away from their non-event levels, they revert to more or less the levels they started from, but it may take some time. This model uses:

$$ds(t) = \lambda(\mu - s(t))dt + \sigma\, s(t)dB_t .$$

As $s(t)$ falls below some mean value μ, the drift term $\lambda(\mu - s(t))$ will become positive. This makes $ds(t)$ more likely to be positive, with $s(t)$ eventually approaching towards μ. The rate of mean reversion is controlled by parameter λ>0. As this parameter becomes greater, the excursions become shorter. Every time the term $\sigma\, s(t)dB_t$ gives the asset price a push away from μ, the drift term acts in such a way that the asset price starts heading back to μ.

3.3.1.3. Ornstein-Uhlenbeck Model

The Ornstein-Uhlenbeck Model is a special case of the Mean Reverting Model. It can be used for asset prices that fluctuate around zero. This model uses:

$$ds(t) = -\lambda\, s(t)dt + \sigma\, s(t)dB_t\,,$$

where λ is the rate of mean reversion. As this parameter increases, *s(t)* reverts towards zero at a faster rate.

3.3.1.4. Geometric Brownian Motion Model

The Geometric Brownian Motion Model is appropriate when there is an exponential trend and the variance increases over time, or the graph of historical data is close to log-normal distribution. This model is somewhat more realistic than other asset price models. The Geometric Brownian Motion model satisfies the following formula:

$$ds(t) = \mu\, s(t)dt + \sigma\, s(t)dB_t\,.$$

The drift of this model is $\mu\, s(t)$, and its diffusion is $\sigma\, s(t)$. These coefficients depend on the information set I_t.

The Geometric Brownian Motion is sometimes referred to as the Wiener Process and has been used in physics to describe the motion of particles that are subject to a large number of molecular shocks. Similarly, asset prices are subject to shocks in the market. This process has two properties: (1) Its instantaneous changes in a small period of time follow a normal distribution, and (2) the values of these changes for any two short, non-overlapping time intervals are independent.

The pricing of financial assets in continuous time may not have extreme changes, as long as one ignores rare events. Historically speaking, however, financial markets demonstrate some extreme rare behavioral shocks from time to time. The analysis of prices over historical data reveals sudden and rare breaks logically accounted for by exogenous events (jumps) on historical data [Guo 2004]. It is clear that changes in asset prices are a function of normal events that occur in a continuous fashion and of rare events that occur infrequently.

Given a Brownian Motion process, the small unexpected price changes occur with a variance $\sigma^2\,dt$, where σ may depend on the available information, and *dt* is a small interval. The distribution of these changes is assumed to be normal. In Brownian Motion, when *dt* approaches to zero, the size of the changes becomes smaller and the probability of significant changes approaches zero. Hence, it is not an appropriate model for situations where, in very short intervals, prices can make significant jumps. The rare events have the following properties:

- At each small interval, at most one rare event can occur.
- The information set I_t can not help us to predict the occurrence or the nonoccurrence of the event in the next interval Δt.

Rare events can be modeled by the Pareto-Beta Jump-Diffusion process. At a given time *t*, the price of a risky asset follows [Ramezani 1998]:

$$\frac{ds(t)}{s(t)} = \mu\,dt + \sigma\,dB_t + (Y^u_{N^u(\lambda^u t)} - 1)\,dN^u(\lambda^u t) + (Y^d_{N^d(\lambda^d t)} - 1)\,dN^d(\lambda^d t)\,,$$

where Y^d and Y^u are random variables for the size of up and down jumps, and N^u and N^d are independent Poisson processes with intensity parameters λ^u and λ^d (*u* and *d* represent up and down jumps, respectively). There are two parts in this formula, one described by a Brownian Motion and the other by a Poisson process. The Brownian Motion process is used for modeling the small changes in asset price, while the Poisson process is used for modeling jumps caused by rare events. This model shows occasional jumps due to the Poisson component, but between jumps, the process is not constant; it fluctuates randomly due to Brownian motion. The noise introduced by the Brownian motion is much smaller than the jumps due to the Poisson process.

3.3.1.5. Square Root Brownian Motion Model

If the variance of asset price does not increase too much when the *s(t)* increases, the Square Root Brownian Motion Model should be applied. This model has the following formula:

$$ds(t) = \mu\, s(t)dt + \sigma\sqrt{s(t)}\; dB_t \,.$$

In this formula, the *s(t)* has an exponential trend and its variance increases proportionally to the *s(t)*. Clearly, the fluctuations of this model are more subdued than those of Geometric Brownian Motion Model, but they have similar trends.

3.4. Asset Allocation

Here, asset allocation is formalized as an SMDP. If the state space is small and an appropriate model of the system is available, the SMDP can be solved by Dynamic Programming. If an accurate model of the environment is not available, Q-learning is a viable option. It learns system behavior through trial and error interactions with its dynamic environment.

3.4.1. Dynamic Programming and Q-learning

The optimal value function for an SMDP is the solution of the following Bellman equation:

$$V^*(S) = \underset{a \in A}{Max} \left\{ \sum_{S' \in SS} p_{S,S'}(a) R(S,S',a) + \sum_{S' \in SS} p_{S,S'}(a) \int_0^\infty e^{-\alpha t} V^*(S') dF_{S,S'}(t|a) \right\}.$$

V^* can be found by using the value iteration algorithm. This algorithm assumes that the expected return function $R(S,S',a)$ and transition probabilities $p_{S,S'}(a)$ are known. Q-learning optimizes the problem by sampling state-action pairs and returns while interacting with the system. Let us assume that the agent executes action *a(t)* at state *S*, and the system moves to state S'. Q-learning uses the following equation:

$$Q^{(k+1)}(S,a) = Q^{(k)}(S,a) + \alpha_k \left[\frac{1-e^{-\alpha\tau}}{\alpha} R(S,S',a) + e^{-\alpha\tau} \max_{a'} Q^{(k)}(S',a') - Q^{(k)}(S,a) \right].$$

The selection of action *a(t)* should be guided by the trade-off between the discovery of the possibilities of the environment, and the utilization of the actions that have been discovered so far. At the beginning, the agent chooses actions randomly, but as it learns, it selects actions with larger Q-values with increasingly higher probability.

3.4.2. Learning

The optimal actions can be learned from rewards and punishments. Learning about optimal actions implies that the agent will eventually acquire the ability to follow the maximally optimal action.

The learning of optimal actions may be clarified by considering a portfolio problem with two assets (*A* and *B*). We assume here that the agent may choose either asset *A* or asset *B*, but not both. After choosing an asset, the agent receives a reward. The reward is generated according to a different probability distribution. Successive rewards are independent of each other. The average rewards given by two assets are different. The agent reward depends only on its assets. At each time, the agent may decide which asset to choose based on the rewards it has received so far. If the agent knows that asset *A* gives a higher reward than asset *B*, then clearly its optimal action is to choose asset *A*. But if the agent is uncertain about the relative mean rewards offered by the two assets, and its objective is to maximize its total discounted reward after time period *T*, then the problem becomes interesting. The point is that the agent should try to choose both assets alternately at first, to determine which asset appears to give higher rewards.

3.5. Experimental Result

In this section, the following cases are investigated:

Case 1: one riskless asset and one risky asset in one period

Case 2: one riskless asset and one risky asset in more than one period

Case 3: one riskless asset and 47 risky assets in more than one period.

3.5.1. Static Portfolio Management for Two Assets

We illustrate our approach by using the following example for a single period. Here we consider a single stock (Wal-Mart stock) versus a bond. A total of 265 weekly stock price data, from April 1, 2002, through November 10, 2008, are used, as shown in Figure 3.2:

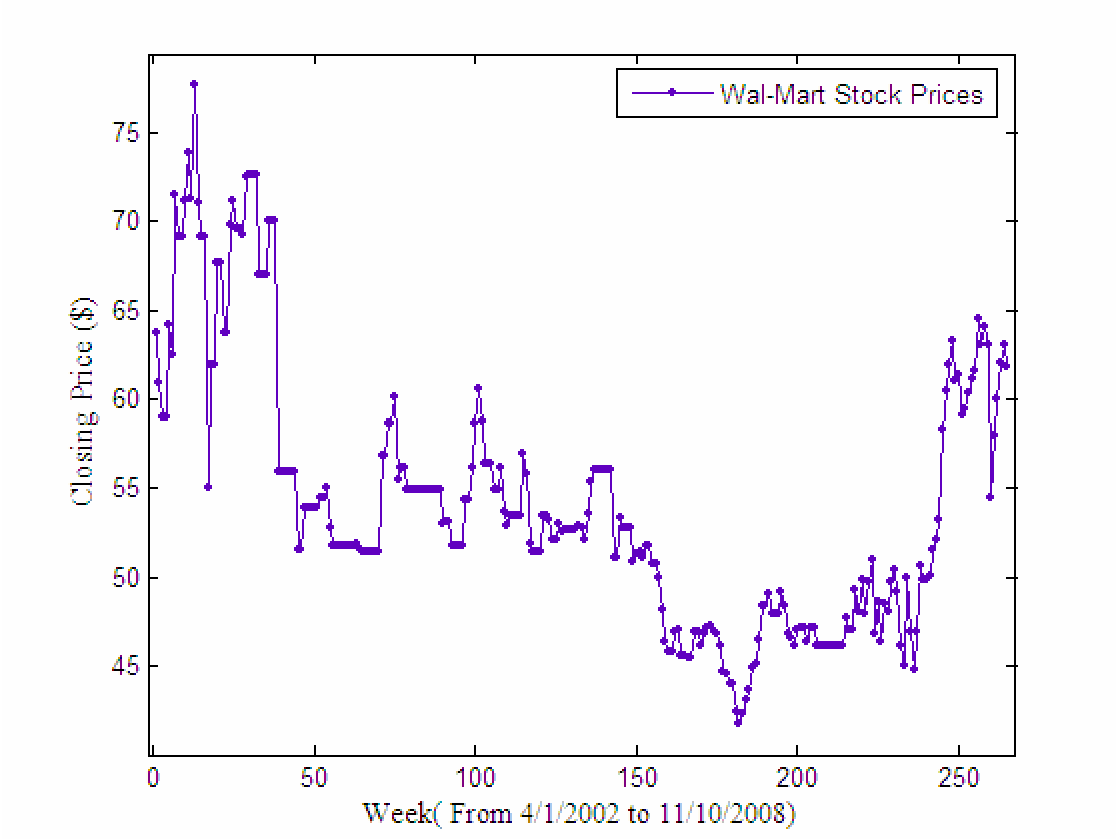


Figure 3.2: Wal-Mart Stock Prices

The first 192 data are used as a training dataset and the remaining points are used as a testing stage. The Curvefit Toolbox of MATLAB is used for fitting the distribution and finding its parameter.

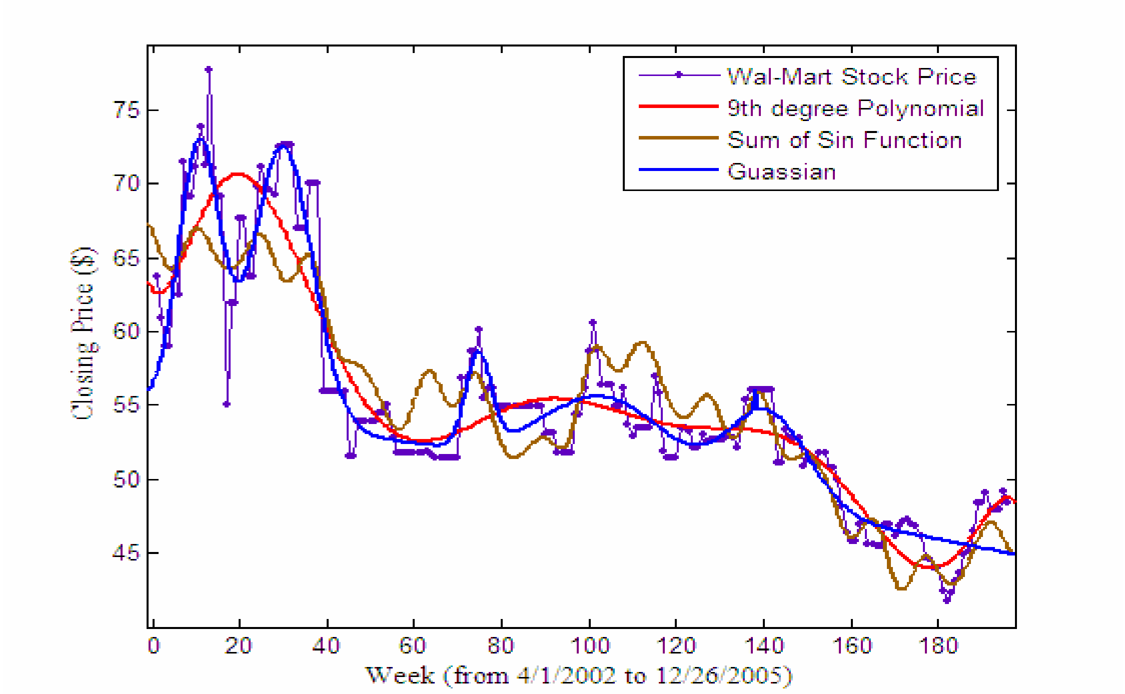


Figure 3.3: Best Fitted Distributions on Wal-Mart Stock Price

Several distributions are examined and fitted in this data set, and three distributions with better fitting results are shown in Figure 3.3. Since there are an exponential trend and some jumps in this data, Geometric Brownian Motion with a jump is also fitted to the data set (Figure 3.4).

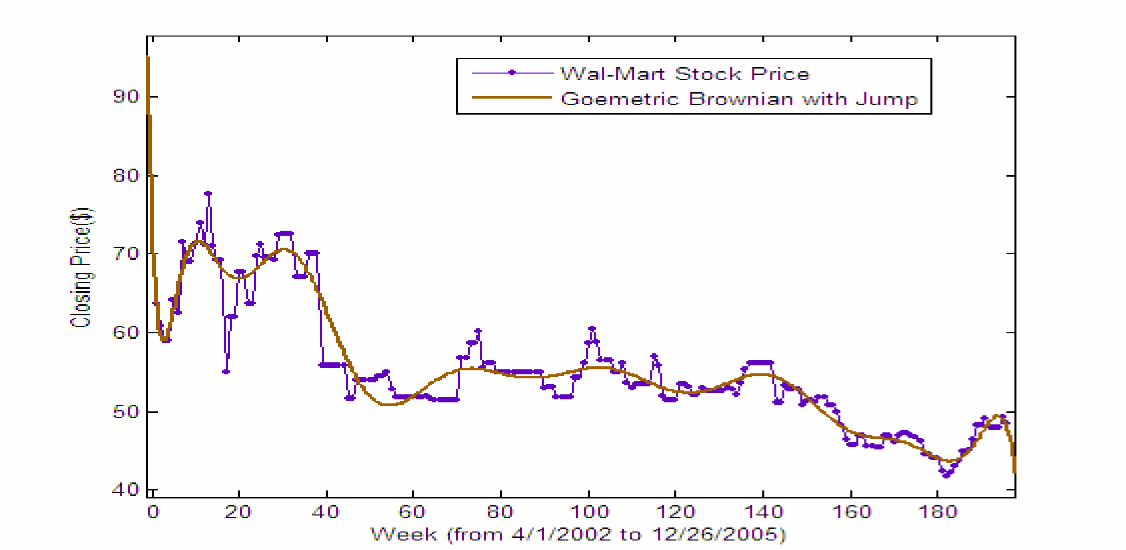


Figure 3.4: Fitted Geometric Brownian Motion with Jump on Wal-Mart Stock Price

The results of Curve fitting are summarized in Table 3.2.

Table 3.2: Curve Fitting Results

Distribution	***SSE***	***R-square***	***Adj R-sq***
Gaussian	946.16	.91392	.9057
Geometric Brownian Motion and Jump	906.29	.91755	.90812
Polynomial	1722.9	.84325	.83567
Sum of Sin	2182.4	.80145	.78249

In this table, SSE means sum of squares due to error. This statistic measures the deviation of the responses from the fitted values of the responses. A value closer to 0 indicates a better fit. The coefficient of multiple determinations is shown as "R-square." This statistic measures how successful the fit is in explaining the variation of the data. A value closer to 1 indicates a better fit. Also, "Adj R-sq" is the degree of freedom of the adjusted R-square. A value closer to 1 indicates a better fit. In this table, the curve fitting of Geometric Brownian Motion with Jump has the best results. Therefore, we choose this distribution for our data.

Suppose the bond has the rate of return .07. The agent should make its decision on investment by choosing the asset (stock) or the bond with the larger Q-value. In this section, the utility function is considered as: $2\ln w_t$. The Q-value is the expected utility function.

By using the Geometric Brownian Motion with Jump model, we can predict that the asset price will decrease in the next period. Obviously, it is preferred to hold the bond during this period. Figure 3.5 shows the Q-values of the holding the bond and the Q-values of the holding the asset. In this figure, the Q-value of the asset is lower than the Q-value of the bond. Therefore, the optimal policy for the next period is holding the bond.

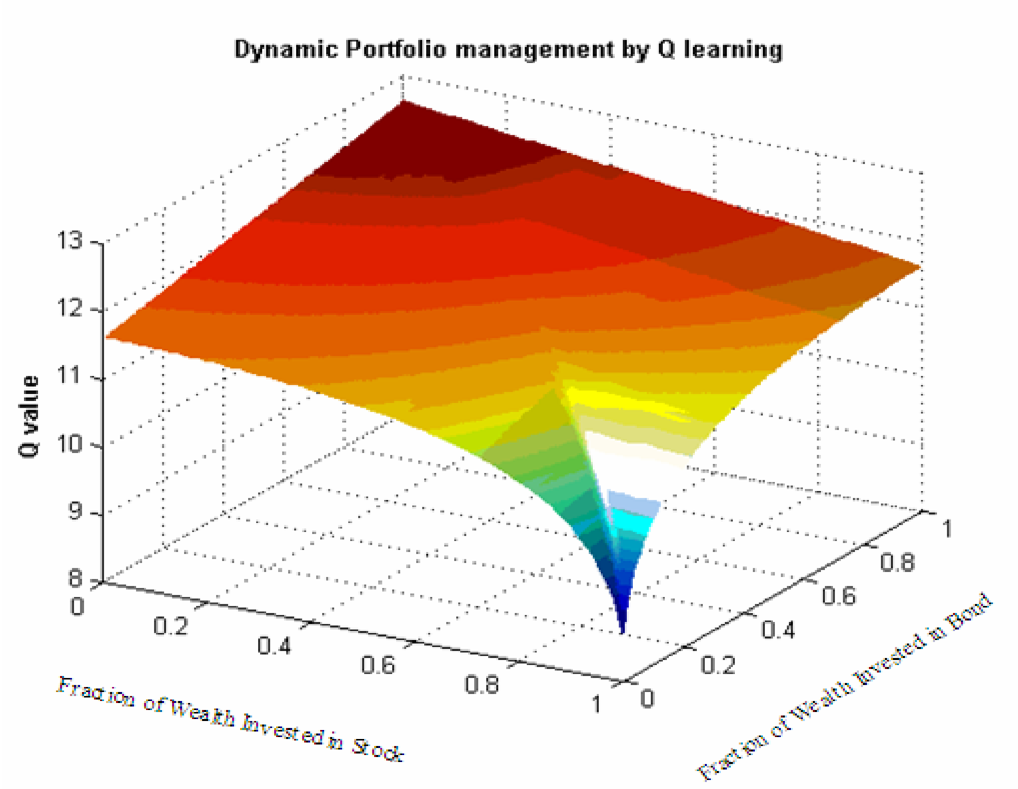


Figure 3.5: Q-value of One Stock and One Bond for One Period

3.5.2. Dynamic Portfolio Management for Two Assets

Within a period of time, the Q-value function will be the expected discounted utility function. Suppose the estimated price of the stock is raised in time periods from t=1 to t=5 and from t=11 to t=15. The optimal actions will be to hold the stock from t=1 to t=5 and from t=11 to t=15, and to hold the bond in the remaining period. It should be noted that the Q-values for this stock vary according to the changes in its price. Therefore, the optimal action based on our model is buying the stock in time 1, selling the stock and buying the bond in time 5, and selling the bond and buying the stock in time 11.

3.5.3. Dynamic Portfolio Management for 47 Assets

In this section, forty-seven assets are considered. These assets are Dow Jones assets in the period from 1992 to 2008. Again, by using the Curvefit Toolbox of MATLAB on 196 points of the data set, we can determine the best fitted distribution and its parameters. The Geometric Brownian Motion with Jump is chosen based on the Curve fitting results. In order to define the best trading policy, we should choose the optimal portfolio in each period. Since we are dealing with thousands of portfolios and it is impossible to illustrate them in one graph, let us consider 250 portfolios with the highest Q-value in each period. These portfolios and their Q-values are shown in Figure 3.6.

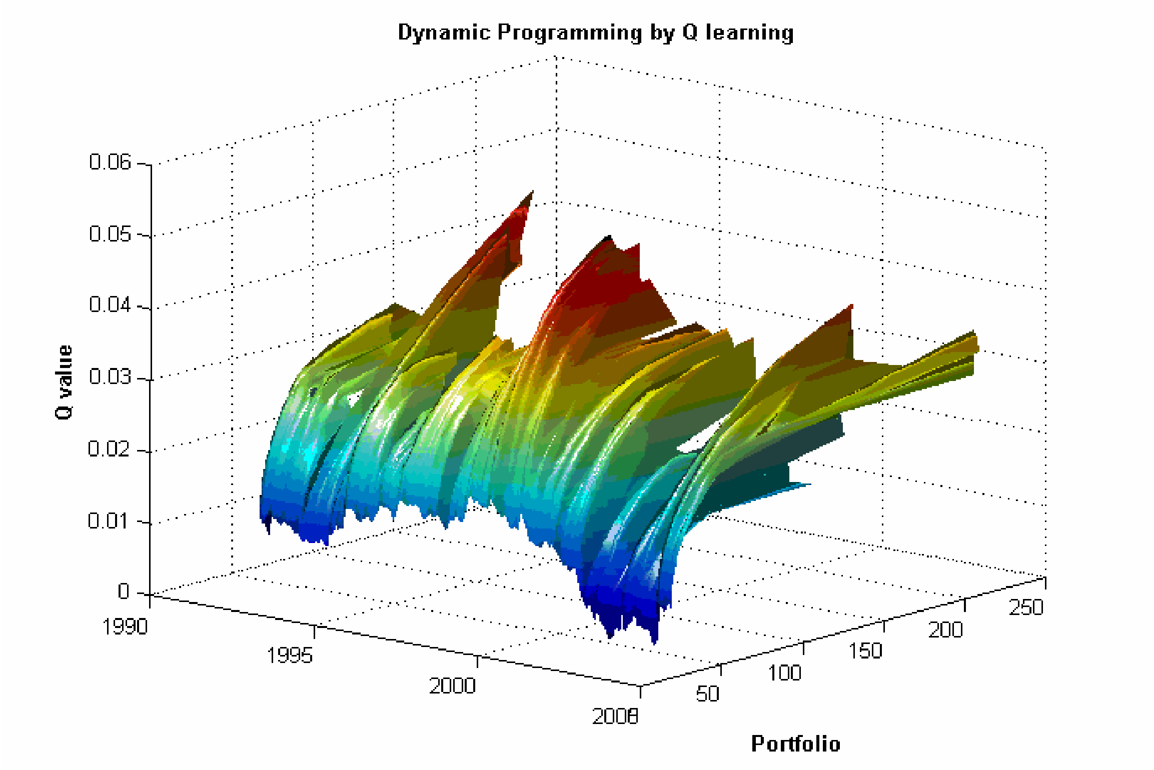
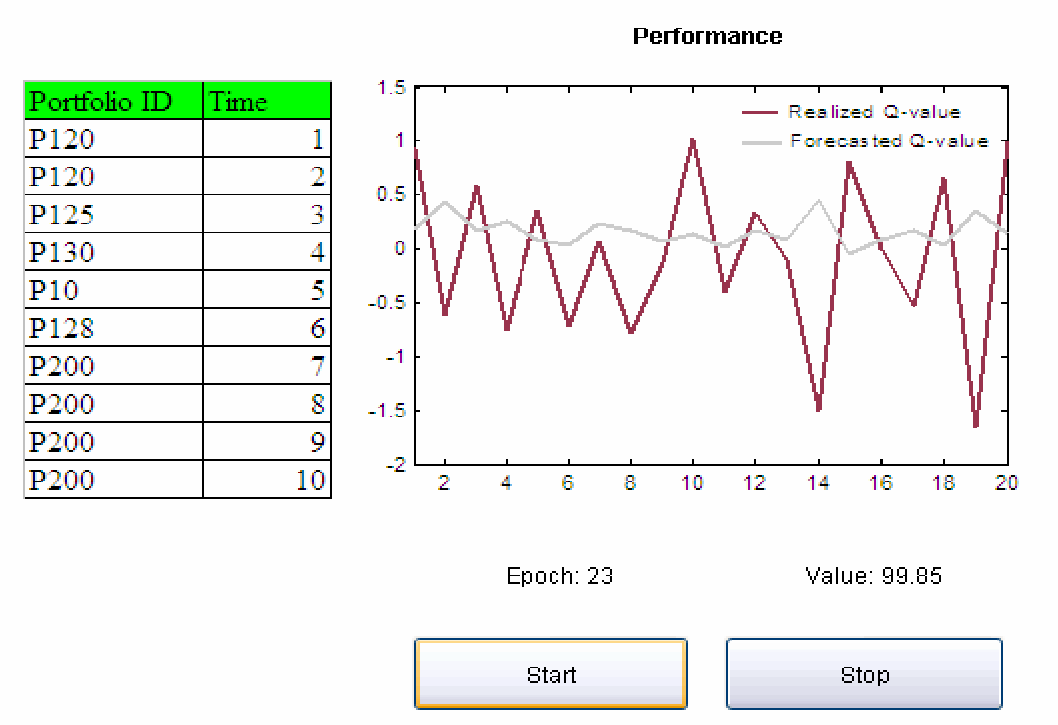


Figure 3.6: Q-value of 250 Optimal Portfolios

By applying Q-learning, we can determine 40 optimal scenarios. Figure 3.7 shows a snapshot of the Q-learning program in MATLAB. In this program, we are trying to optimize the Q-value in 2000 epochs. The 23rd epoch is shown in Figure 3.7. Each epoch is related to one trading policy. On the lefthand side of this figure, you can find the details of this trading policy. The forecasted Q-value and realized Q-value of this trading policy are on the righthand side of this figure. The forecasted Q-value is calculated based on Geometric Brownian Motion with Jump, and the realized Q-value is calculated based on historical data. The difference between the realized and the forecasted Q-values decreases after epoch 1400.



Portfolio ID	Time
P120	1
P120	2
P125	3
P130	4
P10	5
P128	6
P200	7
P200	8
P200	9
P200	10

Epoch: 23

Value: 99.85

Figure 3.7: A Snapshot of Q-learning Program in MATLAB

The realized return of this trading policy is shown on the righthand side of this figure.

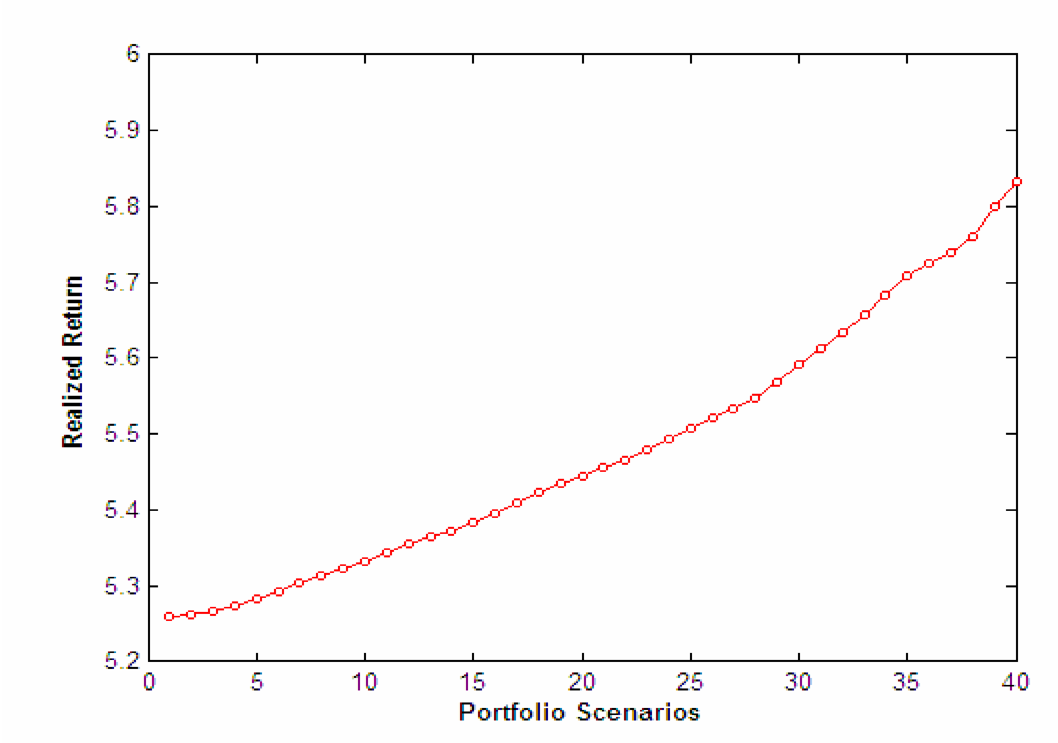


Figure 3.8: Realized Return of Portfolio Scenarios

As you can see in Figure 3.8, each trading policy has a different return.

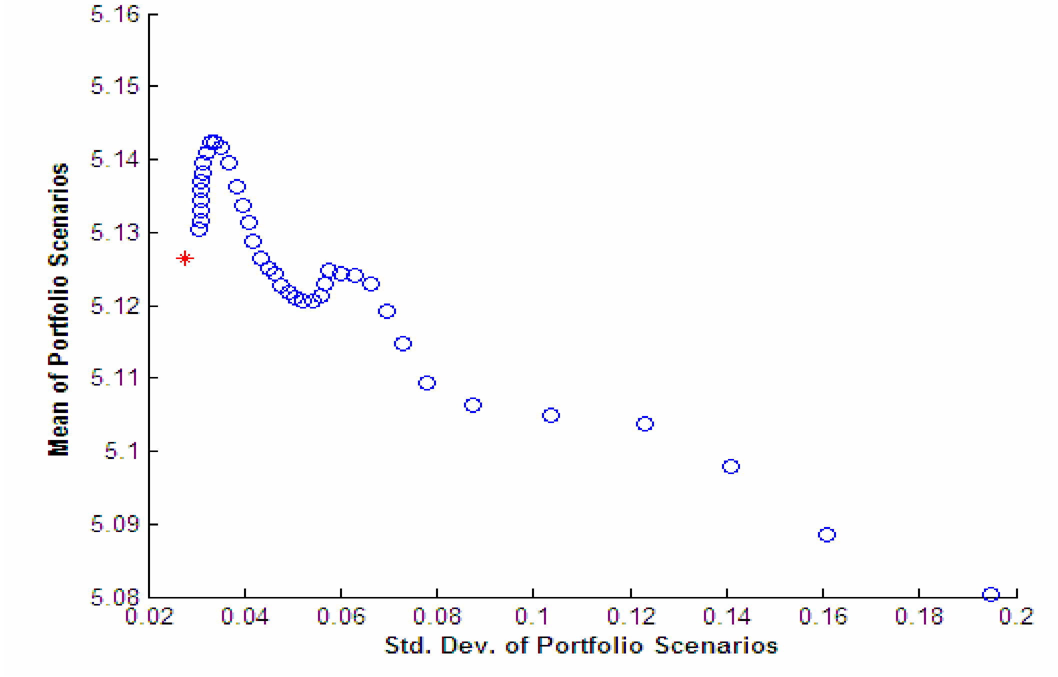


Figure 3.9: Mean-Variance Performance and Comparison with DJIA Benchmark

DJIA can be used as a benchmark, and we can compare our results with this index. Figure 3.9 shows returns and standard deviations of different trading policies and DJIA. DJIA has the lowest standard deviation, but some of the portfolio scenarios have larger returns. Also, the drawdown of DJIA can be compared with drawdowns of portfolio scenarios. This is shown in Figure 3.10.

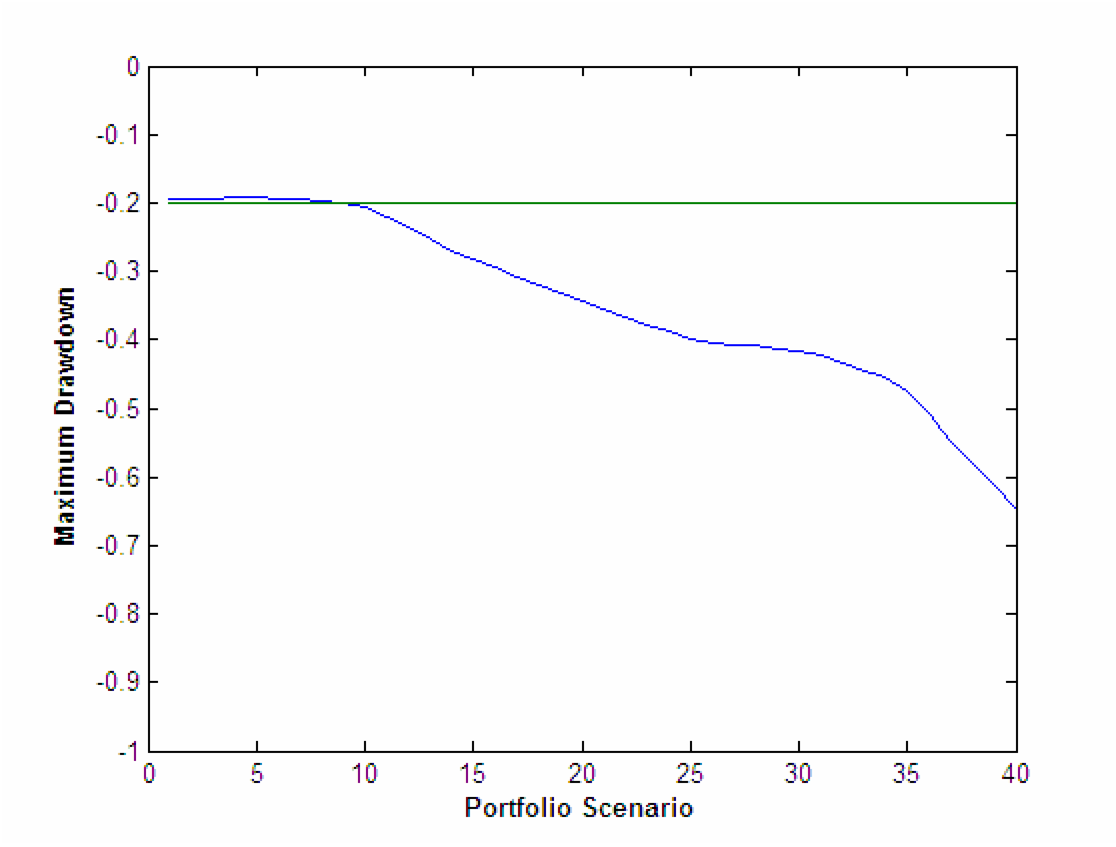


Figure 3.10: Drawdowns of Portfolio Scenarios

As you can see, some of our portfolio scenarios have lower drawdowns than DJIA.

3.6. Conclusion

In this chapter, the problem of Dynamic Portfolio Management was solved by using the Q-learning algorithm. Q-learning was utilized in combination with Geometric

Brownian Motion as an asset price function. We compared our results with the market index. This comparison shows that the trading policy from Q-learning gave us sequential actions with better portfolio returns but not better standard deviations than the market index.

CHAPTER 4

IMPACT OF INCOMPLETE INFORMATION

4.1. Problem Definition

Models of asset pricing generally are made based on the assumption that their parameters are known and observable. However, in general they are not observable and must be estimated. The optimal estimators of these parameters do not yield precise inferences because agents may have noisy and only partial observations. In these situations, it is more realistic to include uncertainty in?/on the value of parameters. This will be referred to as the case of incomplete information.

4.2. Major Contributions

To the best of our knowledge, the sequence of optimal actions (trading) has not been addressed previously in the literature. The literature mainly focuses on a single optimal action. This research compares an investor's sequence of optimal actions under complete and under incomplete information setups and shows how the quality of information influences the agent's sequence of actions. In particular, the following objectives will be pursued here: (1) To obtain optimal estimators for the unobservable drifts using observations of past realized returns, and (2) To examine the impact of uncertainty on the sequence of portfolio actions.

4.3. General Overview

The asset pricing model described here is similar to the one explained in most asset pricing models, with the fundamental difference that agents do not know the exact value of the parameters which determine the state of their systems. There are two parameters in asset pricing models: volatility and drift. It is feasible to obtain a good estimate of the asset's volatility, but it is much harder to estimate drift from noisy observations [Merton 1980]. Therefore, we focus only on the problem of uncertainty about the asset's drift and consider the asset's volatility as a known constant. The unobservable drift is estimated from observations of past realized returns. This estimator is optimized by applying Filtering Theory.

This uncertainty may have an impact on the portfolio choices of agents (investors). Several authors such as: Detemple (1986), Gennotte (1986), Browne (1996), Brennan et al. (2001), Cvitanic et al. (2003), Brandt et al. (2004), Brendle (2005), Lundtofte (2006) and Feldman (2007), have discussed the impact of incomplete information on asset prices and investors' portfolios. Essentially, these authors tried to find the optimal allocation with incomplete information. Because it is not our purpose to give an historical account of the development of optimal portfolio methods with incomplete information, we explore only some directions of research in this area.

The first category of papers in the literature formulates the discrete-time optimal portfolio with incomplete information. For example, Brandt et al. (2004) presented a simulation-based method for solving discrete-time portfolio problems involving a large number of assets with incomplete information on expected rate of return of assets.

Other authors, pioneered by Detemple (1986), extended this idea to continuous-time settings. In most of these papers (e.g., [Brennan 2001]), investors have prior beliefs about the assets' expected returns which are updated according to their observation of prices. Brennan et al. (2001) considered incomplete information about dividend growth rate and measured the effect of learning on portfolio selection. The representative agent in their model is in an early stage of learning, such that any new information may potentially decrease its uncertainty about dividends until a steady state is reached. However, as Gennotte (1986) showed, the accuracy of the estimators does not necessarily increase over time. Since past observations may contain imperfect information, the path of the expected returns will not be perfectly assessed with imperfect information.

In the last category of papers in this area, some authors, like Browne at al. (1996), solved the problem of optimal portfolio allocation in a discrete time and showed its convergence to a continuum time solution.

4.4. Incomplete Information

Drifts and volatilities of rates of return are considered as inputs to asset pricing models. These inputs are generally defined based on historical observations of assets' prices, but the practicality of this assumption is questionable. The reason is twofold: (1) the sample path of an asset price within a time interval is not fully observed; instead, discrete observed statistics of sample paths are used. This means that the sample period *[0, T]* is divided into *m* intervals, and only *m* discrete time price realizations are available to draw the continuous-time model. Therefore, it is not a reasonable estimation of the asset pricing model parameter (drift). (2) Only current and some past assets' prices are observable, and the expected rate of return, μ_s , is unknown. This parameter must be induced from current and past asset prices. The uncertainty in the asset price is generated by Brownian motion, B_t. Since B_t is not observable, and drifts (rates of return) can not be retrieved. Thus, drifts are not fixed and follow a stochastic process.

4.5. Information Structure

All uncertainty is defined over a probability space (Ω, F, P) with a fixed terminal time *T>0*, where Ω denotes a complete description of the environment. Agents are endowed with a common probability measure *P*, and *F* denotes the augmented filtration generated by the Brownian Motion *B*.

Some information, denoted by the symbol $F_{t,}$ is utilized to forecast the value of unobservable parameters. The information utilized could be, and in general is, different from one time to another. If we assume that the agent never forgets past data, the information sets will increase over time:

$$F_0 \subseteq F_1 \subseteq \ldots \subseteq F_T \,, \tag{4.1}$$

where F_t is generated by the observations of the value of the asset up to time *t*, and is augmented.

4.6. Assumptions

It is assumed that the capital market is perfectly competitive. In general, the number of buyer agents and seller agents is sufficiently large. Under this assumption, the agent can buy or sell assets as much as it wants. All agents are small enough relative to the market so that no individual agent can influence an asset's price. This assumption is relaxed in Chapter 5 where the impact of the agent's action is discussed.

Agents are characterized by their risk aversion factors and their initial wealth. Here, an agent is concerned with constant absolute risk aversion in a continuum setting. The investor can observe asset prices, but not instantaneous drifts, which follow a mean-reverting process. By considering these assumptions, the optimal portfolio under partial observation (incomplete information) is determined. The agent's objective function is to maximize its expected utility of wealth at time *T*.

4.7. The Methodology

The main goal of this methodology is to select a sequence of optimal portfolio actions with incomplete information. Before we delve into the details, it will be useful to see an overview of this methodology. Five main steps of this methodology are listed below:

1. Determining the model for asset pricing
2. Determining the model for drift of asset price
3. Applying filtering theory and minimizing noises
4. Learning process
5. Optimizing portfolio

The agent's decision problem is divided into two separate problems [Gennotte 1986]. In Steps 1 to 4, the agent tries to find a good estimate of the asset price. The agent seeks to filter or extract information on unobservable variables from its past observations. In Step 5, it uses its current estimate of asset prices to choose a sequence of optimal portfolio actions. The process of selecting optimal actions with incomplete information is shown in Figure 4.1.

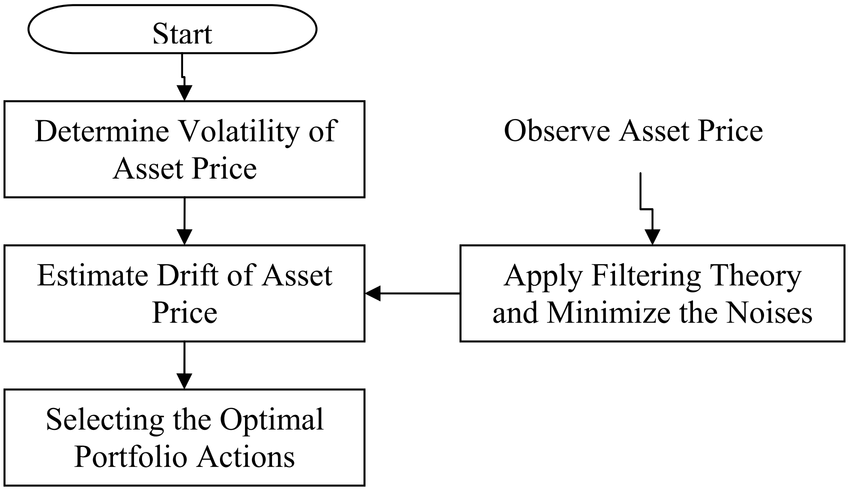

Start
Determine Volatility of Asset Price
Observe Asset Price
Estimate Drift of Asset Price
Apply Filtering Theory and Minimize the Noises
Selecting the Optimal Portfolio Actions

Figure 4.1 : The Process of Selecting Optimal Actions with Incomplete Information

Here, agents who make decisions in continuums time over a finite time horizon $[0,T]$ are considered. At time zero, the agent determines from available data the volatility and drift of asset prices. Based on these parameters, it forecasts the asset prices, and selects the optimal actions which maximize its expected utility of wealth. Whenever the agent observes the new asset prices, it updates its estimation of drift and optimal actions. Since some noises exist in the asset and drift process, the agent filters information to minimize the effects of these noises.

4.7.1. Determining the Model for Asset Pricing

The environment has *n* risky assets and one risk-free asset available for investment. The price *s(t)* of asset at time *t* is interpreted as accumulated its dividend and its price at that time. The price is observable and satisfies the following stochastic differential equation:

$$\frac{ds^i(t)}{s^i(t)} = \mu^i(t)\,dt + \sigma_s^i\,dB_t^i, \tag{4.2}$$

where $\mu^i(t)$ is the expected rate of return of asset i, and σ_s^i is its volatility. σ_s^i is considered equal to zero for the riskless asset. dB_t^i can be interpreted as the infinitesimal change in a Brownian motion over the next instant of time.

At each point in time, the asset price is characterized by the path of B_t^i, and the value of $\mu^i(t)$ prior to that time point. The uncertainty in the asset price is generated by this Brownian motion. B_t^i is defined over a complete probability space (Ω, F, P) with a non-decreasing family of sub-σ-algebras $\{F_t, 0 \le t \le T\}$. The possible paths of B_t^i and $\mu^i(t)$ constitute the set of possible events.

4.7.2. Determining the Model for Drift of Asset Prices

Clearly, the assumption of constant expected drift is inappropriate and needs to be replaced. The estimation of $\mu^i(t)$ is not a particularly easy problem, and estimated errors are likely to be substantial. Like Brendle (2005), Barberis (2000) and Xia (2001), we assume that drifts are modeled by a mean-reverting process, and the $\mu^i(t)$ follows a stochastic differential equation of the form:

$$d\mu^i(t) = \lambda_i(\overline{\mu}^i - \mu^i(t))\, dt + \sigma_\mu^i \, dZ_t^i \qquad \forall i, \quad 1 \le i \le n, \tag{4.3}$$

where $\overline{\mu}^i$ is the mean reversion level, $\lambda_i > 0$ is the reversion rate, and σ_μ^i is the drift volatility of asset *i*. dZ_t^i represents the infinitesimal change in a Brownian motion. The value of λ_i defines how quickly the drift returns to the equilibrium. Parameters $\overline{\mu}^i$, λ_i and σ_μ^i are known and constants.

This equation differs from simple diffusion by the addition of an equilibrium level and a restoring force that pulls subsequent values toward that equilibrium. The first term of this formula brings $\mu^i(t)$ back to some equilibrium level ($\overline{\mu}^i$). Every time the stochastic term ($\sigma^i_\mu \; dZ^i_t$) gives $\mu^i(t)$ a push away from the equilibrium, the deterministic term ($\lambda_i(\overline{\mu}^i - \mu^i(t))\,dt$) will act in such a way that $\mu^i(t)$ will start heading back to the equilibrium. When the deterministic term is negative ($\overline{\mu}^i < \mu^i(t)$), $\mu^i(t)$ is pulled down toward the equilibrium level, $\overline{\mu}^i$. When the deterministic term is positive ($\overline{\mu}^i > \mu^i(t)$), $\mu^i(t)$ is pulled up to the equilibrium value $\overline{\mu}^i$. The end result is that the drift tends to oscillate around the equilibrium. The greater the value of λ_i is, the faster the value of $\mu^i(t)$ returns to the equilibrium level.

4.7.3. Applying Filtering Theory and Minimizing Noises

The agent has to make its investment decisions on the information available at the time of decision. The quality of the portfolio selection obviously depends on the information that can be used by the investor. A better-informed agent (investor) can make a better investment decision.

Here, the agent has partial information in its hands. The drift is not directly observable, but the asset price is. The agent forecasts the asset price (Equation (4.2)) based on the estimated drift (Equation (4.3)). The agent seeks to extract information on future expected returns from its observation of past returns. Equations (4.2) and (4.3) are rearranged as:

$$\mu^i(t) = \lambda_i \overline{\mu}^i dt + \sigma^i_\mu dZ^i_t + \mu^i(t - dt)(1 - \lambda_i dt) \tag{4.4}$$

and

$$\ln s^i(t) = \mu^i(t)\, dt + \sigma_s^i\, dB_t^i + \ln s^i(t - dt)\,. \qquad (4.5)$$

Based on Equations (4.4) and (4.5), the asset price and drift have their own noises, dB_t^i and dZ_t^i. These two noises are independent of each other. The process of predicting the asset price is represented pictorially in Figure 4.2.

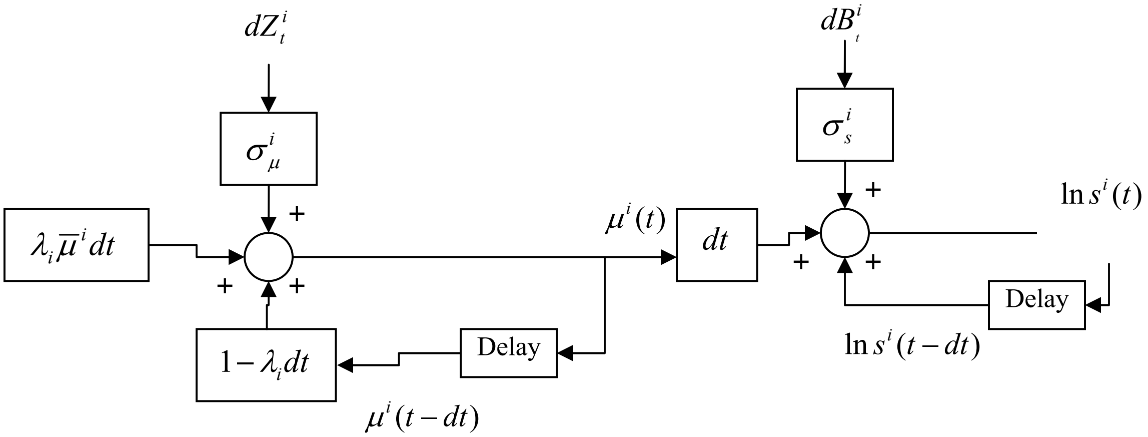


Figure 4.2 : The Process of Predicting Asset Prices

This approach has two glaring weaknesses. First, there is no feedback from observation *(s(t))* to estimate the drift process. Second, there are two noises in the computation, which could cause a large error.

The problem of estimating the state of a dynamic system from noisy observations is an important topic in engineering. Filtering can derive the optimal estimators of the expectation of unobservable parameters conditional on past observations. In simple terms, filtering is based on a recursive algorithm that would find an optimal solution given current measurement data and past system data from the previous iteration (the recursion) [Maybeck 1979].

By applying the filtering theory (Theorem 12.1 of Liptser and Shiryayev [Liptser 2001]), the effect of noises is minimized and estimation of drifts is updated by new observations. This allows us to predict future returns, conditional only on observed noisy data, and not on fixed parameter values. The conceptual overview of Filtering theory is shown in Figure 4.3.

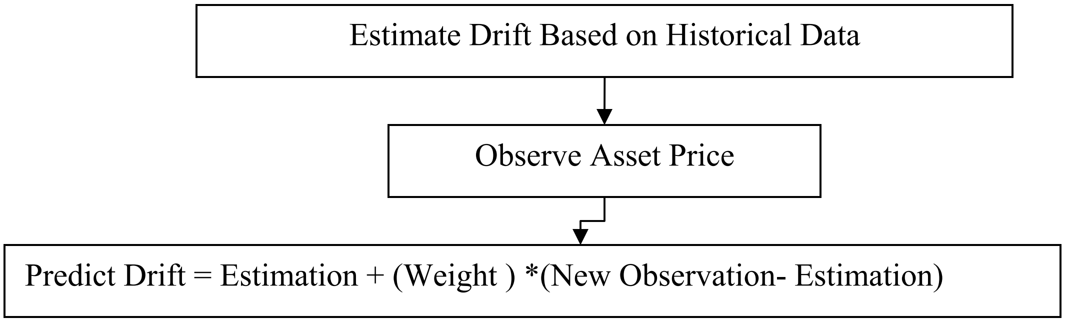

Estimate Drift Based on Historical Data
Observe Asset Price
Predict Drift = Estimation + (Weight) *(New Observation- Estimation)

Figure 4.3: Conceptual Overview of Filtering Theory

At time zero, each agent has a belief about drift. At each time point, the agent continuously updates its beliefs and uses all observable information to learn about the unobservable drift.

4.7.4. Learning Process

It is assumed that the asset price drift is normally distributed and its initial distribution, $\mu(0)$, has mean m_0 and variance v_0. During each time interval $[t, t+dt]$, the agent observes $\frac{ds(t)}{s(t)}$, which is correlated with the drift, $\mu(t)$. The agent does not directly observe $\mu(t)$. Let $m(t) = \mathrm{E}\left[\mu(t)\middle|F_t\right]$ and let $v(t) = E\left[(\mu(t) - m(t))^2\middle|F_t\right]$ denote the expectation and variance of drift at time *t* conditional on observing data up to time *t* (the augmented filtration, F_t). Then, upon observing the asset price, the agent can revise its belief about the true value of the drift. Based on Theorem 12.1 of Lipster et al. (2001), the instantaneous changes in the expected estimated drift, $dm(t)$, and the instantaneous changes in the variance of estimated drift, $dv(t)$, are given by Equations (4.6) and (4.7):

$$dm(t) = \lambda \ (\bar{\mu} - m(t))dt + \frac{\sigma_s \sigma_\mu + v(t)}{\sigma_s^2}(\frac{ds(t)}{s(t)} - m(t)dt) \tag{4.6}$$

and

$$\frac{dv(t)}{dt} = -2\lambda v(t) + \sigma_\mu^2 - (\sigma_\mu + \frac{v(t)}{\sigma_s})^2 . \tag{4.7}$$

The updating process is a recursive procedure as new observations become available. From Equation (4.6), the change in the assessment of drift, $dm(t)$, is equal to the estimated expected drift at time t plus a correction term. This term includes the weighting value and instantaneous change occasioned by the observation of s over the period $[t, t + dt]$. This means that the agent updates $m(t)$ by the surprise component ($\frac{ds(t)}{s(t)} - m(t)dt$) weighted by its relative uncertainty ($\frac{\sigma_s \sigma_\mu + v(t)}{\sigma_s^2}$). The agent raises its assessment of the drift whenever the rate of return is above its current assessment. The weight in Equation (4.6), $\frac{\sigma_s \sigma_\mu + v(t)}{\sigma_s^2}$, determines how much of the new information is incorporated into the updating of $m(t)$. When the quality of data is poor (a high value of σ_s^2), little information can be extracted, and therefore, $m(t)$ is not changed much. When the agent is less confident of its current estimate (a higher $v(t)$) more information can be obtained. In this case, the agent puts more weight on the new information, and revises its beliefs more quickly.

The first two terms of Equation (4.7) correspond to the unobservable variation of $\mu(t)$ over the period $[t, t+dt]$. The last term denotes the reduction in variance when additional information becomes available. It says that the better the quality of data (a low value of σ_s), the more rapidly the agent learns about the current value of the drift. The solution of this equation is:

$$v(t) = \frac{2\sigma_s v_0 (\lambda\sigma_s + \sigma_\mu)}{e^{\frac{2t(\lambda\sigma_s + \sigma_\mu)}{\sigma_s}} \left(2\lambda\sigma_s^2 + 2\sigma\mu\sigma_s + v_0\right) - v_0}. \qquad (4.8)$$

This equation says the uncertainty about drift decreases as time progresses. The limit of *v(t)* as *t* tends to infinity is zero. If we perfectly know m_0 (this means that $v_0 = 0$), the uncertainty of *m(t)* will totally disappear (*v(t)=0*).

The process B_t' is defined as a Brownian motion as follows:

$$dB_t' = \frac{1}{\sigma_s}\left(\frac{ds(t)}{s(t)} - m(t)dt\right). \qquad (4.9)$$

B_t' is the normalized deviation of the rate of return from its estimated mean. B_t' can be inferred from observable processes; therefore, B_t' is observable. Rearranged Equations (4.9) and (4.6) follow:

$$\frac{ds(t)}{s(t)} = m(t)dt + \sigma_s dB_t', \qquad (4.10)$$

and

$$dm(t) = \lambda(\bar{\mu} - m(t))dt + (\sigma_\mu + \frac{v(t)}{\sigma_s})dB_t' = \lambda(\bar{\mu} - m(t))dt + \sigma_m dB_t'. \quad (4.11)$$

This means that the incomplete information is equivalent to complete information with the consideration of expected drift (see [Gennotte 1986]). The solution of Equation (4.11) is:

$$m(t)=e^{-\lambda t}\left(m_0+\lambda\overline{\mu}(e^{\lambda t}-1)+\int_0^t e^{\lambda k}(\sigma_\mu+\frac{v(k)}{\sigma_s})dB'_k\right). \tag{4.12}$$

Based on this equation, if μ is considered as an unknown constant, *m(t)* still fluctuates, but its limit, as *t* tends toward infinity, is $\lambda\overline{\mu}$.

4.7.5. Optimizing the Portfolio

To choose a sequence of optimal portfolio actions, the agent forms its assessment of the drift from the observed asset prices. The objective function of the agent is to maximize the expected utility of wealth at the end of horizon *T.* The agent's utility function at time *t* depends on its current wealth and its risk aversion factor. This utility is assumed to be in the form of:

$$u(w_t)=\frac{w_t^{1-\beta}}{1-\beta}\quad for\ \beta\neq 0\quad . \tag{4.13}$$

The parameter β represents the Arrow-Pratt coefficient of the relative risk aversion factor, which is assumed to be constant for this choice of utility function. w_t is the wealth accumulated at the end of period *t* and is given by:

$$w_t=x(t)s(t)-c(dx(t)), \tag{4.14}$$

where $x(t)$ specifies the number of assets belonging to the agent at time *t*, and $c(dx(t))$ is the transaction cost of action $dx(t)$. For simplicity, $c(dx(t))$ is considered equal to zero.

The agent has to make its investment decision on its actions, *dx(t)*, which is equal to the instantaneous change in $x(t)$:

$$dx(t) = a(t)dt\,. \tag{4.15}$$

Since the environment is perfectly competitive and the instantaneous returns do not depend on the level of investment, the agent can choose its optimal portfolio action based on the estimated expected drifts of asset prices. The impact of uncertainty on this action is discussed in the next section.

4.8. Examining, the Impact of Uncertainty on Agent's Action

The problem of optimal control is to find the sequence of *dx(t) (t=0,..T)* which maximizes the objective function. Following Liu (2007) and Brennan (1998), the stochastic optimal control approach is applied to find the optimal sequence of the agent's actions. The optimal portfolio actions ($a^*(t)$), which ignore parameter uncertainty, are compared with the portfolio actions that take this uncertainty into account. This framework allows us to understand how parameter uncertainty affects portfolio choices. The instantaneous change of wealth is as follows:

$$\begin{aligned} dw_t &= s(t)dx(t) + x(t)ds(t) \\ &= w_t\left\{\frac{dx(t)}{x(t)} + (m(t)dt + \sigma_s dB_t')\right\}. \end{aligned} \tag{4.16}$$

The agent's indirect utility function at time *t*, Iu, depends on wealth, estimated drift, variance of estimated drift and time.

Let $Iu(w,m,v,t)$ be the expected value of utility of wealth, $u(w_T)$, at time *t<T* ($Iu(w,m,v,t) = E\left[u(w_T)\right]$). Based on the stochastic control approach, the optimal policy can be found by $E[dIu] = 0$. dIu is formulated by considering Ito's lemma as:

$$\begin{aligned} dIu = {} & \partial Iu_{w_t} dw_t + \partial Iu_t dt + \partial Iu_{m(t)} dm(t) + \partial Iu_{v(t)} dv(t) \\ & + \frac{1}{2}\partial^2 Iu_{w_t} dw_t^2 + \frac{1}{2}\partial^2 Iu_{m(t)} dm(t)^2 + \frac{1}{2}\partial^2 Iu_{v(t)} dv(t)^2 \qquad , \\ & + \partial Iu_{w_t\, m(t)} dw_t dm(t) + \partial Iu_{w_t\, v(t)} dw_t dv(t) + \partial Iu_{v(t)m(t)} dv(t) dm(t) \end{aligned} \tag{4.17}$$

where ∂Iu_{w_t}, ∂Iu_t, $\partial Iu_{v(t)}$ and $\partial Iu_{m(t)}$ denote partial derivatives with respect to w_t , *t*, *v(t)* and *m(t)*, respectively. Similar notation is used for higher derivatives and mixed derivatives. Since the Brownian Motion includes $e(t)\sqrt{dt}$ (*e(t)* is a normal random variable with zero mean and unit variance), the terms with dt^2 and dB_t' are discarded and $dB_t'^2$ is considered as *dt*. Equation (4.17) is simplified as:

$$\begin{aligned} dIu = {} & \partial Iu_{w_t} w_t \left(\frac{a(t)}{x(t)} + m(t) \right) + \partial Iu_t + \partial Iu_{m(t)} \lambda(\bar{\mu} - m(t)) \\ & + \partial Iu_{v(t)} dv(t) + \frac{1}{2}\partial^2 Iu_{v(t)} dv^2(t) + \frac{1}{2}\partial^2 Iu_{m(t)} \left(\lambda^2 (\bar{\mu} - m(t))^2 dt + \sigma_m^2 \right) \\ & + \frac{1}{2}\partial^2 Iu_{w_t} w_t^2 \left(\frac{a^2(t)}{x^2(t)} dt + m^2(t) dt + 2\frac{a(t)}{x(t)} m(t) dt + \sigma_s^2 \right) \qquad . \\ & + \partial Iu_{w_t\, m(t)} w_t \left[\sigma_s \sigma_m + (\frac{a(t)}{x(t)} + m(t)) \lambda(\bar{\mu} - m(t)) dt \right] \\ & + \partial Iu_{v(t)m(t)} \lambda(\bar{\mu} - m(t)) dv(t) + \partial Iu_{w_t\, v(t)} w_t \left(\frac{a(t)}{x(t)} + m(t) \right) dv(t) \end{aligned} \tag{4.18}$$

The investor's indirect utility function $Iu(w,m,v,t)$ is given by (Theorem 3 of Brennan [Brennan 2001b]):

$$Iu(w,m,v,t) = \frac{w_t^{1-\beta}}{1-\beta} f(m,v,t) . \tag{4.19}$$

The principle of optimality leads to:

$$\underset{a(t)}{Max}\left\{\begin{array}{l} w_t^{1-\beta}\left(\dfrac{a(t)}{x(t)}+m(t)\right)f+\partial f_{m(t)}\lambda(\overline{\mu}-m(t))+\partial f_{v(t)}dv(t)+\partial f_t \\ +\dfrac{1}{2}\partial^2 f_{v(t)}dv^2(t)+\dfrac{1}{2}(-\beta)w_t^{1-\beta}\left(\begin{array}{l}\dfrac{a^2(t)}{x^2(t)}dt+m^2(t)dt \\ +2\dfrac{a(t)}{x(t)}m(t)dt+\sigma_s^2\end{array}\right)f \\ +\dfrac{1}{2}\partial^2 f_{m(t)}\left(\lambda^2(\overline{\mu}-m(t))^2dt+\sigma_m^2\right) \\ +\partial f_{v(t)m(t)}\lambda(\overline{\mu}-m(t))dv(t)+\partial f_{v(t)}w_t^{1-\beta}\left(\dfrac{a(t)}{x(t)}+m(t)\right)dv(t) \\ +\partial f_{m(t)}w_t^{1-\beta}\left[\sigma_s\sigma_m+(\dfrac{a(t)}{x(t)}+m(t))\lambda(\overline{\mu}-m(t))dt\right] \end{array}\right\}=0\,, \quad (4.20)$$

with the boundary condition: $Iu(w,m,v,T)=\dfrac{w_T^{1-\beta}}{1-\beta}$.

$a^*(t)$ is derived by the first order condition of Equation (4.20), and it is given by Equation (4.21). This defines the sequences of optimal portfolio actions. The sign of a^* shows whether the optimal action is to sell or to buy.

$$a^*(t)dt=x(t)\left\{\begin{array}{l}\dfrac{1}{\beta}+\dfrac{\partial f_{m(t)}}{f*\beta}[\lambda(\overline{\mu}-m(t))]dt \\ +\dfrac{\partial f_{v(t)}}{f*\beta}\left(-2\lambda v(t)+\sigma_\mu^2-(\sigma_\mu+\dfrac{v(t)}{\sigma_s})^2\right)dt-m(t)dt\end{array}\right\}. \quad (4.21)$$

Given the agent's information set, its optimal portfolio action at time *t* is characterized by its current asset level, its current assessment of drift, its risk aversion factor, the volatility of asset price and the drift of asset price. Note that in a complete information scenario, $\partial f_{v(t)}$ is zero, and the optimal agent's action is:

$$a^*(t)dt=x(t)\left\{\dfrac{1}{\beta}+\dfrac{\partial f_{m(t)}}{f*\beta}[\lambda(\overline{\mu}-m(t))]dt-m(t)dt\right\}. \quad (4.22)$$

Consider an agent with relative risk aversion factor greater than zero. Whether the agent's optimal trading volume under incomplete information is more or less than that under complete information depends on its estimate of drift, *m(t).* Based on Equation (4.21), when the agent assessment of drift is lower than the mean of rate of return, the action of complete information may be higher than the action of incomplete information. Otherwise, it is always less than the trading volume of incomplete information.

4.9. Summary and Conclusions

Previous chapters discuss cases where the uncertainty of parameters is ignored. This means the agents allocate their portfolios taking the parameters as fixed at their computations. Here, these parameters are considered as stochastic variables. This is a quite realistic assumption, since the only available information for agents at each time is the prices of the assets up to that time, and the underlying Brownian Motion and the drift process of the asset prices are not directly observable. The optimal estimators for the unobservable rate of returns were obtained by applying Filtering theory.

The main contribution of this chapter is to find the optimal sequence of actions of a dynamic portfolio. These actions were defined based on estimated drifts. To establish the results, the structure of the optimal sequence of actions with incomplete information was formally defined in mathematical terms. Then, we compared the optimal actions of an investor who takes into account the error of predicted drift of asset prices, with the optimal actions of an investor who is blind to this error. In other words, the effect of the uncertainty of parameters was defined by comparing the solution of fixed parameters with the solution of uncertain parameters. This comparison showed that the uncertainty of parameters usually forces risk-averse agents to choose a higher trading volume. However, these trading volumes may be lower in cases where the agent’s assessment of drift is lower than the mean of drift.

CHAPTER 5

IMPACT OF AGENT ACTIONS

5.1. Problem Definition

Different sources of empirical evidence indicate that large investors have striking impacts on prices through their trading strategies. These agents often consider the question of how to choose their trading strategies when taking into account their price impact. They choose trading strategies in order to maximize their objective functions. An agent's objective function depends on its own action as well as on other agents' actions.

5.2. Major Contributions

In previous chapters, all agents were considered as price taking investors. However, observations of today's asset markets reveal the ever-increasing importance of non-price-taking investors. A non-price-taking investor influences market prices with its large order flow. Therefore, a non-price-taking agent is often called a large trader or a large agent. This agent is particularly important in small environments because of its significant effect on prices. Here, a model based on game theory is developed to figure out the optimal actions of non-price-taking agents with and without debt constraint.

5.3. General Overview

Numerous papers have been written about price impact, which was considered mainly in the market with private information and imperfect competition [Pritsker 2005]. The number of works is too great to mention all the relevant papers. Glosten et al. (1985), Basak (1995), Demarzo et al., Urosevic (2002), Foster et al. (1996) and Holden et al. (1996) studied price impacts based on private or asymmetric information. Lindenberg

(1979) and Kihlstrom (2001) examined the impacts of agent actions on the imperfect competition market. Lindenberg (1979) considered the static portfolio optimization problem with many large investors. Kihlstrom (2001) considered a dynamic portfolio with only one large investor. In this chapter, the price impact of agents' actions is modeled on Game Theory. This model attempts to figure out the Nash Equilibrium of games when each agent acts to optimize its own expected utility. The game model has the following form: first the players (agents) simultaneously choose actions (trading volumes); then they receive payoffs (rewards) that depend on the combination of the actions just chosen. In this game, each agent's payoff depends on the others' actions. Also, an agent considers both the fact that its action has an effect on its payoff at that period, and the trading opportunities available in the future.

This chapter is organized as follows: In Sections 5.4 through 5.6, game theory, and the market and asset pricing model are introduced. Section 5.7 discusses a game in a market without any constraint. Then, in that framework, Sections 5.8 and 5.9 present a game in the market with and without debt constraint. Section 5.10 concludes.

5.4. Game Theory

A large number of financial assets are held and managed by non-price-taking investors, whose order flows may change asset prices. Game Theory is applied to consider the impact of actions by non-price-taking agents. "Game theory is a modeling approach which drops competition's assumption that individuals are price-takers and instead requires them to behave strategically, taking into account that their actions will alter the behavior of the rest of the market" [Rasmusen 1992]. This theory first appeared with the publication of Von Neumann and Morgenstern's book, *The Theory of Games and Economic Behavior* [Von Neumann 1944].

Each agent has to consider some constraints imposed by its information. The limitations on information are important in understanding an agent's behavior, because such limitations induce the agent to alter its behavior. There are different kinds of games based on available information.

Games are classified according to their level of limitation of information. If the historical moves of a game are not accessible to all players, the game is said to be an *imperfect information game*; otherwise, it is a *perfect information game*. The *incomplete information game* is interpreted as a game where a player lacks full information about utility functions and the available strategies of other players. If this information is available to a player at the time of a move, the game is said to be a *complete information game*.

Games can also be classified based on the level of collaboration between agents. A game may be played by agentscooperatively or noncooperatively. In a cooperative game, agents jointly maximize their expected utilities, and in a non-cooperative game, each agent selfishly maximizes its own expected utility. Cooperative games concentrate on the efficient payoffs of agents under the assumption that the agents are allowed to communicate and make binding commitments, whereas non-cooperative games seek efficient payoffs when the players are not able to communicate. In the real world, non-cooperative games are more common than cooperative ones. In this research, each agent plays a non-cooperative, perfect, and complete-information game with other large agents.

5.5. Market

The market is populated by large agents and “noise” traders. The noise trader is an asset trader who makes decisions to buy, sell or hold without fundamental analysis. We consider only one noise trader in our model. This assumption may capture the essence of multi-noise traders if we consider this noise trader as an aggregate of all noise traders.

Market makers are not considered in this model; thus, the trading is done according to the supply and demand of assets. Also, there is no private information in this model, and all news is public knowledge. The other assumptions of this market are as follows:

- The time horizon is infinite, and trading in the market can happen in each time epoch.
- There are always buyers and sellers for the assets, in the sense that almost any amount of an asset can be bought or sold immediately. There is no delay in selling or buying orders.
- The random quantity traded by each noise trader is distributed independently from present or past quantities traded by other agents.
- The quantity traded by a large agent at time *t* is independent of the quantities traded by this agent at other times.
- When an agent chooses the quantity to hold or trade, it observes past prices of assets.
- Large agents have the ability to affect the price of assets by varying their trading volumes, while noise traders do not have this ability.

5.6. Asset Price Model

Suppose that there exist *n* risky assets traded in continuous time. In Chapter 2, the risky asset price was modeled by a stochastic differential equation, ($s(t) = s(0)\, e^{(\mu - \frac{\sigma^2}{2})t + \sigma B_t}$). In this model, the price impact of an agent's order flow and its inventory of assets were not considered. Let us now consider the price of asset *i* as $\Pr^i(s^i(t), X^i(t), A^i(t))$, where $X^i(t)$ is the inventory variable which measures the aggregate amount of asset *i* that the large agents hold at time *t*. $X^i(t)$ is defined by $\sum_{j=1}^{k} x_j^i(t)$, where $x_j^i(t)$ denotes the number of holding asset *i* by agent *j* at time *t*. The holding asset at the start of period *t* should be right continuous ($x_j^i(t-) = \lim_{\tau \uparrow t} x_j^i(\tau)$).

The aggregate supply $Z^i > 0$ of risky asset *i* at any time *t* is divided between the large agents' holdings, $X^i(t)$, and the noise trader's holding, $O^i(t)$, such that $Z^i = X^i(t) + O^i(t)$. As $X^i(t)$ increases, the supply available to the noise trader decreases and the value of $\Pr^i(s^i(t), X^i(t), A^i(t))$. Where $A^i(t)$ is the total traded volume of asset *i* at time *t*. We can consider $A^i(t)$ as $\sum_{j=1}^{k} a_j^i(t)$, where $a_j^i(t) = dx_j^i(t)$ denotes the trading volume of asset *i* by agent *j* at time *t*. $a(t)$ with positive value is the purchasing amount of the asset and $a(t)$ with negative value is the selling amount of the asset. If many agents want to sell their holding assets, the value of $A^i(t)$ will increase. This leads to a decrease in value of $\Pr^i(s^i(t), X^i(t), A^i(t))$. In other words, a high value of $A^i(t)$ leads to a lower value of $\Pr^i(s^i(t), X^i(t), A^i(t))$.

The asset price models with linear function of total action and inventory have been widely used in both the theoretical and the empirical literature. This widespread application is not due to the opinion that those linear functions are particularly realistic, but instead, is due to the simplicity of analysis and the possibility of having a well behaved solution. Here, a linear function of action and inventory is considered as follows:

$$\mathrm{Pr} = s(t) - \gamma(Z - X(t)) + \lambda A(t)\,, \tag{5.1}$$

where γ and λ are considered as parameters. This asset price model is justified by an empirical work by Cheng et al. (1997).

5.7. The Game in the Market

The behavior of the market can be viewed as a two-sequential-step game. In Step One, all agents simultaneously choose their actions. Agents' possible actions are buying, holding and selling an asset. When making this choice, an agent does not observe future prices, or current and future quantities traded by other agents. In Step Two, the aggregate quantities are traded, and then the market has price fluctuations as a consequence of these aggregate order flows. The timing of the game is shown in Figure 5.1.

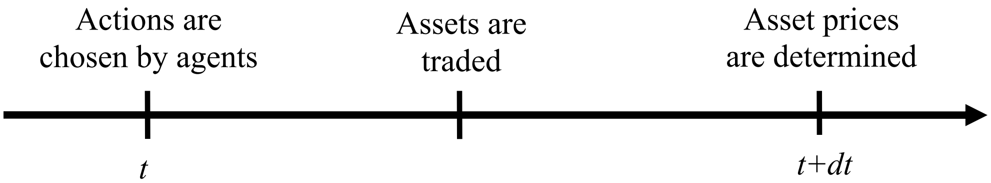


Figure 5.1: The Timeline of the Game

Each agent in the market is trying to optimize its own reward function. This reward function directly depends on assets price, which in turn depends on all agents' actions. Therefore, agents choose actions which optimize their reward functions.

When a large agent decides to buy a large amount of one asset, the demand will increase; hence, a higher price will emerge. If it decides to sell the asset, this time the supply will increase, causing the asset price to fall. If the agent decides to hold the asset, since neither the demand nor the supply changes, the asset price will remain unaffected. The large agents form a tight Cournot oligopoly over order flow in the market. This game has a Nash Equilibrium.

5.7.1. Definition: Nash Equilibrium

A Nash Equilibrium is a profile of strategies such that each player's strategy is an optimal response to the other players' strategies. In other words, if each agent has chosen a strategy, and no agent can benefit by changing its strategy, the current set of strategy choices is a Nash Equilibrium. This means that no player has a reason to choose a strategy other than its equilibrium strategy.

5.8. Game without any Constraint

The strategic interaction of agents is modeled as a Cournot model in continuous time. Our framework borrows heavily from the Cournot model of repeated games, as developed by Cournot in 1838 [Gibbons 1992]. First, we consider a very simple version of Cournot's model, and then, in each subsequent section, some variations on the model. In this model, we assume that agents choose their actions simultaneously since it is easier to describe a game in continuous time with simultaneous action. Also, we assume that agents do not engage in forms of competition other than quantity of trading or price.

5.8.1. The Model

The normal form representation of a game is specified by its players, available actions and the payoff of each player in different states. Here, *k+1* players are considered. These players are either rational large agents or non-rational noise traders. We consider *k* large agents and one noise trader. Large agents trade to maximize their rewards, whereas a noise trader agent trades without a particular goal. The trading volume of asset is considered to be the agent's action: $a_j^i(t) = dx_j^i(t)$. Basically, the agent at each time chooses to buy $(a_j^i(t) > 0)$, sell $(a_j^i(t) < 0)$, or hold $(a_j^i(t) = 0)$, the asset. Each large agent chooses its action without knowledge of the others' actions. The payoff of a player in this game is equal to its reward function.

The pay off function of each agent depends on that agent's holding and trading volume of its assets and the price of assets. The asset price model (Equation (5.1)) has three parts: $s(t)$, $\gamma(Z - X(t))$ and $\lambda A(t)$. *s(t)* follows the differential stochastic equation and shows the effects of an uncertain environment on the fortunes of asset price. The last two terms, $(\gamma(Z - X(t))$ and $\lambda A(t))$, measure the impact of order flow and inventory on the asset price. These two terms can be controlled by agents' actions. However, *s(t)* cannot be controlled. The effect of *s(t)* on the asset price turns out to be very important in our analysis. The agents trade the advantage of a high level of an asset when *s(t)* is high, against the disadvantage when *s(t)* is low. The Cournot game with uncertainty can be applied to model this trading volume game in the market.

5.8.1.1. Nash Equilibrium

The Nash Equilibrium of the game is the set of strategies $(a_1^*(t),\ldots,a_k^*(t))$ if for each agent i, $a_i^*(t)$ is at least tied for agent i's best response to the strategies specified for the k other agents:

$J_i(a_1^*(t),\ldots a_{i-1}^*(t),a_i^*(t),a_{i+1}^*(t),\ldots,a_k^*(t)) \geq J_i(a_1^*(t),\ldots a_{i-1}^*(t),a_i(t),a_{i+1}^*(t),\ldots,a_k^*(t)) \forall a_i(t)$,

where J_i is the reward function of agent *I*, and $a_i^*(t)$ is the solution of:

$$max\ J_i(a_1^*(t),\ldots a_{i-1}^*(t),a_i(t),a_{i+1}^*(t),\ldots,a_k^*(t)) . \tag{5.2}$$

Basically, each agent's best action is determined based on Equation (5.2), taking $(a_1^*(t),\ldots a_{i-1}^*(t),a_{i+1}^*(t),\ldots,a_k^*(t))$ as given.

5.8.2. Static Portfolio

Let the risk aversion factor of agent *i* be zero and let players be playing for *one* trading period. The agent's reward function is then given by:

$$J_i = x_i(t)\Pr(s(t),X(t),A(t)) - c(a_i(t)) . \tag{5.3}$$

We assume that $0 \leq x_j^i(t) < \infty$ for all *i* and *j*. $x_j(t)$ is the vector ($x_j^1(t)$, $x_j^2(t)$,…, $x_j^n(t)$), which lies in $(0,\infty)^n$. In the Equation (5.3), each agent's payoff is determined by the agent's action and its holding position of assets and global quantity *A(t)* and *X(t)*.

In the market, the agent is playing against other agents. Although it is assumed in this game that all agents choose their actions at the same time, it does not mean that the parties necessarily act simultaneously. After agents make decisions about their trading volumes, then the market determines how the prices of assets change.

The game in the market is defined as:

Players: Agent 1,…, Agent *k,* and noise trader

Natural number *n*: number of assets in the market

Protocol:

For *t=1*

Agent *i* selects *a_i(t)*

Noise trader selects *O(t)*

Market selects price *(Pr)*

Each agent seeks to maximize its reward function. The agent's objective function is:

$$\underset{a_i(t)}{Max}\ E\big(x_i(t)s(t)-\gamma\, x_i(t)(Z-X(t))+\lambda\, x_i(t)A(t)-c(a_i(t))\big).$$

Each agent may buy or sell various assets at set prices at the beginning of the trading period, and the market determines the assets' returns at the end of that period. An agent is allowed to distribute its current wealth across all *n* assets in one period.

5.8.2.1. Illustrative Example

The purpose of this section is to present a simple example to illustrate the game in the market. Let us consider two large agents in the market who are playing a game in the market for one period of time. a_1 and a_2 denote the quantities traded by Agent 1 and 2, respectively. The asset price follows Equation (5.1): $\Pr = s(1)-\gamma(Z-X(0))+(\lambda+\gamma)A$. Agent *i*'s total cost for trading a_i is $c(a_i)$. We assume that the agents choose their actions simultaneously. In this game, the strategies available to each agent are the different quantities they may trade. We assume that trading volume is continuously divisible and that an agent can trade any fraction of an asset. An agent cannot hold and trade more than asset supply *Z* and does not hold any cash or debt. For player *i*, a_i^* must solve the following optimization problem:

$$\underset{a_i}{Max}\; E((x_i(0)+a_i)\left(s(1)-\gamma(Z-X(0))+(\lambda+\gamma)(a_i+a_j^*)\right)-c(a_i))$$
$$St:\sum_j a_i^j s^j(0)=0$$
.

Let us consider $E(s(1))=\mu$ and $c(a_i)=0$. These formulas optimize the expected utility function of an agent by considering the optimal action of other agents. This constraint shows that everything is reinvested and there is no consumption or labor income during the investment horizon. After applying the Lagrange multiplier method, we have:

$$\underset{a_i}{Max}\; (x_i(0)+a_i)\left(\mu-\gamma(Z-X(0))+(\lambda+\gamma)(a_i+a_j^*)\right)-\theta a_i s(0)\,.$$

The first order equation yields:

$$a_i^* = \frac{1}{2(\lambda+\gamma)}(\gamma(Z-X(0))-\mu+\theta s(0))-\frac{a_j^*}{2}-\frac{x_i(0)}{2}\,.$$

Thus, if the quantity pair (a_1^*, a_2^*) is a Nash equilibrium, the agent's action must be:

$$a_i^* = \frac{1}{3(\lambda+\gamma)}(-\mu+\gamma(Z-X(0))+\theta s(0))+\frac{-2x_i(0)+x_j(0)}{3}\,.$$

The quick insight behind this equilibrium is simple. If the agent is alone in the market, it would choose a_i to optimize $J_i(a_i,0)$ as:

$$a_m = \frac{1}{2(\lambda+\gamma)}(\gamma(Z-X(0))-\mu+\theta s(0))-\frac{x_m(0)}{2}\,.$$

If there are two agents, the optimal action will be: $a_1^* = \frac{a_2}{2}$. Since these actions have lower qualities than actions in the Cournot equilibrium, so the associated price is higher, and so the temptation to increase action is increased. In the Cournot equilibrium, in contrast, the agents do not want to increase or decrease their actions.

5.8.3. Dynamic Portfolio

In this section, the model of a static portfolio is generalized by examining a model in which a number of rounds of trading take place sequentially. Trading begins at time *t=0* and ends at time *t=T*. The agent owns different assets during this time horizon. The resulting dynamic model is structured so that equilibrium prices at each round of trading reflect the information contained in the past and current portfolios.

The game in the market is defined as:

Players: Agent 1,…, Agent *k* and noise trader

Natural number *n*: number of assets in the market

Natural number *T*: investment period

Protocol:

For *t= (0, T)*

Agent *i* selects *a_i(t)*

Noise trader selects *O(t)*

Market selects price *(Pr)*

The payoff function is $E\left[\int_0^T e^{-\alpha t} \frac{\left(x(t)\left(s(t)-\gamma(Z-X(t))+\lambda A(t)\right)-c(a(t))\right)^{1-\beta}}{1-\beta} dt\right]$, and each agent seeks to maximize its payoff function. Each trade by a large agent impacts the price. It therefore acts strategically and takes its price impact into account when trading.

Each agent may buy various assets at set prices at the beginning of each trading period, and the market determines the assets' returns at the end of that period. An agent is allowed to redistribute its current wealth across all *n* assets in each round. We call the set of all possible sequences, $\{a_i(t), \Pr(s(t), A(t), X(t))\}$, the sample space of the game, and we designate it as Ω. We call any subset of Ω an event. All agents are also allowed to redistribute their current wealth across all *n* assets in each round, but the moves of the other agents are not recorded in the sample space. They do not define events.

5.9.Game with a Debt Constraint

The agent's obligations to debt holders are usually ignored in modeling the strategic interaction between agents in the market. Ignoring the financial aspect of an agent that depends on external financing misses a significant constraint on the agent's actions in the real world. The agent who uses external financing can be forced to liquidate when its debts are recalled. This forces the agent to sell its assets with less utility value than if it held them. Moreover, this forced liquidation leads the asset price to drop in the market.

Agents undertake debt levels that restrict their actions later in a game. The first central insight is that higher debt levels tend to increase an agent's desire to sell assets. As discussed, *s(t)* follows the stochastic differential equation. A lower realization of *s(t)* corresponds to cases where the asset price is low. In this situation, an agent should sell more assets to satisfy its debt holders. In particular, the agent sells its assets even at a low price. The more debt the agent has, the more aggressive it becomes.

Large agents fear a forced liquidation, especially if their cash needs are known by other agents [Brunnermeier 2005]. If rival agents know that an agent needs to sell something quickly, they will also sell the same asset and subsequently buy it back. This kind of trading is called predatory trading [Brunnermeier 2005]. Cramer (2002) described predatory trading as follows: "When you know that one of your numbers is in trouble…. you try to figure out what he owns and you start shorting those stocks…" Goldman, Sachs & Company and other counterparties to Long Term Capital Management (LTCM) did exactly that in 1998 [Brunnermeier 2005]. Predatory trading tends to drive the price down even faster and reduces the liquidation value for distressed agents.

The model of this section provides a framework for predatory trading. During each period, a liquidity event may occur in which an agent is required to trade a large block of an asset in a relatively short time period. This is especially important when many large agents are financially distressed and have to quickly liquidate some or all of their positions to meet cash needs. The need for liquidity is observed by other agents; hence, they may choose to predate. Predation trading may cause an adverse price impact on the market.

Suppose many large agents are playing the game in the market. This game has the following form: first, each agent chooses its own debt level, $D_i(t)$. Then the liquidation event (obligation to pay debt claims) may occur. After agents simultaneously choose their trading volume, the market experiences price fluctuations. The timing of the model is depicted in Figure 5.2.

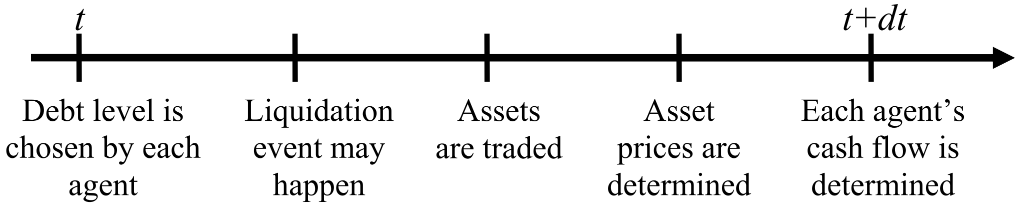

t
t+dt
Debt level is chosen by each agent
Liquidation event may happen
Assets are traded
Asset prices are determined
Each agent's cash flow is determined

Figure 5.2: The Timeline of Game with a Debt Constraint

Here, we do not examine the earlier decision of how much debt the agent should take on. Agents are always obliged to pay debt claims out of their own cash. If the agent is unable to meet its debt obligations, it should take actions which lead to lower levels of its utility. If after these actions the agent still cannot meet debt obligations, the agent goes bankrupt. An increase in debts means that the range of the states over which the firm becomes bankrupt is expanded.

The model presented here is the simplest model of connectivity between financial decisions and actions. Our model is closest in spirit to Brunnermeier (2005). He discussed predatory trading; trading that induces and/or exploits the need of other agents to reduce their positions. The objective function of his model was to maximize the expected terminal wealth, whereas our objective function is to maximize the discounted expected utility function from time zero to *T*.

There are important linkages between debt levels, trading volume decisions and an agent's utility. One possible linkage between them is the rivals' bankruptcy effect. The firm's fortunes will usually improve if one or more of its rivals are driven into bankruptcy. The agent can lead its rival with high debt levels to bankruptcy by predatory trading.

5.9.1.Model

The debt level of agent *i* is represented by $D_i(t)$. Given debt levels, the agent should choose actions with the objective of maximizing the discounted expected utility from time zero to time *T*.

All agents should reduce their asset levels if they have debts. Compared to a zero debt case, agents will receive lower rewards. It is assumed that the debt levels are chosen before actions are decided upon.

5.9.2. Structure and Notation

Risky assets have a supply of *Z*. The agent's position in risky assets, $x(t)$ and $o(t)$, cannot be unlimited, and $x(t)$ and $o(t)$ are less than or equal to *Z*. The trading mechanism works as follows: Each agent trades the asset by choosing its action. At time T the agent's position in the risky assets is:

$$x(T) = x(0) + \int_0^T a(\tau)\,d\tau\ .$$

Large agents are subject to a risk of financial distress. We denote the set of all large agents by L and the set of distressed agents by L^d $(\mathsf{L}^d \in \mathsf{L})$. In this game, large agents are either distressed or they are predators. If an agent is troubled, it must liquidate its position in the risky asset $(a_i(t) \le 0)$ to pay back its debts. A distressed agent is required to sell a large block of assets in a short time period. Hence, at time *t* its action is restricted to:

$$\sum a_i(t)\,\mathrm{Pr}(t) \le -D_i(t)\ .$$

This relationship indicates that a troubled agent must liquidate its position by at least as much as $D_i(t)$. After trading occurs, the agent should pay creditors $D_i(t)$ out of its available cash. For simplicity, we assume that cash is completely invested in assets, so that creditors can collect only earnings from current sales of assets.

The predators are informed of the trading requirement and they trade strategically in the market in order to drop the price of the distressed agents' selling power. At the start of the game, each agent chooses its own actions over the period *[0,T]* to maximize its own expected utility function, assuming the other agents will do likewise. Subject to their initial wealth and debt level constraints, they solve the following model:

$$Max\ E\left[\int_0^T e^{-\alpha t}\ u(\,w_t)\ dt\right]$$

Subject to $x_i(0)\,\Pr(0) = W_i(0)$

$\sum a_i(t)\,\Pr(t) \le -D_i(t)$.

$\Pr(s(t), A(t), X(t))$ is the price of assets at time *t*. The high value of *A(t)* (buy more) corresponds to a higher value of $\Pr(s(t), A(t), X(t))$. An increase in Agent *i's* debts causes a decrease in $a_i(t)$. In particular, higher levels of debts make it optimal for Agent *i* to sell more regardless of any information from its rival Agent *j*. An agent becomes bankrupt when it cannot cover its debt obligations, which happens to be the state in which the marginal returns to extra output are very low.

5.10. Conclusion

In recent years a literature has emerged that studies non-cooperative game theory from a decision-theoretic point of view [Aumann 1995]. Following this approach, we discuss a portfolio management game in terms of the rationality of the players and their cognitive states, that is, what they know or believe about the game and about each other's actions. The model describes the uplink of a market that consists of *k+1* agents investing in *n* assets.

CHAPTER 6

CONCLUSION

The dynamic portfolio model distributes the available funds accessible to or owned by an agent (investor) to different assets and opportunities over time in an uncertain environment. The general problem is how to enable an agent to maximize its expected utility by taking actions (selling or buying assets) in an uncertain environment. Most asset returns are uncertain, and investors do not know the probabilities of different future returns. This uncertainty must be considered since we do not have full knowledge about other agents' actions or events within the environment.

The agent should choose actions to change the portfolio based on the environment. If the agent does not take any action, events (observable and unobservable) will change the market and the value of the portfolio.

In this research, we illustrated the potential of describing and optimizing dynamic portfolio selection under the framework of Semi-Markov Decision Processes. The detail of the model was discussed in Chapter 2, where we gave a rigorous mathematical formulation of the problem. It was assumed that a single agent was trying to strategize its actions in order to maximize; its objective function, defined as capturing the agent's level of risk aversion. At any time, the agent keeps track of the positions and prices of all the possible assets in the market. The asset prices are modeled as a Geometric Brownian Motion process. The Q-learning technique is applied to define the solution.

In Chapter 3 we introduced the Q-learning approach and presented some numerical results to the above problem for:

- one risk-free asset and one risky asset in one period of time

- one risk-free asset and *n* risky assets in one period of time
- one risk-free asset and *n* risky assets in more than one period of time

As we indicated above, the agent's state of information includes positions and prices for all assets in the market. While it is possible for the agent to limit this to a small subset of the market, there are no provisions made in the model to account for the partiality of the information about the rest of the market. On the other hand, the agent can include in its state of information, all the possible options. This would certainly result in incomplete information of the system states. In Chapter 4, we tried to address these shortcomings of Chapter 2's model and discussed the impact of incomplete information on portfolio choices. Furthermore, in Chapter 5 we have discussed the impact of agent actions on portfolio selection.

REFERENCES

[Akian 2001] Akian, M., Sulem, A. and Taksar, M.I., Dynamic Optimization of Long-term Growth Rate for a Portfolio with Transaction Costs and Logarithmic Utility, Mathematical Finance, 11(2), 2001

[Armano 2005] Armano, G., Marchesi, M. and Murru, A., A Hybrid Genetic-neural architecture for stock indexes forecasting, Information Sciences, 170(1), 2005

[Aumann 1995] Aumann, R. and Machler, M., Repeated Games with incomplete information, MIT Press, 1995

[Baek 2005] Baek, S.K., Jung, W., Kwon, O. and Moon, H., Transfer Entropy Analysis of the Stock Market, ArXiv Physics e-prints, 2005

[Barberis 2000] Barberis, N., Investing for the Long Run when Returns are Predictable, Journal of Finance, 55, p. 225-264, 2000

[Bartmann 1980] Bartmann, D., Mittelfristige Productionsplanung Bei Ungewissen Szenarieb, Opns Res. 28, B187-B204, 1984

[Basak 1995] Basak, S., A General Equilibrium Model of Portfolio Insurance, Review of Financial Studies, 8, 1995

[Basalto 2005] Basalto, N., Bellotti, R., De Carlo, F., Facchi, P. and Pascazio, S., Clustering Stock Market Companies Via Chaotic Map Synchronization, Physica A: Statistical Mechanics and its Applications, v. 345, 2005

[Bellman 1957] Bellman, R, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957

[Beltratti 1992] Beltratti, A. and Margarita, S., Evolution of Trading Strategies among Heterogeneous Artificial Economic Agents, MIT Press, Cambridge, MA, 1992

[Bertsekas 1996] Bertsekas, D. and Tsitsiklis, J., Neuro-Dynamic Programming, Athena Scientific, Belmont, MA, 1996

[Brandt 2004] Brandt, M.W., Goyal, A., Santa-Clara, P. and Storud, J., A Simulation Approach to Dynamic Portfolio Choice with an Application to Learning About Return Predictability, NBER Working Papers 10934, National Bureau of Economic Research, Inc., 2004

[Brendle 2005] Brendle, S., Portfolio Selection under Incomplete Information, Stochastic Processes and their Applications, 116, p. 701-723, 2005

[Brennan 1998] Brennan, M.J., The Role of Learning in Dynamic Portfolio Decisions, European Finance Review, 1, p. 295-306, 1998

[Brennan 2001] Brennan, M.J. and Xia, Y., Stock Price Volatility and Equity Premium, Journal of Monetary Economics, 47, p. 249-283, 2001

[Brennan 2001b] Brennan, M.J. and Xia, Y., Assessing Asset Pricing Anomalies, The Review of Financial Studies, Vol. 14, No. 4, p. 905-942, 2001

[Browne 1996] Browne, S. and Whitt, W., Portfolio Choice and the Bayesian Kelly Criterion, Advances in Applied Probability, 28, p. 1145-1176, 1996

[Brunnermeier 2005] Brunnermeier, M.K. and Pedersen, L.H., Predatory Trading, The Journal of Finance, Vol 4, 2005

[Cheng 1997] Cheng, M. and Madhavan, A., In search of Liquidity: Block Trades in the Upstairs and Downstairs Markets, Review of Financial Studies, 10, p. 175-203, 1997

[Cramer 2002] Cramer, J., Confessions of a Street Addict, Simon and Schuster, New York, 2002

[Cox 1989] Cox, J. and Huang, C.F., Optimal Consumption and Portfolio Polices when Asset Prices Follow a Diffusion Process, Journal of Economic Theory, 49, 1989

[Cvitanic 2003] Cvitanic, J., Goukasian, L. and Zapatero, F., Monte Carlo Computation of Optimal Portfolios in Complete Markets, Journal of Economic Dynamics and Control, 27, p. 971-986, 2003

[Damodaran 2008] Damodaran, A., Strategic Risk Taking: a Framework for Risk Management, Wharton School Pub., 2008

[Davis 1990] Davis, M.H.A. and Norman, A., Portfolio Selection with Transaction Costs, Mathematics of Operations Research, 15, 1990

[Derman 1984] Derman, C., Liberman, G.J., Ross, S.M., A Stochastic Sequential Allocation Model., Opns Res. 32, 1984

[Detemple 1986] Detemple, J.B., Asset Pricing in a Production Economy with Incomplete Information, Journal of Finance, 41(2), p. 383-391, 1986

[Detemple 2003] Detemple, J.B., Garcia, R. and Rindisbacher, M., A Monte Carlo Method for Optimal Portfolios, The journal of Finance, 58(1), 2003

[Domenica 2007] Domenica, N.D., Mitraa, G., Valentea, P. and Birbilisa, G., Stochastic Programming and Scenario Generation within a Simulation Framework: An Information Systems Perspective, Decision Support Systems, Vol 42, Issue 4, 2007

[Downes 2006] Downes, J. and Elliot Goodman, J., Finance and Investment Dictionary of Finance and Investment Terms, 7th edition, Barron's Educational Series, Inc, 2006

[Duba 2000] Duda, R., Hart, P., and Stork, D., Pattern Classification, Wiley-Interscience, 2000

[Elton 1997] Elton, E.J. and Gruber M.J., Modern Portfolio Theory: 1950 to Date, Journal of Banking & Finance 21, p. 1743-1759, 1997

[Fama 1970] Fama, E.F., Multiperiod Consumption-investment Decisions, American Economics Review 60, p. 163-174, 1970

[Fan 2001] Fan, A. and Palaniswami, M., Stock Selection using Support Vector Machines, Proceedings of the International Joint Conference on Neural Networks, 2001

[Feldman 2007] Feldman, D., Incomplete Information Equilibria: Separation Theorems and other Myths, Annals of Operations Research, 151, 1, p. 119-149, 2007

[Franklin 1996] Franklin, S. and Graesser, A., Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents, Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages, Springer-Verlag, 1996

[Fu 1970] Fu, K.S., Learning Control Systems- Review and Outlook, IEEE Transactions on Automatic Control, 1970

[Gao 2000] Gao, X. and Chan, L., An Algorithm for Trading and Portfolio Management Using Q-learning and Sharpe Ratio Maximization, Proceedings of the International Conference on Neural Information Processing, 2000

[Gennotte 1986] Gennotte, G., Optimal Portfolio Choice under Incomplete Information, Journal of Finance, 61, p. 733-749, 1986

[Gibbons 1992] Gibbons, R., Game Theory for Applied Economists, Princeton University Press, 1992

[Gode 1993] Gode, D.K. and Sunder, S., Allocative Efficiency of Markets with Zero Intelligence Traders, Journal of Political Economy 101, 1993

[Grossman 1976] Grossman, S.J., On the Efficiency of Competitive Stock Markets where Traders have Diverse Information, Journal of Finance 31, 1976

[Grossman 1981] Grossman, S.J. and Shiller, R.J., The Determinants Of The Variability Of Stock Market Prices, American Economic Review 71, p. 222-227, 1981

[Grossman 1996] Grossman, S.J. and Zhou, Z., Equilibrium Analysis of Portfolio Insurance, Journal of Finance 51, 1996

[Guo 2004] Guo, W. and Xu, C., Optimal Portfolio Selection when Stock Prices Follow an Jump-diffusion Process, Mathematical Methods of Operations Research, 60, 2004

[Han 2005] Han, J., An Approximate Linear Programming Approach, Dissertation of Doctor of Philosophy, Stanford University, 2005

[Harrison 1981] Harrison, M. and Pliska, S., Martingales and Stochastic Integrals in the Theory of Continuous Trading, Stochastic Process Appl. 11, 1981

[Kaelbling 1996] Kaelbling, L.P., Littman, M.L., and Moore, A.W., Reinforcement Learning: A Survey, Vol 4, p. 237-285, 1996

[Karatzas 1987] Karatzas, J.L. and Shreve, S.E., Optimal Portfolio and Consumption Decisions for a Small investor on a Finite Horizon, SIAM Journal of Control and Optimization, 25(6), 1987

[Kim 1996] Kim, S. and Omberg, E., Dynamic Nonmyopic Portfolio Behavior, The Review of Financial Studies, 9(1), 1996

[Korn 1998] Korn, R., Portfolio Optimization with Strictly Positive Transaction Costs and Impulse Control, Finance and Stochastics, 2, 1998

[Krawczyk 2000] Krawczyk, J.B., A Markovian Approximated Solution to a Portfolio Management Problem, Computing in Economics and Finance 233, Society for Computational Economics, 2000

[Kullmann 2004] Kullmann, L., Kert´esz, J., and Mantegna, R.N., Identification of Clusters of Companies in Stock Indices via Potts Super-paramagnetic Transitions, Physica A: Statistical Mechanics and its Applications, v. 287, issue 3-4, 2004

[Kuo 1998] Kuo, R.J., A Decision Support System for the Stock Market through Integration of Fuzzy Neural Networks and Fuzzy Delphi, Applied Intelligence, 6, 1998

[Kveton 2006] Kveton, B., Hauskrecht, M., and Guestrin, C., Solving Factored MDPs with Hybrid State and Action Variables, Journal of Artificial Intelligence Research , 27, p. 153–201, 2006

[Labbi 2007] Labbi, A. and Berrospi, C., Optimizing Marketing Planning and Budgeting Using Markov Decision Processes: An Airline Case Study, Business Optimization Vol. 51, Number 3/4, 2007

[LeBaron 2000] LeBaron, B., Agent Based Computational Finance: Suggested Readings and Early Research, Journal of Economic Dynamics and Control, 24, 2000

[LeBaron 2006] LeBaron, B., Agent-based computational finance, Handbook of Computational Economics, 2006

[Lee 2002] Lee, J.W. and O, J., A multi-agent Q-learning Framework for Optimizing Stock Trading Systems, Proceeding of the International Conference on Database and Expert Systems Applications, p. 153-163, 2002

[Lee 2004] Lee, R.S.T., iJADE Stock Advisor: An Intelligent Agent Based Stock Prediction System using Hybrid RBF Recurrent Network, IEEE Transactions on Systems, Man, Cybernetics, Part A: Systems and Humans, 34(3), 2004

[Lettau 1997] Lettau, M., Explaining the Facts with Adaptive Agents: The Case of Mutual Fund Flows, Journal of Economic Dynamics and Control, 21, 1117, 1997

[Lioui 2001] Lioui, A. and Poncet, P., On Optimal Portfolio Choice Under Stochastic Interest Rates, Journal of Economic Dynamics and Control, 25, 2001

[Liptser 2001] Liptser, R.S. and Shiryayev, A.N., Statistics of Random Processes II: Applications, New York, Springer, 2001

[Liu 2007] Liu, J., Portfolio Choice in Stochastic Environments, The Review of Financial Studies, Vol. 20, Issue 1, p. 1-39 , 2007

[Luenberger 1997] Luenberger, D.G., Investment Science, Oxford University Press, 1997

[Lundtofte 2006] Lundtofte, F., The Effect of Information Quality on Optimal Portfolio Choice, The Financial Review, 41, p. 157–185, 2006

[Magil 1976] Magil, M. and Constantinides, G., Portfolio Selection with Transaction Costs, Journal of Economic Theory, 13, 1976

[Markowitz 1952] Markowitz, H., Portfolio selection, Journal of Finance 7, pp. 77-91, 1952

[Maybeck 1979] Maybeck, P.S., Stochastic Models, Estimation, and Control, Vol. 1, Academic Press, Inc. 1979

[Mendel 1966] Mendel, J.M., A Survey of Learning Control Systems, ISA Transactions, 1966

[Merton 1969] Merton, R.C., Lifetime Portfolio Selection under Uncertainty: The Continuous-time Case, Review of Economics and Statistics, 51, 1969

[Merton 1980] Merton, R.C., On Estimating the Expected Return on the Market: an Exploratory Investigation, Journal of Financial Economics, 8, p. 323-362, 1980

[Merton 1990] Merton, R.C., Continuous Time Finance, Basil Blackwell, Oxford, 1990

[Minsky 1954] Minsky, M. L., Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model Problem, Dissertation of Doctor of Philosophy, Princeton University, 1954

[Moody 2001] Moody, J. and Saffell, M., Learning to trade via direct reinforcement, IEEE Transactions on Neural Networks, 12 (4), p. 875–889, 2001

[Mulvey 2006] Mulvey, J.M., Simsek, K.D. and Zhang, Z., Improving Investment Performance for Pension Plans, Journal of Asset Management, 7, 2006

[Neuneier 1998] Neuneier, R., Enhancing Q-learning for Optimal Asset Allocation, Advances in Neural Information Processing Systems 10, MITPress, Cambridge 1998

[Neuneier 1999] Neuneier, R. and Mihatsch, O., Risk Sensitive Reinforcement Learning, Advances in Neural Information Processing Systems 11, p. 1031-1037, MITPress, Cambridge, 1999

[Norman 1965] Norman, J.M. and White, D.J. ,Control of Cash Reserves, Opl Res. Q. 16, p. 309-328, 1965

[O 2006] O, J., Lee, J., Lee, J.W., and Zhang, B., Adaptive Stock Trading With Dynamic Asset Allocation Using Reinforcement Learning, Information Science, 176, p. 2121-2147, 2006

[Øksendal 2002] Øksendal, B. and Sulem, A., Optimal Consumption and Portfolio with both Fixed and Proportional Transaction Costs, SIAM J. Control Optim., 40(6), 2002

[Osborne 1972] Osborne, M.F.M, Random Nature of Stock Market Prices, Journal of Economics and Business, 6, p. 220-233, 1972

[Pafka 2004] Pafka, S., Potters, M. and Kondor, I., Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization, 2004

[Petrie 1996] Petrie, C.I., Agent-based engineering, the Web, and intelligence, IEEE Expert: Intelligent Systems and Their Applications, v.11 n.6, p. 24-29, December 1996

[Pliska 1986] Pliska, S.R., A Stochastic Calculus model of Continuous Trading: Optimal Portfolio, Mathematics of Operations Research 11, 1986

[Pritsker 2005] Pritsker, M., Large Investors: Implications for Equilibrium Asset Returns, Shock Absorption, and Liquidity, FEDS working paper No. 2005-36, 2005

[Puterman 1994] Puterman, M., Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, New York, NY, 1994

[Ramezani 1998] Ramezani, C.A. and Zeng, Y., Maximum Likelihood Estimation of Asymmetric Jump-Diffusion Processes: application to security prices, 1998, http://www.stat.purdue.edu/~ntuzov/niks%20files/papers/jum_diff/Ramezani%20Zeng98.pdf

[Rasmusen 1992] Rasmusen, E.B., Game Theory in Finance, The New Palgrave Dictionary of Money and Finance. Ed. John Eatwell, Murray Milgate, and Peter Newman. New York: Stockton Press, 1992

[Richard 1975] Richard, S., Optimal Consumption, Portfolio, and Life Insurance Rules for an Uncertain Lived Individual in a Continuous-time Model, Journal of Financial Economics, 2, 1975

[Russell 1995] Russell, S.J. and Norvig, P., Artificial Intelligence: A Modern Approach, Englewood Cliffs, NJ: Prentice Hall, 1995

[Saad 1998] Saad, E.W., Prokhorov, D.V., and Wunsch, D.C., Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks, IEEE Transactions on Neural Networks, 9(6), p. 1456-1470, 1998

[Samuelson 1969] Samuelson, P.A., Lifetime Portfolio Selection by Dynamic Stochastic Programming, Review of Economics and Statistics, 51, 1969

[Smith 1994] Smith, D.C., Cypher, A. and Spohrer, J., KidSim: Programming Agents without a Programming Language, Communications of the ACM, 37, 7, p. 55-67, 1994

[Snarska 2006] Snarska, M. and Krzych, J., Automatic Trading Agent RMT Based Portfolio Theory and Portfolio Selection, Acta Phys. Pol. B 37, 2006

[Sutton 1998] Sutton, R.S. and Barto, A.G., Reinforcement Learning: An Introduction. MIT Press, 1998

[Sycara 2003] Sycara, K., Giampapa, J.A., Langley, B.K. and Paolucci, M., The RETSINA MAS, a case study. In SELMAS, vol. LNCS 2603, edited by A.Garcia, C. Lucena, F. Zambonelli, A. Omici, and J. Castro, pp. 232-250. New York: Springer-Verlag, 2003

[Taksar 1988] Taksar, M.I., Klass, M.J. and Assef, D., A Diffusion Model for Optimal Portfolio Selection in the Presence of Brokerage Fees, Mathematics of Operations Research, 13(2), 1988

[Tapiero 1998] Tapiero, C.S., Applied Stochastic Models and Control for Finance and Insurance, Boston: Kluwer Academic Publishers, 1998

[Vlassis 2003] Vlassis, N., A Concise Introduction to Multiagent Systems and Distributed AI, Introductory Text, University of Amsterdam, 2003

[Von Neumann 1944] Von Neumann, J. and Morgenstern, O., Theory of Games and Economic Behavior, New York: John Wiley and Sons, 1944

[Wachter 2002] Wachter, J., Portfolio and Consumption Decisions under Mean-reverting Returns: An Exact Solution for Complete Markets, Journal of Financial and Quantitative Analysis, 37(1), 2002

[Waltz 1965] Waltz, M.D. and Fu, K.S., A Heuristic Approach to Reinforcement Learning Control Systems, IEEE Transactions on Automatic Control, 1965

[Wessels 1980] Wessels, J., Markov Decision Processes; Implementation Aspects, Memorandum COSOR 80-14, Department of Mathematics, Eindhoven, 1980

[White 1993] White, D.J., A Survey of Applications of Markov Decision Processes, J. Opl Res. Soc. 44, 1993

[Wooldridge 1995] Woolbridge, M. and Jennings, N.R., Agent Theories, Architectures, and Languages: a Survey, in Wooldridge and Jennings Eds., Intelligent Agents, Berlin: Springer-Verlag, pp. 1-22, 1995

[Xia 2001] Xia, Y., Learning about Predictability: the Effect of Parameter 2001

[Xiu Uncertainty on Dynamic Asset Allocation, The Journal of Finance, LVI, 1, 2000] Xiu, G. and Laiwan, C., Algorithm for Trading and Portfolio Management Using Qlearning and Sharpe Ratio Maximization, Proceedings of ICONIP 2000, Korea, p. 832–837, 2000

[XufreCasqueiroa 2006] Xufre Casqueiroa, P. and Rodrigues, A.J.L., Neuro-Dynamic Trading Methods, European Journal of Operational Research ,Volume 175, Issue 3, 16 December 2006

[Zhao 2000] Zhao, Y., Dynamic Investment Models with Downside Risk Control, Dissertation of Doctor of Philosophy, The University of British Columbia, 2000

CURRICULUM VITA

HALEH VALIAN

EDUCATION

Oct. 2009 Ph.D., Industrial and Systems Engineering
Rutgers, The State University of New Jersey
New Brunswick, New Jersey

May 2006 M.S., Industrial and Systems Engineering
Rutgers, The State University of New Jersey
New Brunswick, New Jersey

Jan. 2000 M.S., Industrial Engineering
Iran University of Science and Technology
Tehran, Iran

Jan. 1997 B.S., Industrial Engineering
Amirkabir University of Technology
Tehran, Iran

WORK EXPERIENCE

Jan. 2004-
May 2008 Graduate Research and Teaching Assistant, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, New Jersey

Feb 1997-
Feb 2004 Manager, Operational Methods Department, Bank of Industry and Mine, Tehran, Iran

PUBLICATIONS

Jan 2009 D. Golmohammadi, R. Creese and H. Valian, Application of Backpropagation Learning Rule in Neural Networks for Supplier Ranking, International Journal of Product Development, Vol 8, No. 3

Jan. 2007 H. Valian and M.A. Jafari, Optimization of Open Water Disposal Site for Dredged Material Using NLP, Journal of dredging Engineering, Vol. 8, No.1

Feb. 2006 H. Valian, T. Williams, and M.A. Jafari, Optimization Open-water Disposal Sites by Using NLP, Proceedings of GeoCongress 2006, Atlanta, Georgia