Statistical and Financial Aspects of Research and Development Portfolios

Michael Andrea Talias

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Statistical Science

University College London

January 2000

Αφιερωμένο στους γονείς μου Ανδρέα και Βαρβάρα

Abstract

Sequential evaluation and decision problems must frequently be solved under uncertainty. The sequential nature of activities like research, development or exploration, requires optimal funding criteria that take into account the fact that further funding decisions will be made throughout the future.

In this thesis, we examine several sequential and parallel strategies for R&D project selection and capital budgeting problems. Some of these problems have as a solution a prioritisation index. We pay particular interest to the Pearson and Gittins indices. We relate the Pearson index to the Neyman-Pearson lemma and state clearly the kind of problems the Pearson index solves. We reformulate this problem using non-linear utility function and show how to solve it for different utility functions.

These kind of indices may need to have a forecast for R&D rewards or costs. We discuss adaptive prediction, we derive the forecasting rule for various data generating processes, and study the behaviour of unconditional and conditional forecast variances. Furthermore, we study the connection of R&D projects with real options theory, and discuss the suitability of this methodology and its fundamental principle of economic rationality or no-arbitrage.

Finally, the multi-armed bandit problem is introduced and is reconciled with the option pricing. We prove that an additional condition is required for an index policy to be optimal when two projects are selected simultaneously with criterion the sum of their indices to be maximum.

Acknowledgements

I would like to thank my supervisors Prof. A. P. Dawid and Prof. S. Senn for their invaluable support and guidance during the last 4 years.

I am grateful to the Department for partially supporting my studies with a 3 years Departmental Research Studentship provided by CIBA-Geigy (Novartis).

Finally, my deepest gratitude goes to my parents for their encouragement, understanding and financial support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The objective of this thesis is to examine issues related to various decision analytic approaches to sequential choice and implementation of projects with special motivation from and application to the Pharmaceutical industry. More specifically we study prioritisation indices such as the Pearson index (Pearson, 1972) and the Gittins index (Gittins, 1979), the problems solved by these indices and other statistical and financial aspects related to these problems.

These indices can have an application to prioritize or finance Research and Development (R&D) projects, which should be understood as multi-stage long term financial investments meaning that investment decisions are made sequentially, the initial investment may be very large, and they are characterized by high uncertainty and high rewards when they are completed successfully. For example, Pharmaceutical research and development (Gittins, 1996, 1997) has a time scale of at least 15 years or more, where costs are incurred, but it can be easily be more than 30 years. Also, no financial benefit accrues until a drug is marketed. Thus, great uncertainty characterizes the duration of the research and development process and the ultimate technical success or failure of research process.

In Chapter 2, we begin by introducing the R&D selection problem and discuss the existence of optimal priority index rules. We study some priority indices which although they originated from sequential decision process problems, have an application to portfolio selection problems. We discuss the capital budgeting problem as a Linear programming problem and its relation to the knapsack problem. The innovation of this chapter is that we relate the Neyman-Pearson lemma to the Pearson index and provide a comparison of the Pearson index with the Gittins index. We also examine sequential decision processes like the secretary problem and the search problem.

In Chapter 3, we review some Portfolio models. We develop a stochastic resource allocation model based on an extension of the new interpretation of the Pearson index given in Chapter 2. We maximize non-linear utility functions (exponential and negative exponential) and assume the Normal distribution for random rewards. Part of the solution of these optimization problems is to study the Kuhn-Tucker conditions and the quadratic knapsack problem.

In Chapter 4, the theme is adaptive prediction. Future rewards and costs change continuously and their actual magnitude determine the ranking of each project. Because of the high uncertainty in the duration of the R&D procedure, future rewards need to be predicted based on the information gained by observing Market conditions up to any current time. We assume that the data generating process is specified by an Ito stochastic differential equation. Examples are Geometric Brownian motion, Ornstein-Uhlenbech process etc. We want to predict the value of the process some steps ahead, at a certain time point in the future. We derive the forecasting rules for this kind of prediction for each of these processes. We also study the behaviour of conditional and unconditional variance of forecast. An example of forecasting system with autoregressive conditional heteroscedastic model is given. We present the theory of Option pricing (Martingale approach) and explain its relation to forecasting.

In Chapter 5, we discuss the theory of Real Options. Real options is the application of financial option pricing theory to real investment, such as the valuation and management of an $R\&D$ project, taking account of flexibility on resource allocation decisions. We compare the real option approach and the decision analysis approach to the capital budgeting problem. We give emphasis to the no-arbitrage condition and discuss its usefulness. We set up an example to explain the concept of option value and how it arises due to strategic options.

In Chapter 6, the Multi-Armed Bandit problem is introduced. We prove that it is optimal to add two Gittins indices and select two projects simultaneously if the sum of their indices is maximum under certain assumptions. We reconcile the Multi-Armed Bandit problem with option pricing.

In Chapter 7, we summarize the results and discuss future lines of research.

Chapter 2

The R&D Selection Problem

2.1 Introduction

In this chapter we discuss several sequential and parallel strategies for the R&D project selection and capital budgeting problems. Evaluation and selection of R&D projects is done to aid the best use of limited resources, since development of the set of all available projects is not usually attainable.

We focus especially on ranking indices, and more specifically on the Pearson Index (Gittins, 1996; Senn, 1996, 1997, 1998), which is used as a simple method for evaluating R&D projects in the Pharmaceutical Industry. It is defined as the ratio of expected net present value of the reward of an R&D project to its expected development cost. This productivity index (Regan & Senn, 1997) is used by practitioners as a measure of project value, and also as an indication of the relative of the relative values of available projects (ranking index).

The Pearson Index is defined (and used) without any reference to the R &D manager's objectives. For example, when the manager's goal is to maximise the net present value of the expected profit stream from any project undertaken, it is not clear why one should try to maximize the ratio of expected reward to the expected cost, and not, for example, to seek maximization of their long-run difference. In this chapter we address the issues of which project selection problem the Pearson index solves, and what is its relationship to Portfolio theory and resource allocation models.

Another relevant question is how the selection strategy for prioritisation of projects would differ if the optimisation were to be implemented sequentially over time. Sequential decision problems are formulated as optimal stopping problems and can have an application to the decision making problem of whether to continue an R&D project which will gain an uncertain reward, or to abandon it. We consider a class of sequential decision problems, such as hiring a secretary, selling an asset, the two-armed bandit problem, search problem etc, and show how their solutions are related to the Pearson Index.

The R&D and capital budgeting problems can be related to statistical decision theory, for example it provides us with search theory application in the capital budgeting problem. We also present an interpretation to the Pearson Index using the Neyman-Pearson lemma. The likelihood ratio test is optimal for choosing between two hypotheses in the sense of maximising the power of the test for a specific significance level. Likewise, the Pearson Index can be seen as the optimal rule for selecting a subset of projects under a budget constraint.

In the next section the R&D selection problem is discussed and emphasis is given to the parallel and series selection methods. In section (2.3) the Capital budgeting problem is studied. Then section (2.4) refers to Pearson index, its origin and related problems. Sequential decision processes are examined in the section (2.5) and a conclusion is presented in the last section (2.6).

2.2 The R&D selection problem

In this section, an introduction is given to parallel and sequential strategies in R&D projects. It is shown how the R&D project selection and capital budgeting problem can be related to optimal statistical decision theory. The R&D selection problem is concerned with how to evaluate and identify the best subset of projects among several proposed ones under some constraints. The argument under consideration is to whether tasks should be proceeded in parallel or in series.

The parallel selection method is addressed with the following question. Suppose that G is the set of the available projects. How should one separate these projects into two sets in order to decide which projects will or will not be developed? To get an effective analysis of the decision problem of choosing projects, a criterion to choose amongst projects is needed. When the target is to maximize some measure of reward or utility subject to a budget constraint, the problem may be formulated using a linear programming approach.

In the series system, the projects are selected in a sequential manner, that is, try one project and work on it. Then observe the result, and choose another one and so on. To model problems which involve sequential decisions over time, assuming that only one task can be undertaken at a time, a dynamic programming approach is used. Therefore the optimization is with respect to time.

To connect the R&D selection problem with optimal statistical decision problems the following problems are considered in this chapter. These are "the secretary problem" , "discrete search problems", and "the job sequencing problems". A characteristic of such sequential decision problems is that the decision maker must make an irrevocable choice from a number of applicants (jobs or tasks) whose values are revealed only sequentially.

Parallel and Series systems of tasks are considered. The problem is to determine an optimal sequence to implement these projects so as to optimize an objective function. In some cases the solution to these optimisation problem is an index.

Methods like ranking procedures, scoring or rating methods, decision analysis and optimization techniques such as linear programming and dynamic programming were used in the past. Different scenarios could be of interest such as a series system of n tasks, a parallel system of n alternatives etc. Some structures are considered in the next section.

2.2.1 Probability structure of different strategies

A Series System

Consider, for example, an R&D project which is composed of several tasks. The first case is a series system of tasks where the tasks are performed sequentially over time. One stage is implemented at each time and the next task is initiated if and only if the current task is completed successfully. The selection process is terminated as soon as one of the tasks is failed. If there are n possible independent stages and the general stage $j (j = 1, \dots n)$ has a probability of success p_j then, the probability of success of the system is:

$$\text{Probability of completing the procedure} = \prod_{j=1}^{n} p_j.$$

Parallel system of alternatives

A more complicated system of sequential selection processes than the series system is the *parallel system of alternatives.* One stage can be thought of as parallel system of alternatives, that is the situation in which several alternatives trials can be attempted until the first success is achieved (see Figure 2.1) and thus a stage is completed.

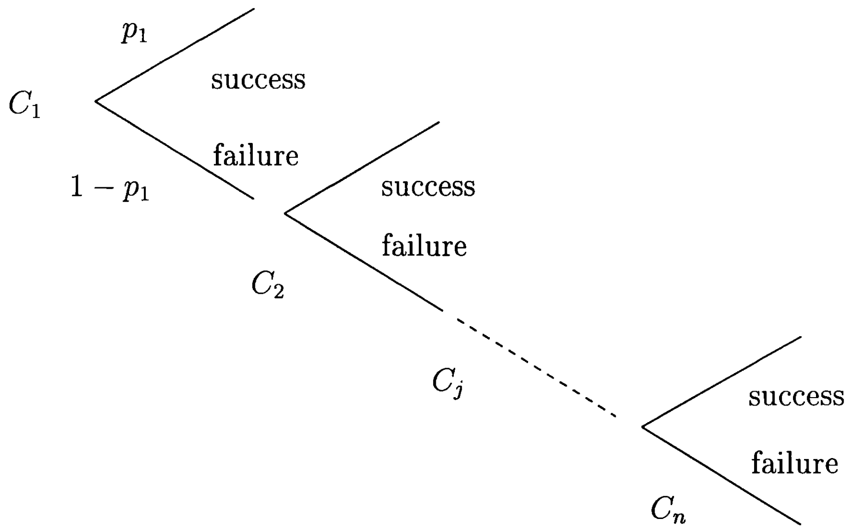

Figure 2.1: Decision tree for parallel system of alternatives

In each stage, the decision process continues until one of the alternatives is completed successfully. Given the probabilities of success p_j of the general stage j for $j = 1, \dots n$, the probability of success of the system is given by

$$\text{Probability of completing the procedure} = 1 - \prod_{j=1}^{n} (1 - p_j).$$

2.2.2 Expected cost structure of different strategies

Let $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ denote an ordering strategy, where $\pi_i \in \{1, 2, \dots, n\}$ for $i = 1, 2, \dots, n$, and when $\pi_j = i$, the ordering strategy π requires the decision maker to undertake task i at the j^{th} stage.

The expected total cost $\mathrm{E}(c(\pi))$ in a series system is given by

$$\mathrm{E}(c(\pi)) = c_{\pi_1} + \sum_{i=2}^{n} \left[\prod_{j=1}^{i-1} p_{\pi_j} \right] c_{\pi_i}$$

which is dependent on the order $\pi = (\pi_1, \pi_2 \ldots, \pi_n)$ of tasks where p_{π_j} and c_{π_j} are the probability of success and the cost of task s. It is not difficult to prove based on interchange argument that the expected cost is minimized if and only if the order of tasks is arranged in the order of their increasing ratios $c_s/(1-p_s)$ *i.e.,* start with the task with lowest ratio (Ben-Dov, 1981)

$$\frac{c_1}{1-p_1} \leq \frac{c_2}{1-p_2} \leq \ldots \leq \frac{c_n}{1-p_n}. \tag{2.1}$$

To show how the interchange argument yields ordering relations, we have:

Let the expected cost of ordering strategy $\pi = (\pi_1, \pi_2 \ldots, \pi_n)$ denoted by $E\{c(\pi_1, \pi_2 \ldots, \pi_n)\}$ be

$$\mathrm{E}\{c(1,2,3\ldots,n)\} = c_1 + c_2 p_1 + c_3 p_1 p_2 + \ldots + c_n p_1 p_2 \ldots p_{n-1}.$$

Now for the expected cost for ordering strategy $\pi_1 = 2, \pi_2 = 1, \pi_3 = 3, \ldots \pi_n = n$

$$\mathrm{E}\{c(2,1,3\ldots,n)\} = c_2 + c_1 p_2 + c_3 p_1 p_2 + \ldots + c_n p_1 p_2 \ldots p_{n-1}.$$

Suppose that the least expensive sequence is the first one. Then, we have:

$$\mathrm{E}\{c(1,2,3\ldots,n)\} \leq \mathrm{E}\{c(2,1,3\ldots,n)\}.$$

Therefore,

$$\mathrm{E}\{c(2,1,3\ldots,n)\} - \mathrm{E}\{c(1,2,3\ldots,n)\} = c_2 + c_1 p_2 - c_1 - c_2 p_1 \geq 0. \tag{2.2}$$

From (2.2) we get:

$$\frac{c_1}{1-p_1} \leq \frac{c_2}{1-p_2}. \tag{2.3}$$

Now suppose we interchange any two adjacent elements m and $m+1$ of the original sequence. Similarly we get

$$\begin{aligned}\mathrm{E}\{c(1,2,\ldots,m-1,m+1,m,m+2\ldots n)\} - \mathrm{E}\{c(1,2,3\ldots,n)\} &= \\ = p_1 p_2 \ldots p_{m-1}(c_{m+1} + c_m p_{m+1} - c_m - c_{m+1} p_m) \geq 0.\end{aligned} \tag{2.4}$$

From (2.4) we get:

$$\frac{c_m}{1-p_m} \leq \frac{c_{m+1}}{1-p_{m+1}} \qquad m = 1, \ldots, n-1. \tag{2.5}$$

Clearly, if (2.5) holds then minimum expected cost occurs when the tasks are arranged in the order of their increasing ratios $c_s/(1-p_s)$ ratios.

Alternatively, the expected cost $\mathrm{E}(c(\pi))$ for a parallel systems of alternatives (Ben-Dov, 1981) is given by:

$$\mathrm{E}(c(\pi)) = c_{\pi_1} + \sum_{i=2}^{n} \left[\prod_{j=1}^{i-1} (1 - p_{\pi_j}) \right] c_{\pi_i}.$$

The cost is minimized if and only if tasks are chosen with the following criterion,

$$\frac{c_1}{p_1} \leq \frac{c_2}{p_2} \leq \ldots \leq \frac{c_n}{p_n}, \tag{2.6}$$

that is, the tasks are ordered by increasing values of the above index. Comparing results (2.1) and (2.6) it is concluded that the minimization of the expected cost in parallel and sequential scenarios is the ratio of the cost to the probability of success or the ratio of cost to the probability of failure of the task.

Therefore a ratio criterion gives the optimal solution to optimization problems with either series or parallel system structure.

2.2.3 Existence of optimal priority index sequencing rules

In this section, examples of sequencing rules are presented. Suppose that there are n alternatives tasks or jobs. There are $n!$ possible ordering strategies. Parameters which might be taken into account are the probability of successful completion of each stage, the completion time, the cost of the stage, the discounting factor and the reward obtained upon the successfully completion of the stage.

Job Sequencing Problem

In the job scheduling problem (Walrand, 1988), there are n jobs that require independent random service times $S_1, \ldots, S_n$, respectively. A single server processes these jobs, one at a time, non-preemptively. That is, once a job starts, it cannot be interrupted until completion. The parameters associated with each job j are the processing time t_j, that is the time at which the processing of the job j is completed $(1 \leq j \leq n)$ and the delay cost rate c_j which is paid for time interval $0 \leq t \leq t_j$ for each $j = 1, \ldots, n$. The cost of the job j is $c_j \times t_j$ where c_j is a positive constant. The sum $\sum_j c_j t_j$ is called total weighted flow time.

The problem is to find in which sequence the jobs should be processed so as to minimize the expected weighted flowtime

$$\mathrm{E}\left\{\sum_j c_j t_j\right\}.$$

One should note that the value of the expected weighted flowtime depends on the order in which jobs are processed through the t_j.

It can be shown based on interchange argument that the total expected weighted flow time $\sum_j c_j t_j$ is minimized if the jobs are processed in decreasing order of the following index δ_j,

$$\delta_j = \frac{c_j}{\mathrm{E}(S_j)}.$$

Rothkopf (1966) incorporated a continuous discount rate factor β $(0 < \beta < 1)$, for the cost c_i paid at future time t. Thus the present value of cost c_i paid at future time t is $c_j \exp(-\beta t_j)$, and he derived the following priority index as an extension to the above index:

$$\delta_j = \frac{c_j \beta e^{-\beta t_j}}{1 - e^{-\beta t_j}}.$$

Rothkopf and Smith (1984) proved that there are no undiscovered priority index solutions to the job sequencing problem. The only two cases are:

1. the delay cost function is linear $k_j + c_j t$, *i.e.*, the cost of delaying a task is proportional to the length of the delay,

2. exponential delay cost function $k_j + c_j\{1 - \exp(-rt)\}/r$,

where c_j indicates the cost of deferring the completion of task j until time $t > 0$, r is the discounting rate and k_j is the constant cost for j^{th} task, for all $j = 1, \ldots n$.

Discrete Search Problem

In the discrete search problem (DeGroot, 1970), an object is hidden in one of n possible locations. We assume that the prior probability that the object is at the j^{th} location is p_j ($\sum p_i = 1$). We also denote by α_j the probability that the object will not be found in a particular search of the location j even though the object is actually in j^{th} location. This probability is called the overlook probability and remains the same for every search of location j. A cost c_j must be paid for the search of location j.

The objective is to minimize the expected search cost of the strategy until the object is found. The solution to the discrete search problem is to examine the locations in decreasing order of the index

$$\delta_j = \frac{p_j(1-\alpha_j)}{c_j}.$$

The cost c_j can be replaced by t_j, the time taken to search the location j if the objective is to minimize the time spent to identify the hidden item rather than the total expected search cost.

R&D Project Selection Problem

According to Chun (1994), Dean (1966) appears to be the first who considered the optimality of sequencing strategies in terms of the development cost c_j and the probability of success p_j. He considered a series system model and he derived the index

$$\delta_j = \frac{1-p_j}{c_j},$$

which is the same result as the index (2.1).

Joyce (1971) examined a similar problem in which he assumed that a research project consists of several sub-projects. Each sub-project must be successfully completed for the total research project. Assume that there are n alternative approaches for task i denoted by $\pi_{i,1}; \pi_{i,2}; \ldots ; \pi_{i,n}$. Let $c_{i,j}$ be the cost of the j^{th} alternative of the i^{th} task and $p_{i,j}$ is the probability of its success. Let c_i be the cost of carrying out the alternatives in the order $\pi_{i,1}; \pi_{i,2}; \ldots ; \pi_{i,n}$ until one succeeds or until all have been tried. The expected cost c_i of the i^{th} task is minimized if its alternatives are pursued in the order of increasing value of their $c_{i,j}/p_{i,j}$ ratios, that is if

$$\frac{c_{i,j}}{p_{i,j}} < \frac{c_{i,j+1}}{p_{i,j+1}}, \quad j = 1, \dots, n-1.$$

then c_i is minimized.

Kwan and Yuan (1988) proposed an index for a parallel system of investment alternatives, that is, there are n mutually exclusive projects and project j is chosen. Let x_j be the reward if the project is successfully completed and y_j is the benefit otherwise. The probability of success of the project j is equal to p_j. Projects are chosen sequentially (see Figure 2.2), and the objective is to maximize the expected net present value of cashflow of the chosen project. Projects are chosen in decreasing order of

$$\delta_j = (x_j - y_j) - \frac{c_j - y_j}{p_j}.$$

This is an index for sequential choice of projects which denotes which project should be undertaken first.

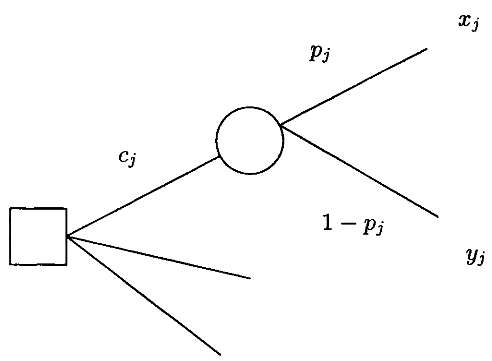

Figure 2.2: Sequential choice of projects

Series system of tasks

Chun (1994) considered a series system of tasks in the R&D project selection problem with ordering strategy $\pi = \{\pi_1, \pi_2, \cdots, \pi_n\}$. His objective was to derive a priority index in order to find the optimal ordering strategy which maximizes the expected discounted reward $T(\pi)$ which, in this case, is equal to,

$$T(\pi) = \sum_{i=1}^{n} \left[\prod_{j=0}^{i-1} p_{\pi_j} (1+\beta)^{-t_{\pi_j}} \right] R(\pi_i),$$

where the empty product is taken to be one, the reward $R(\pi_i)$ is gained if stage i is completed successfully, with discounting factor β and time taken to complete the stage j, is t_{π_j}, then, the priority index is,

$$\delta_i = \frac{R(\pi_i)}{1 - p_i(1+\beta)^{-t_i}} \, .$$

The optimal policy chooses projects in descending order of the above index.

Precedence Restriction

Now, suppose that there are two tasks j and k in a series system of tasks and that task j must precede task k. If, also, the task k directly follows task j then the two tasks can be thought of as a single joint task i as far as the ordering is concerned. Chun (1994) proved that the ordering index of the single joint task i is given as follows:

$$\delta_i = \frac{R(\pi_j) + p_j(1+\beta)^{-t_j} R(\pi_k)}{1 - p_j p_k (1+\beta)^{-(t_j+t_k)}} \, ,$$

and projects are chosen in the descending order of δ_i.

Parallel System of Alternatives

Chun (1994) studied the problem for parallel system of alternatives with objective to maximize the expected discounted reward $T(\pi)$, which is equal to

$$T(\pi) = \sum_{i=1}^{n} \left[\prod_{j=0}^{i-1} (1 - p_{\pi_j})(1+\beta)^{-t_{\pi_j}} \right] R(\pi_i).$$

The priority index related to this optimization problem is given by

$$\delta_i = \frac{p_i R(\pi_i)(1+\beta)^{-t_i}}{1 - (1-p_i)(1+\beta)^{-t_i}} \qquad i = 1, \dots, n, \tag{2.7}$$

where the optimal strategy is to perform tasks in descending order of the above index.

Parallel system of alternatives occurs when several projects are chosen sequentially, one at each time until the first success is achieved. The expected present value of reward for any selection strategy π is expressed as follows

$$T(\pi) = \sum_{i=1}^{n} \left[\prod_{j=0}^{i-1} (1 - p_{\pi_j}) \right] [p_{\pi_i} R(\pi_i) - c_i] \exp \left[-\beta \sum_{j=1}^{i} t_{\pi_j} \right],$$

where $p_{\pi_0} = 0$.

Chun and Platt (1992) showed that the appropriate priority index is

$$\delta_i = \frac{p_{\pi_i} R(\pi_i) - c_i}{p_{\pi_i} + \dfrac{1 - \exp(-\beta t_i)}{\exp(-\beta t_i)}} \qquad i = 1, \dots, n. \tag{2.8}$$

Priority index (2.8) can be derived from (2.7) when $(1+\beta)^{-t_i}$ and $R(\pi_i)$ are substituted by $\exp(-\beta t_i)$ and $\{p_{\pi_i} R(\pi_i) - c_i\}$ respectively.

The above example is very interesting because it shows us how the functional form of priority index changes when the discounting factor β is taken into the account. It could be the case that the priority index will give different ordering results if undiscounted values for rewards and costs are used compared to the ordering strategy one would get, if discounted values are used and the discounting factor is equal to zero.

2.3 Capital Budgeting in a firm

In the previous section we discussed how priority indices can solve parallel and sequential selection problems. In this section an account is given how project selection problems are formulated using a mathematical programming approach, with the main emphasis on linear programming .

A firm has available an amount C of investment capital. There are n different projects which are competing for the available funds. Project j requires an investment of c_j and yields future profit which has present value R_j. A project must either be taken (lincenced out) or not. We introduce a binary variable y_j for each project j to denote whether the project j is selected or not. Let

$$y_j = \begin{cases} 1 & \text{if project } j \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

The objective is to maximize the total discounted return $Z = \sum_{j=1}^{n} y_j R_j$ and spend exactly an amount C. Furthermore, $c_j \geq 0$ and $R_j \geq 0$ for every $j = 1 \ldots n$. The solution to this project selection problem is determined by the following integer linear programming problem

$$\text{Maximize} \quad \sum_{j=1}^{n} y_j R_j \tag{2.9}$$

$$\text{subject to} \quad \sum_{j=1}^{n} y_j c_j = C, \tag{2.10}$$

$$y_j \in \{0,1\} \qquad \text{for} \quad j = 1, \cdots, n. \tag{2.11}$$

However, trying to maximize the objective function with the equality constraint is not only unrealistic but it may be an infeasible problem if no combination of the available projects exist such that their total cost is equal to C.

Let us forget the constraint (2.11) and consider the linear programming problem:

$$\begin{aligned} \text{Maximize} \quad & \sum_{j=1}^{n} y_j R_j \\ \textit{subject to} \quad & \sum_{j=1}^{n} y_j c_j = C, \end{aligned}$$

Consider also its dual program, that is

$$\text{Minimize} \quad Cx$$

$$\text{subject to} \quad x c_j \geq R_j, \quad j = 1, \cdots, n \tag{2.12}$$

where x denotes the achievable profit per unit cost. From (2.12), one can conclude that $x \geq R_j/c_j$ for all j and therefore the value of the objective function $U(C)$ at the optimum point as a function of the available capital C is given by

$$U(C) = \max_j \left\{ \frac{R_j}{c_j} \right\} C.$$

A different approach to the capital budgeting problem is to reformulate the above problem with inequality constraint $\sum_{j=1}^{n} y_j c_j \leq C$.

Being able to spend less than the available capital, that is the budget constraint becomes an inequality, the optimal solution which maximizes the benefit might be such that the projects chosen did not exhaust all the available resources. However, if, the projects can be chosen in any rate between zero and one, which means that y_j is not restricted to values 0 and 1 only, then the optimum solution spends all the available funds if the rewards gained are increasing function with respect to the intensities of the available projects. The mathematical formulation is a linear programming problem with bounds

$$\begin{array}{lll} \text{Maximize} & \sum_{j=1}^{n} y_j R_j & \\ \textit{subject to} & \sum_{j=1}^{n} y_j c_j \leq C & j = 1, \cdots, n \\ & 0 \leq y_j \leq 1 & j = 1, \cdots, n \end{array}$$

The solution is given by comparing the ratio R_j/c_j and 1, since y_j must be bounded by 1.

Multiple stage projects

Suppose, now, that each project has k multiple stages, which must be implemented in some order with success and the reward is gained at the end of the last stage. Let c_{ij} be the capital required for project j in period i. The indicator function for selecting project j is y_j. Now, we introduce an indicator function for each stage of a given project. Let y_{ij} be the binary variable which is 1 when the stage i of the project j has been implemented and 0 otherwise. To consider the time sequencing for flow of funds in each project, consider the following mathematical program. To maximize the profit

$$\begin{array}{lll} \text{Maximize} & \sum_{j=1}^{n} y_j R_j & \\ \text{subject to} & \sum_{j=1}^{n} \sum_{i=1}^{k} y_{ij} c_{ij} \leq C & \\ & y_{ij} \geq y_j & i = 1, \cdots, k \quad j = 1, \cdots, n \\ & y_{ij} = 0 \text{ or } 1, & i = 1, \cdots, k \quad j = 1, \cdots, n \\ & y_j = 0 \text{ or } 1 & j = 1, \cdots, n. \end{array}$$

This formulation requires that for a given project j whose indicator function is y_j, its reward R_j is gained if all the stages i happen with total cost $\sum_{i=1}^{k} c_{ij}$. A project j is chosen, if only if all stages can be funded as it is required by constraint $y_{ij} \geq y_j$. If not all the stages i happen for a given project j, then some y_{ij} are equal to zero and therefore $y_j = 0$, that is the reward R_j is not gained.

Groups of projects

Consider the situation where there is a set S of candidate projects, and each project consists of K stages. S is partitioned into disjoint subsets $S_1, S_2, \cdots, S_K$, where the subset S_j are all the projects that their first $j-1$ stages have been implemented. Let us suppose that there is a collection of different budgets with amount C_j for $j = 1, \ldots K$ where C_j is the budget investment for all projects which their next stage to be implemented is the j^{th} stage. To maximize the profit, we have:

$$\begin{array}{lll} \text{Maximize} & \displaystyle\sum_{j=1}^{n} y_j R_j & \\ \text{subject to} & \displaystyle\sum_{j=1}^{n}\sum_{i=1}^{k} y_{ij} c_{ij} \leq C & \\ & \displaystyle\sum_{j=1}^{n} y_{ij} c_{ij} \leq C_i & i = 1, \cdots, k \\ & y_{ij} \geq y_j & i = 1, \cdots, k \quad j = 1, \cdots, n \\ & y_{ij} = 0 \,\text{or}\, 1 & i = 1, \cdots, k \quad j = 1, \cdots, n \\ & y_j = 0 \,\text{or}\, 1 & j = 1, \cdots, n. \end{array}$$

The first constraint says that the total budget available is equal to C. The second constraint $\sum_{j \in S_k} y_{ij} c_{ij} \leq C_i$ determines the budgets of each subset(category) of projects denoted by S_i for all $i = 1, \ldots, K$.

It can be easily seen that in every different project selection scenario, a different linear programming problem may be constructed. However, people in practice found it more convenient to have a basic approach a productivity index, namely, the Pearson index (Regan & Senn, 1997).

2.4 The Pearson Index

The Pearson Index (Pearson, 1972) was suggested as a ranking formula in R&D projects by Pearson in 1972. It was argued that all the indices existing before Pearson were giving misleading results because they failed to take into account the multistage characteristics of R&D projects. Therefore, it seemed to be necessary that the new index would be based on a decision tree type approach, *i.e.,* by the technique of backward induction (see, *e.g.,* Raiffa 1968), we calculate the value of the project at the first decision of the tree. Examples of indices (see, Pearson 1972, p. 69) before Pearson index are:

$$\text{Index 1} = \frac{P_t \times P_c \times (p - c) \times V \times L}{\text{Total Cost}} \quad \text{and} \tag{2.13}$$

$$\text{Index 2} = \frac{P_t \times P_c \times \frac{I_1}{(1+i)} + \frac{I_2}{(1+i)^2} + \cdots + \frac{I_n}{(1+i)^n}}{\text{Total discounted R\&D Cost}}, \tag{2.14}$$

where P_t and P_c are the probability of technical and commercial success respectively, p and c are the price and the cost, V and L are the sales volume per year and the life of the product respectively, $(1 + i)$ is the discounting factor and I_n refers to the net income in the n^{th} year of the project's life.

To explain the definition of the Pearson Index, consider the following Figure (2.3). The general form of the Pearson index is the ratio of the expected net

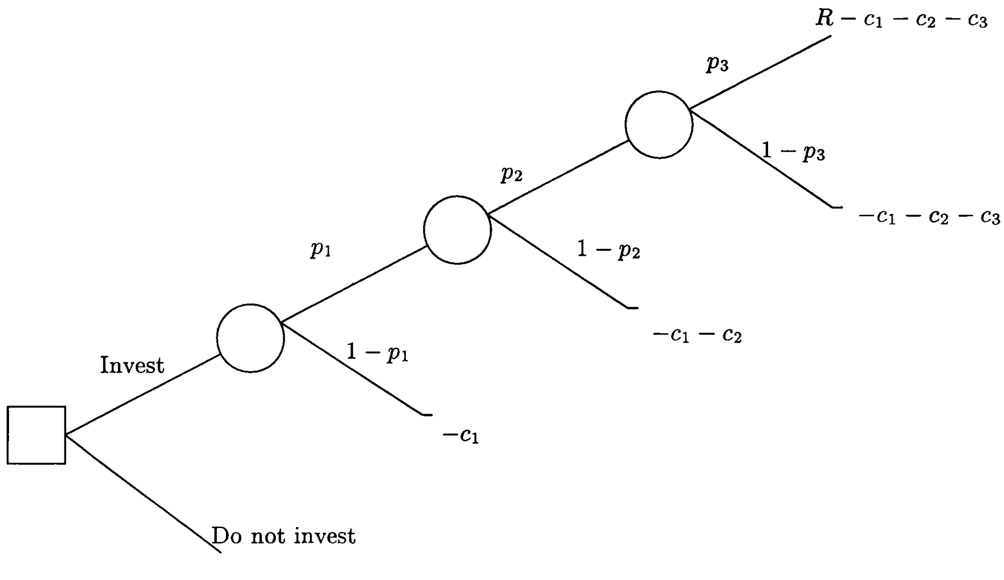

Figure 2.3: Pearson Index decision tree for a three stage project

reward to the expected development cost. Consider a project with n stages with fixed order, where stage i has cost c_i, probability of success p_i given success at the previous stages and, if successful at all stages, final reward R. Thus, the decision to invest to an R&D project consists of a series of cash inflows and outflows. The *Net Present Value* (NPV) rule of an investment is the difference between the net present value of all cash inflows rate, and, the net present value of all cash outflows, everything discounted at some interest rate. Assuming that the reward R all costs c_i for $i = 1, \dots, n$ are discounted and applying the NPV rule that the expected value of the decision to invest is equal to the expected value of the benefits less the expected value of the cost, we get:

$$\text{Expected net reward} = R\prod_{i=1}^{n} p_i - \sum_{i=1}^{n} c_i \prod_{j=0}^{i-1} p_j.$$

The expected cost of the decision to invest is:

$$\text{Expected cost} = \sum_{i=1}^{n} c_i \prod_{j=0}^{i-1} p_j.$$

Pearson Index is defined to be:

$$\frac{R\prod_{i=1}^{n} p_i - \sum_{i=1}^{n} c_i \prod_{j=0}^{i-1} p_j}{\sum_{i=1}^{n} c_i \prod_{j=0}^{i-1} p_j} \tag{2.15}$$

where $p_0 = 1$. Its meaning is the expected net reward per unit expected cost.

Suppose that one has to value the decision to invest in an n-stage project in which the first $k-1$ stages have been implemented. We denote the Pearson index by $P_n(i)$ for $i = 1, \dots, n-1$, where i indicates the next stage to be run

$$P_n(k) = \frac{R\prod_{i=k}^{n} p_i - c_k - \sum_{i=k+1}^{n} c_i \prod_{j=k}^{i-1} p_j}{c_k + \sum_{i=k+1}^{n} c_i \prod_{j=k}^{i-1} p_j}$$

An important property of Pearson Index is its higher ratio property, that is, $P_n(i) \geq P_n(j)$ for any $i \geq j$.

Pearson concluded that:

1. Simple rank indices will inevitably lead to a "bias" in the calculation of the expectation from a particular project.

2. The "bias" will increase as the degree of uncertainty of the initial stages of the project increases.

3. The "bias" can be removed by the use of a modified form of ranking index based on a decision tree type analysis.

Pearson did not explain why his index should have a ratio form and did not give any definition of the term "bias". Also, he did not account at all for its higher ratio property and how it is related to its portfolio selection problem. An account for these two issues will be given later. The Pearson index solves a selection problem subject to linear constraint of the type of Knapsack problem (Martello & Toth, 1997).

Pearson derived his index by folding backwards a simple decision tree (see Figure 2.3) and as a result he derived an index which is based on a net present value rule. It is well known that NPV rules and other discounted cash flow techniques for capital budgeting may be inappropriate to build a portfolio of research because they may favour short term projects in relatively certain markets over long term and relatively uncertain markets. This will not happen in a dynamical model in which uncertainty unravels over time creating flexibility for decision makers, who behave optimally (in some sense) at each point in time.

Flexibility has a value which should be quantified. This flexibility is due to variability of future reward and costs. Stochastic dynamic programming is a possible solution to the problem of quantification of flexibility. In the Pearson index case, one should note that there is variability in rewards and costs. The probabilities of the Pearson index refer to technical success and not to the uncertain future payoffs of the R&D project.

In order to build an appropriate framework to value such risky R&D projects, a model is essential to give us an idea how the information about the future value of the project evolves. The framework provided by the theory of real options can price a project in a better way a than the Pearson index.

Real Option is the application of Financial option theory applied to real in-

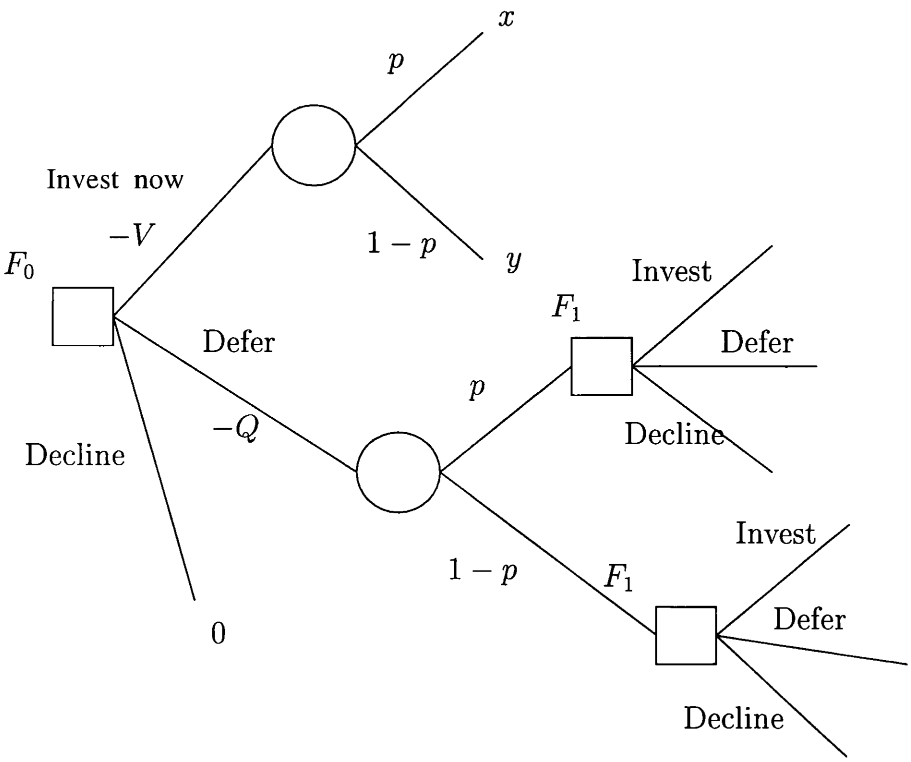

Figure 2.4: Decision tree with option value

vestment such as the valuation and management of an undeveloped oil field or an R&D project. Given an optimal way to make decisions at some future time t, the decision maker makes a decision in such way that the future decisions will be made in the given optimal way.

Consider the Figure (2.4). The decision maker has three possible choices. These are to "invest now", "defer" the investment and "decline" the investment. The decision to "invest now" yields expected reward $px + qy$, where p is the probability to get reward x and $q\,(= 1 - p)$ is the probability to get reward y. This decision has cost V. Denote by Ω_0 the value associated with the decision to invest now or never, *i.e.*, the option to delay the investment decision whether to invest or not is not available. We set

$$\Omega_0 = \max\left(px + qy - V, 0\right),$$

that is, the maximum between the net present values of the two decisions, namely, the decision to invest now and the decision to reject the project. Denote by F_0, the value of the opportunity to invest now at time $t = 0$ or later at time $t = 1$.

The "defer" option costs Q. We set

$$\begin{aligned} F_0 &= \max(\text{NPV of the decision to invest now}, \text{NPV of the decision to invest later}) \\ &= \max\left\{px + qy - V, \frac{1}{(1+r)}\mathrm{E}_0(F_1) - Q\right\}, \end{aligned}$$

where F_1 is the random variable which denotes the value of the decision of type "invest or not" at time $t = 1$ and $\mathrm{E}_0(F_1)$ is the expected value of F_1 given the information at time $t = 0$. All future values are discounted at rate r. We define the random variable $W_0 = F_0 - \Omega_0$ to be the option to postpone the decision to "invest now" or never. The option to postpone the decision to invest has value:

$$W_0 = \begin{cases} 0 & \text{if } px + qy - V \geq \frac{1}{(1+r)}\mathrm{E}_0(F_1) - Q \geq 0 \\ \frac{\mathrm{E}_0(F_1)}{(1+r)} - Q - (px + qy - V) & \text{if } \frac{1}{(1+r)}\mathrm{E}_0(F_1) - Q \geq px + qy - V \\ \frac{1}{(1+r)}\mathrm{E}_0(F_1) - Q & \text{otherwise.} \end{cases}$$

The weakness of the Pearson Index which is inherent in standard NPV approach is in the treatment of the contingent cash flow and values arising from implicit or explicit 'options' which arise as a project evolves. The option to postpone the decision to invest is a random variable and it depends on the level on uncertainty and how it is resolved in the future.

To discuss in more depth the concept of options, a sequential decision framework is needed to model the uncertainty. For the analysis of the next section, the concept of contingent cash flows is ignored, in order to study which problem the Pearson Index solves. In the next section, the relation of the knapsack problem and the Pearson Index is studied.

2.4.1 The knapsack problem and the Pearson index

The classical knapsack problem is to pack a knapsack of integer volume V with objects from K different classes in order to maximize profit. There are K different classes, $j = 1, \cdots, K$, and each object from a given class j, consumes c_j integer units of the knapsack and produces profit P_j. We also assume that the class j consists of b_j items ($j = 1, \ldots, K$).

The problem has a simple solution: fill the knapsack entirely, if possible, with objects from class j that has the highest profit to volume ratio P_j/c_j. If the knapsack volume ratio is not an integer multiple of the object volumes, then the problem can still be solved with dynamic programming. In terms of linear programming, the Knapsack problem is formulated as follows:

Let y_j denote the number of items selected from j^{th} class. The $0-1$ Knapsack problem is

$$\begin{array}{lll} \text{Maximize} & \displaystyle\sum_{j=1}^{K} P_j y_j & \\ \textit{subject to} & \displaystyle\sum_{j=1}^{K} c_j y_j \leq V, & \\ & y_j \in \{0,1\} & j = 1, \cdots, K, \end{array}$$

where, y_j is 1 if j object is selected and 0 otherwise. Suppose now, that one can select up to b_j items for the j^{th} class. Then, we have

$$\begin{array}{lll} \text{Maximize} & \displaystyle\sum_{j=1}^{K} P_j y_j & \\ \textit{subject to} & \displaystyle\sum_{j=1}^{K} c_j y_j \leq V, & \\ & 0 \leq y_j \leq b_j & j = 1, \cdots, K, \\ & y_j \quad \text{integer} & j = 1, \cdots, K, \end{array}$$

and is known as a bounded knapsack problem. We can now relax the constraint that $y_j\,(j = 1, \ldots, K)$ are integers and have the continuous version of the Knapsack problem, *i.e.*, we are allowed to choose items partially, that is:

$$\begin{array}{lll} \text{Maximize} & \displaystyle\sum_{j=1}^{K} P_j y_j & \\ \textit{subject to} & \displaystyle\sum_{j=1}^{K} c_j y_j \leq V, & \\ & 0 \leq y_j \leq 1 & j = 1, \cdots, K. \end{array}$$

Its solution is given by:

Order the items according to decreasing values of their ratios, namely, profit per weight. The items are inserted consecutively until the first item, s, is found which does not fit. This is called the *critical item* $s = min\{i : \sum_{j=1}^{i} c_j \geq V\}$. Then the optimal solution is given by:

$$\begin{array}{lll} \overline{y}_j = 1 & \text{for} & j = 1, \cdots, s-1 \\ \overline{y}_j = 0 & \text{for} & j = s+1, \cdots, K \\ \overline{y}_s = \frac{\overline{V}}{c_s} & \text{where} & \overline{V} = V - \sum_{j=1}^{s-1} c_j. \end{array}$$

The ranking property gives the solution to the selection problem given the linear equality constraint. It is obvious that all the above wholly or partially chosen projects, have equal or greater ratio than the critical item. This property can be found in any mathematical program of the form (Zipkin, 1980),

$$\begin{array}{llll} Maximize & \sum_{j=1}^{n} R_j(y_j) & & \\ subject\ to & \sum_{j=1}^{n} y_j \leq B & y_j \geq 0 & j = 1, \cdots, n \end{array}$$

provided that, $B \geq 0$, and each $R_i : \mathbf{R} \to \mathbf{R}$ is continuously differentiable and strictly concave.

The Pearson index can be thought of as a critical ratio for the selection problem. If one wants to select a subset of the available projects in a way which maximises the total expected net return of the selected projects given a fixed total expected cost of the selected projects, then, this can be achieved with an index of the form

$$\frac{E(NetReward)}{E(cost)} \geq \lambda \tag{2.16}$$

where λ is constant.

2.4.2 Neyman-Pearson Lemma

A similar constrained maximization problem occurs in the Neyman-Pearson theory of hypothesis testing (Berger, 1980).

A test of a statistical hypothesis is based on the evidence of the observed data x. A decision has to be made whether to reject H_0, the null hypothesis under consideration, or to accept it. Rejecting H_0 when it is true is called a Type I error, and not rejecting H_0 when it is false is called a Type II error. The sample space S can be partitioned into two regions, W and $S - W$, such that whenever

x falls in W(critical region), the null hypothesis H_0 is rejected, and whenever x is in region $S - W$, H_0 is not rejected.

For a simple hypothesis H_0, let $P(W|H_0)$ be the probability of committing a type I error. For a simple alternative hypothesis H_1, let $P(S - W|H_1)$ be the probability of committing a type II error. The power of this test is $1 - \beta = P(W|H_1)$, that is, the probability of accepting the alternative hypothesis H_1 when it is true.

The Neyman-Pearson problem can be stated as:

$$\begin{aligned} &\text{Maximize} \quad 1-\beta = 1 - P(S - W|H_1) \\ &\text{subject to} \quad P(W|H_0) = \alpha \text{ (a given value)}. \end{aligned} \tag{2.17}$$

W^* is optimal for this problem if there exists a real number k such that

$$W^* = \{x|P(x|H_1) \geq kP(x|H_0)\} \text{ and } P(w^*|H_0) = \alpha. \tag{2.18}$$

In a randomized test, the probability of the rejecting H_0 is specified, on the basis of the observed data. In general, the sample space S is partitioned into three non-overlapping regions, W_1, W_2 and $S - W_1 - W_2$.

The decision rule is:

Reject H_0 if $x \in W_1$

Accept H_0 if $x \in W_2$

Reject H_0 with probability $\Phi(x)$ $(0 \leq \Phi(x) \leq 1)$ if $x \in S - W_1 - W_2$.

In statistical decision theory one tends to think in terms of losses rather than gains and the above problem is reformulated as follows:

$$\text{Minimize } \alpha + \lambda\beta = \int \phi(x)[\lambda f_0(x) + f_1(x)]dx$$

for fixed $\lambda > 0$ and $0 \leq \phi \leq 1$. This is solved by using the likelihood ratio test with "cut-off" λ: i.e. reject H_0 if $\frac{f_1(x)}{f_0(x)} > \lambda$, accept H_0 if $\frac{f_0(x)}{f_1(x)} > \lambda$ and be indifferent if $\frac{f_0(x)}{f_1(x)} = \lambda$.

Therefore, the three possible decisions are:

$$\begin{aligned} \Phi(x) &= 0 && \text{if } \lambda f_0(x) > f_1(x) \\ \Phi(x) &= 1 && \text{if } \lambda f_0(x) < f_1(x) \\ \Phi(x) &= \gamma(x) && \text{if } \lambda f_0(x) = f_1(x) \end{aligned}$$

for some $\lambda \geq 0$ and $0 \leq \gamma(x) \leq 1$. If $\lambda f_0(x) = f_1(x)$ has positive probability, this is not unique. Choose any such rule δ^*, it will have some size α^* and power $1 - \beta^*$. Let δ be any rule at all, with size α and power β. Then, it is concluded that

$$\alpha^* + \lambda\beta^* \leq \alpha + \lambda\beta,$$

which is

$$\lambda\beta^* \geq (\alpha^* - \alpha) + \lambda\beta.$$

Now, if $\alpha \leq \alpha^*$, $\beta^* \geq \beta$ then δ^* solves the problem of maximizing the power for size equal or less than α^*.

2.5 Sequential Decision Processes

In a sequential decision problem the decision maker(Statistician) is looking at a sequence of observations one at a time and he has to decide after each observation whether to stop sampling and take an action immediately or continue sampling and postpone taking action to some later time.

A sequential decision function has two components, namely, a *sampling plan* (or stopping rule) and *a decision rule.* To explain these two terms, we shall assume that the distribution of the sequence $\mathbf{x}$ of observations depends on a parameter W whose values are in a parameter space Ω. A decision space $\mathcal{D}$ consists of all possible decisions d which might be made by the statistician. A loss function $\mathbf{L}$ is defined on the product space $\Omega \times \mathcal{D}$, that is $\mathbf{L} = L(w, d)$ and represents the loss for any point $(w, d) \in \Omega \times \mathcal{D}$ when the value of the parameter W is w and the statistician chooses decision d. In the sampling plan, the statistician specifies whether a decision in $\mathcal{D}$ should be chosen without any observation or whether an observation (at least one) should be taken. In case at least one observation is to be taken, the statistician specifies, given the sequence of observation, whether sampling should stop and a decision in $\mathcal{D}$ should be chosen or whether more observations should be taken. A decision rule $\delta(\mathbf{x})$ is specified by the statistician for each possible set of observed values $\mathbf{x}$.

We shall assume that, given a specific value for the parameter $W = w_i$, the observations for random variables $X_1, X_2 \dots$ are independent and identically distributed. Let the conditional p.d.f of each observation X_i when $W = w_i$ be $f(.|w_i)$. We denote by c_i the cost to be paid for observing the value of X_i.

For example, suppose that a sequential random sample $X_1, X_2, \dots$ can be taken from a Bernoulli distribution with unknown parameter W. We suppose that, $\Omega = \{w_1, w_2\}$ has just two points and that $D = \{d_1, d_2\}$ has two points. The loss function L is specified as follows:

$$\begin{aligned} L(w_1, d_1) &= L(w_2, d_2) = 0, \\ L(w_1, d_2) &= L(w_2, d_1) = b > 0. \end{aligned}$$

We suppose further that each observation costs 1 unit. The prior distribution of W is specified by $\xi = \Pr(W = w_1) = 1 - \Pr(W = w_2)$.

Under a sequential decision procedure, the total number of observations N that are taken before a decision in $\mathcal{D}$ is chosen is a random variable. The problem is to determine a sequential procedure that minimizes the expected terminal loss.

In the next sections, we shall consider sequential decision processes in which there are two choices at any stage. The decision maker may have to decide whether to continue experimenting or to terminate the process. If he decides to continue he may have to choose one of two or more random variables that are available at each stage. Random variables may represent experiments or another item of interest. The statistician can exercise some control over the distributions of the observations generated during the process and in this sense, over the distribution of his rewards and costs.

A discrete time sequential decision model

Consider a dynamic system evolving in discrete time according to the equation

$$x_{k+1} = f_k(x_k, u_k, \epsilon_k), \qquad k = 0, 1, \dots, N-1,$$

where x_k denotes the state of the system, that is the variable of interest, u_k a control input which determines the decision made after the most recent observations, and ϵ_k is a random variable. Suppose also that the function f_k are given and that x_k, u_k, ϵ_k are elements of the appropriate sets. The system operates over a finite number of stages N (finite horizon problem). We shall assume that ϵ_k is characterized by probability measure $p_k(.|x_k, u_k)$ defined on a collection of events in space which ϵ_k belongs to, and that ϵ_k depends on the current state x_k and control input u_k, but does not depend on the values of the prior uncertain parameter $\epsilon_0, \epsilon_1, \ldots, \epsilon_{k-1}$.

A characteristic of this process is that it is a *forward induction process* as opposed to the backward dynamic programming. A problem in which the notion of forward induction plays an important role is the Bandit problem.

2.5.1 Two-Armed Bandit Problem

One of the most interesting problems in stochastic control problem is the bandit problem. Consider two random variables X and Y. The distribution of X depends on the value of a parameter W_1 ($\omega_1 \in \Omega_1$) and the distribution of Y depends on the value of another parameter W_2 ($\omega_2 \in \Omega_2$). Also, suppose that the statistician will take a fixed number n of observations at most, sequentially. If at some stage i ($i < n$) the random variable X is chosen for observation, the conditional p.d.f when $W = w_i$, $f_X(.|w_i)$, (for $i = 1, 2$), are independent of the choices and outcomes of the observations in the previous stages. Let ξ be the prior joint distribution of the parameter W_1 and W_2. The statistician's concern is to find a sequential procedure that maximises the expected value of the sum of the n observations.

Let $V_n(\xi)$ denote this maximal expected sum of n observations. If the first observation is made on X, the expected sum of all n observations is $E\{X + V_{n-1}[\xi(X)]\}$ where $V_{n-1}[\xi(X)]$ is the expected sum of the remaining $n-1$ observations with $\xi(X)$ being the posterior joint distribution of W_1 and W_2. Similarly, if the first observation is made on Y, then the expected sum of the n observations is $E\{Y + V_{n-1}[\xi(Y)]\}$ with $V_{n-1}[\xi(Y)]$ to denote the sum of the $n-1$ observations when an optimal procedure is adapted. Then, the optimal sequential procedure must be the decision rule which maximises the expected reward, that is,

$$V_n(\xi) = \max\{E[X + V_{n-1}\{\xi(X)\}], E[Y + V_{n-1}\{\xi(Y)\}]\}$$

where the expectations are computed with respect to the prior distribution ξ.

Bandit problems appear to have been first proposed by Thompson (1933, 1935) in sequential analysis for determining which of two drugs is superior. Bellman (1956) introduced the discounted Bayesian setting. Bradt, Johnson, and Karlin (1956) gave us one of the early treatments of the bandit problem.

A generalization of the two-armed bandit problem is the N-armed bandit problem which has as a solution the Gittins index. The bandit problem is solvable only for infinite time horizon.

2.5.2 The Gittins Index

The multi-armed bandit problem is concerned with the question of how to dynamically allocate a single resource among several alternatives. A bandit problem in statistical decision theory consists of $N(N \geq 2)$ independent stochastic process which represent arms (projects, machines or treatments.) that can be pulled (chosen) in any order. Each time only one of these arms is selected. Each pull from a bandit process results in a random reward. The problem for the decision maker is to find the optimal strategy that maximizes the expected values of reward over an infinite time of horizon.

Bandit processes can be used to model problems where a sequence of choices has to be made between a collection of fixed alternatives, for example, the scheduling of jobs on a single machine and the design of sequential clinical trials, job search and labour market analysis in economics etc.

In general, the decision maker faces the conflict between taking those actions which yield immediate reward and those whose benefit will come only later. This is an important characteristic of the problem. If the long-term performance is important to the decision maker, not choosing an action which yields immediate benefit may be justified if the gain of extra information about the projects (jobs or treatments) is worth more than the immediate reward.

Suppose that there are N independent projects, each divisible into stages, and only one project can be worked on at each time. Project i has state $x_i(t)$ at time $t = 1, 2 \ldots,$ for $i = 1, 2, \ldots N$. At each time t, one must operate exactly one project. If project i is selected, it gives immediate reward $R_i(x_i(t))$, a function of the current state $x_i(t)$ of the chosen project i, and its state $x_i(t)$ then changes according to a stationary Markov transition rule. The states of the other projects remain frozen. The states of all projects are observed, and the problem is to schedule the order in which the projects are operated so as to maximize the expected present values of the sequence of immediate rewards, that is,

$$\mathrm{E}_\pi \left\{ \sum_{t=1}^{\infty} \alpha^t R(t) \right\}$$

where $0 < \alpha < 1$ is a fixed discounting factor, $R(t)$ denotes the reward gained at time t and π is the strategy used for choosing between projects. Gittins (1979; Gittins & Jones, 1974) proved that the solution to this problem is to associate to each project i an index $v_i(x_i(t))$, which is a function only of its state, and at each time operate the project with the largest current index. Gittins' index (see Gittins, 1979) has the following form. Its numerator is the expected discounted reward for a certain project up to the chosen stopping time τ, and its denominator is the expected discounted time up to the stopping time τ:

$$v_i(x_i) = \max_{\tau > 1} \frac{\mathrm{E}\left\{ \sum_{t=1}^{\tau-1} \alpha^t R_i(x_i(t)) \;\middle|\; x_i(1) = x_i \right\}}{\mathrm{E}\left\{ \sum_{t=1}^{\tau-1} \alpha^t \;\middle|\; x_i(1) = x_i \right\}}, \tag{2.19}$$

where the maximization is over the set of all stopping times $\tau > 1$. Gittins called his index a *Dynamic Allocation Index.* It is interpreted as the maximum expected discounted reward per unit of expected discounted time.

This index is sequential in the sense that it needs to be recalculated at each decision point in order to guide the reallocation process. The theory of Gittins' index, and more generally of any priority index, is based on the idea that the index for each project depends on the past history of the given (or chosen) project only and not on the history of other projects. Effort is allocated to the project with the highest current index value.

Bergman and Gittins (1985) refer to two different versions of dynamic allocation index related to Pearson Index. These are given by (2.20) and (2.21):

$$\frac{Rp_1p_2p_3 - c_1 - c_2p_1 - c_3p_1p_2}{1 - (1-p_1)D_1 - p_1(1-p_2)D_2 - p_1p_2D_3}. \tag{2.20}$$

The index given by (2.20) is for a three stage project, with final reward R, costs c_1, c_2, c_3 for stage one, two, and three respectively, probability of success of each stage p_1, p_2 and p_3. D_i denotes the present discounted value of one monetary unit at time t_i for $i = 1, 2, 3$. The numerator is the expected net profit and the denominator expresses the discounted probability that all stages of the project are implemented. The above can be approximated by the following index as the discounting rate approaches zero:

$$\frac{Rp_1p_2p_3 - c_1 - c_2p_1 - c_3p_1p_2}{t_1 + t_2p_1 + t_3p_1p_2}. \tag{2.21}$$

One should observe that the Pearson index (2.15) and the index given by (2.21) differ only in their denominator. However, there are differences between the Pearson and Gittins indices and a comparison is made below.

Comparison of Gittins and Pearson Index

The multi-armed bandit problem can be thought of as a mathematical programming problem. There are N statistically independent stochastic process. The reward gained from the i^{th} process is denoted by $R_i := R_i\{x(t)\}_{t=1}^{\infty}$, a bounded real stochastic process on $(\Omega, \mathcal{F})$. $\mathcal{F}^i = \{\mathcal{F}_t^i, t = 0, 1, \ldots, \infty\}$, the information process associated with arm i, is a non-decreasing family of sub σ-fields of $\mathcal{F}$ and Ω is the sample space of the reward process. We also assume independence of the N reward processes, and that the expected total discounted reward is finite, that is,

$$\mathrm{E}\sum_{t=1}^{\infty} \alpha^t |R_i(x_i(t))| < \infty, \qquad i = 1, 2, \ldots N.$$

Selecting a process, say i, results in a reward $R_i\{(x(t))\}$, a function of the state $x_i(t)$ at time t of the selected process. . The states of the other processes remain frozen and yield no reward. The objective of the decision maker is to find the spliced sequence which results in the maximum expected discounted reward. The problem (Ishikida & Varaiya, 1994) is as follows:

Problem 1

$$\underset{\Delta x(t)}{\text{Maximize}} \quad \mathrm{E}\sum_{i=1}^{N}\sum_{t=1}^{\infty}\alpha^t R_i(x_i(t))\Delta x_i(t)$$

$$\text{subject to} \quad \sum_{i=1}^{N}\Delta x_i(t) = 1, \qquad t = 1, 2, \ldots, \tag{2.22}$$

$$\Delta x_i(t) \in \{0, 1\}, \qquad i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, \tag{2.23}$$

$$\Delta x(t) \text{ is } \bigvee_{i=1}^{N}\mathcal{F}^i(x_i(t-1)) - \text{measurable}, \quad t = 1, 2, \ldots, \tag{2.24}$$

where the maximization is over $\Delta x(t) := (\Delta x_1(t), \Delta x_2(t), \ldots, \Delta x_N(t))$, and $\Delta x_i(t)$ takes value one if the project i is chosen at time t whose state is $x_i(t)$ and zero otherwise. Constraint (2.22) indicates that at any given time $t = 1, 2, \ldots,$ only one project is chosen; the next constraint (2.23) says that projects can not be chosen partially. Also, the indicator function $\Delta x(t)$ is dependent on or measurable with respect to all the available information about the all rewards of the projects up to most recent decision time $t-1$.

The Pearson index can have the following representation,

Problem 2

$$\textit{Maximize} \quad \sum_{j=1}^{n} y_j \mathrm{E}(R_j), \tag{2.25}$$

$$\textit{subject to} \quad \sum_{j=1}^{n} y_j \mathrm{E}(c_j) \le B \qquad y_j \ge 0 \qquad j = 1, \cdots, n. \tag{2.26}$$

where $\mathrm{E}(R_j)$ is the expected reward to be gained from project j and $\mathrm{E}(c_j)$ is the expected cost of the project j. As was mentioned earlier the solution to these mathematical programs are the Gittins and Pearson Indices, which have the following mathematical forms:

$$\max_{T>t} \frac{E[\text{Net discounted reward up to stopping time T}|\text{process state at time } t]}{E[\text{Discounted time up to stopping time T}|\text{process state at time } t]}$$

$$\text{Pearson Index} = \frac{E(\text{Net discounted reward }|\text{starting state})}{E(\text{cost}|\text{starting state})}.$$

One difference is that the Gittins index is a function of time whereas the Pearson index is not. The Multi-armed bandit problem, as an allocation problem, is concerned with the sharing of limited resources. The resource which one is allocating is one's time or effort. This reflects that the Gittins index solves a sequential selection problem since the optimization is with respect to time and the decision maker is looking for that stopping time which maximizes the expected reward per unit expected discounted time for a certain period. The allocation is being varied in time to meet changing conditions.

The decision maker who uses the Pearson index is not concerned about this type of maximization and selects projects without maximising the objective function with respect to time. Therefore, the Pearson index maximises the expected reward instantly as opposed to Gittins index which maximises the expected reward sequentially.

The difference is due to the principle of forward induction which in its simplest form is termed a one-step look ahead policy. The decision maker applies one-step look ahead policy when he compares stopping immediately with stooping after one period. The Gittins index is understood as the solution to the following problem:

In sequential selection, the decision maker has to think how far into the future he is required to use a certain project in order to achieve the maximum attainable reward up to the chosen time into the future.

The Pearson index is used to classify projects into two subsets, namely, projects which will be developed and projects which will not be considered.

2.5.3 Search problem

An object is located in one of the n possible locations ($n \geq 2$). Let p_i be the prior probability that the object is in the location i, where $p_i > 0$ ($i = 1, \ldots, n$.) and $\sum_{i=1}^{n} p_i = 1$. The decision maker is allowed to search only one location at a time. Let c_i, $c_i > 0$ be the cost to search location i. It is assumed that when the object is in location i it can be overlooked with probability $\alpha_i (0 \leq \alpha_i < 1)$ for $i = 1, \ldots n$. The objective of the decision maker is to discover the object at the minimal expected cost. One needs to devise a sequential search procedure which specifies how the decision maker should choose at each stage which location is to be searched next.

The above problem has been approached as follows. Suppose that the location j is searched first, and the object is not found. Then the posterior probability p_i^* that the object is in location i is

$$p_{i|j}^* = \begin{cases} \frac{p_j \alpha_j}{p_j \alpha_j + 1 - p_j} & \text{for } i = j \\ \frac{p_i}{p_j \alpha_j + 1 - p_j} & \text{for } i \neq j. \end{cases}$$

If the first location to be searched is j the probability that the object will be found in the first search is $p_j(1 - \alpha_j)$ and the probability that the object will not be found in the search is $p_j \alpha_j + (1 - p_j)$. Let $V(p_1, \dots, p_n)$ denote the minimal expected cost function. Then the optimality equation is given by

$$V(p_1, \dots, p_n) = \min_{j=1,\dots,n} \{c_j + (p_j \alpha_j + 1 - p_j) V(p_{1|j}^*, \dots, p_{n|j}^*)\}$$

where $V(p_{1|j}^*, \dots, p_{n|j}^*)$ is the expected cost of the remainder of the searching process when an optimal procedure is adopted after the first search had been completed unsuccessfully.

The strategy which minimizes the expected search cost (see page 20) is to examine the locations in descending order of the following priority index:

$$\delta_j = \frac{p_j(1 - \alpha_j)}{c_j}.$$

A more generalized result is that the expected cost of the search is minimized if the k^{th} search of the location j is in place i *i.e.,* i^{th} in order, if among the numbers $p_j \alpha_j^{k-1}(1 - \alpha_j)/c_j$, the (k, j) is the i^{th} largest (Black, 1965).

Search theory and its applications

Assume that there are n possible projects. Project j corresponds to a location j in the Search problem, for all j. Suppose now that the project j is divided into two consecutive tasks, denoted by j_1 and j_2 and that the probability of success of task j_i is p_{j_i}, independently for $i = 1, 2$. In this case the expected cost c_j of project j can be expressed as $c_j = c_{j_1} + p_{j_1} c_{j_2}$. Furthermore, the overlook probability α_j represents whether the project will be reconsidered which we set to zero for every project when the projects are not reconsidered. For simplicity, assume that all projects consist of two stages each. Let all the project give unit reward. We wonder in which order projects should be attempted in order to minimize the expected cost. Taking into account these changes the index becomes

$$\delta_j = \frac{p_j(1-\alpha_j)}{c_j} = \frac{p_{j_1}p_{j_2}}{c_{j_1}+p_{j_1}c_{j_2}} = \frac{p_{j_1}p_{j_2} - c_{j1} - p_{j1}c_{j2}}{c_{j_1}+p_{j_1}c_{j_2}} + 1.$$

This index gives the same rankings to a set of projects as the Pearson index for a two stage project with reward 1.

2.5.4 The Secretary Problem or the Search for the Best

In the secretary problem, an employer will interview n candidates sequentially in order to hire an individual to fill a vacancy for a secretarial position. After interviewing an individual, the employer must decide whether to accept (and terminate the process) or reject the current individual and continue the process. Once a candidate is rejected the candidate is no longer eligible. The only information available to the decision maker at any time is the relative rank of the current candidate compared with the previous candidates. The decision maker does not know how the current candidate compares with the candidates he has not seen yet.

The decision maker in the secretary problem wishes to appoint a candidate who ranks highly. The question is, when to take the positive decision of appointment. The difficulty related with this decision is its timing. If the decision is made too early in the sequence, one is neglecting the possibility of good candidates not being considered. On the other hand if made too late, the field of candidates remaining may not include the best candidate.

Maximization of the probability to appointing the best

One version of the objective is to maximize the probability of selecting the best candidate when all $n!$ orderings of the candidates are assumed to be equally likely. Let $V(r,\alpha)$ denote the maximum expected probability of choosing the best item just after the r^{th} interview when α is the relative rank of the r^{th} candidate. The next state of the process will be $V(r+1,b)$ where b is equally likely to be any one of the values $1, 2, \ldots, r+1$. The probability that the best candidate is realized if the r^{th} candidate is accepted assuming best of those seen is given by

$$\begin{aligned} P(r) &= P(\text{offer is best of } n | \text{offer is best of first } r) \\ &= \frac{\frac{1}{n}}{\frac{1}{r}} = \frac{r}{n}. \end{aligned}$$

A dynamic programming approach yields the equations

$$V(r,1) = \max\left\{\frac{r}{n}, \frac{1}{r+1}\sum_{b=1}^{r+1} V(r+1,b)\right\} \tag{2.27}$$

$$V(r,\alpha) = \frac{1}{r+1}\sum_{b=1}^{r+1} V(r+1,b) \qquad (\alpha = 2,3,\ldots,r) \tag{2.28}$$

with $V(n,\alpha) = 1$ if $\alpha = 1$ and 0 otherwise. Equation (2.27) equates the probability of appointing the best candidate when r candidates have been observed and not appointed and the r^{th} candidate having the first rank so far, $V(r,1)$, with the maximum of the following probabilities:

$P(r) = P(\text{offer is best of } n | \text{offer is best of first } r) = r/n$, or

$E\{V(r+1,b) | W_r\,;\, r^{\text{th}} \text{ is rejected}\}$ where the expectation is with respect to the ranking of the next candidate given W_r, that is, the information for the first r candidates and the r^{th} has been rejected.

Minimization of the expected rank

A different objective, yielding a second version of the problem, is to maximize the expected utility which has value $n-i$ when the best i^{th} candidate is accepted. This maximization corresponds to minimization of the expected rank of the accepted candidate.

Let $V(r,\alpha)$ be the expected utility of the optimal continuation when r candidates have been interviewed and the r^{th} has been found to have relative rank α. Let $V_0(r,\alpha)$ be the expected utility if the r^{th} candidate is accepted and the interview procedure is terminated.

Now consider the probability that the candidate which has rank α among the first r candidates has actually rank i among all n candidates

$$P_{n,i}(r,\alpha) = \frac{\binom{i-1}{\alpha-1}\binom{n-i}{r-\alpha}}{\binom{n}{r}} \qquad i = \alpha,\ldots,n+\alpha-r.$$

Therefore the expected utility is

$$V_0(r,\alpha) = \sum_{i=\alpha}^{n+\alpha-r} (n-i)P_{n,i} = n - \frac{n+1}{r+1}\alpha.$$

Also, the expected utility of interviewing the $(r+1)^{\text{th}}$ candidate having rejected the first r candidates and then continuing in an optimal way is

$$\frac{1}{r+1}\sum_{b=1}^{r+1} V(r+1,b)$$

In terms of Dynamic programming, we get:

$$\begin{aligned} V(r,\alpha) &= \max\left\{ V_0(r,\alpha)\,,\ \frac{1}{r+1}\sum_{b=1}^{r+1} V(r+1,b)\right\} && (2.29)\\ V(n,\alpha) &= n-\alpha \qquad (\alpha = 2,3,\dots,r). && (2.30)\end{aligned}$$

The optimal procedure is to continue the interviews if $V(r,\alpha) > V_0(r,\alpha)$ and to stop when $V(r,\alpha) = V_0(r,\alpha)$.

2.5.5 Dealing with Random rewards

Projects with equal expect rewards need to be ranked in a different way than the Pearson index if the decision maker has to allow for the fact that these projects might realize different gains. Suppose, that the reward n-stages before the end is equal to r. At the next stage, the information arrives that this project will be worth: either $\{r-\delta\}$ or $\{r+\delta\}$ with equal probability. Let us impose the condition that $0 < \delta < r$. If one is trying to maximize the expected reward, or maximize the probability of gaining the highest reward, which selection strategy should be followed?

Denote by $F_s(r)$ the maximal expected reward s-stages before the end of the project. Then, one can write the following optimality equation

$$F_s(r) = \max\left[r, \frac{1}{2}F_{s-1}(r-\delta) + \frac{1}{2}F_{s-1}(r+\delta)\right] = \max\left[r, EF_{s-1}(r)\right]$$

with terminal condition $F_0(r) = r$.

The optimal policy is either to choose the project with reward r or to choose both projects with rewards $r+\delta$ and $r-\delta$, *i.e.*, the maximal expected reward is the same whichever policy is followed.

Let us change the probabilities. The project with reward $r + \delta$ happens with probability p and the project with reward $r - \delta$ happens with probability $1 - p$.

Rewriting the problem as

$$F_s(r) = \max_{0 \le \delta \le r} \left[pF_{s-1}(r - \delta) + (1 - p)F_{s-1}(r + \delta) \right].$$

Now, we suppose that $p \ge \frac{1}{2}$.

Let $s = 1$, and terminal condition $F_0(r) = U(r)$ where $U(r)$ is a given utility function then,

$$F_1(r) = \max_{0 \le \delta \le r} \left[pU(r - \delta) + (1 - p)U(r + \delta) \right].$$

If $U(r)$ is a convex function, then consider a linear combination of two points of the convex utility function

$$pU(r + \delta) + (1 - p)(r - \delta) \ge \frac{1}{2}U(r + \delta) + \frac{1}{2}(r - \delta) \ge U(r)$$

where $p \ge \frac{1}{2}$. In this case the expected utility is greater than the utility gained for a single project which gives reward r. If, however, the utility is a concave function, then,

$$pU(r + \delta) + (1 - p)(r - \delta) \le \frac{1}{2}U(r + \delta) + \frac{1}{2}(r - \delta) \le U(r)$$

The solution to the above equation will be of the form: if $R < \alpha^\star$ then go for the project which gives reward r, otherwise select the project which will yield rewards $\{R + \delta\}$, $\{R - \delta\}$ with probability $\frac{1}{2}$.

2.6 Discussion

In a stochastic scheduling problem, which can be thought of as a case of sequential experimentation, the Gittins Index gives the solution of how to allocate one's effort over projects sequentially in time so as to maximize expected total discounted reward. However, the optimization is over an infinite horizon, and Gittins's result does not give an optimal solution to the finite horizon optimization problem. The Pearson Index might be more appropriate for the parallel selection method, since it is not clear that it solves any sequential optimization problem as the Gittins index does. Despite this, the Pearson index incorporates features of sequential decision procedures.

Chapter 3

Portfolio issues

3.1 Introduction

In the previous chapter, it was concluded that the Pearson Index can be used to solve a version of the optimal asset allocation problem when the objective is to maximize the expectation of a linear utility function subject to linear inequality constraints.

The above problem will be extended to that of maximizing the expected value of a non-linear utility function, to yield results which can be regarded as an extension of the Pearson index. Before the above problem is formulated, some general background on the portfolio theory is presented.

In the next section (3.2) Portfolio models are presented. We give an account of a non-linear resource allocation model, the fractional resource allocation model and the maximization of the utility of terminal wealth. In section (3.2.1) we refer to multi-period selection models. In section (3.3) we define the general form of the non-linear optimization problem, which is the extension of the optimal asset allocation problem solved by Pearson index. In section (3.4) different type of utility functions are presented. Then, in section (3.5) we formulate the optimisation problem using negative exponential utility function and we assume Normal distribution for the random profit. In section (3.6), we study the optimisation problem but we use exponential utility function.

3.2 Portfolio models

A Portfolio problem involves selecting a strategy for allocating the initial wealth among the available competing investment opportunities.

In Portfolio theory, a risky asset is an asset whose return is a random variable, as opposed to a riskless asset, which has a constant return. One can consider a riskless asset as a risky asset in which the return has variability zero. Thus, without loss of generality one can talk about risky assets only.

In order to identify an "optimum portfolio" one needs to define a criterion to measure the quality of a portfolio. Markowitz (1990) explains that:
"a portfolio analysis is characterized by.

1. the information concerning securities upon which it is based;
2. the criteria for better and worse portfolios which set the objectives of the analysis; and
3. the computing procedures by which portfolios meeting the criteria in (2) are derived from the inputs in (1)",

and that the result of a portfolio analysis is no more than the logical consequences of its information concerning securities. Instead of securities, we can have some projects.

One can think of an available budget C, to be allocated among a collection of risky projects. Let $X_i\,, i = 1, ...n$ be the amounts allocated to n alternatives R&D projects. From the i^{th} project, an uncertain payoff of size $r_i(X_i)$ will be received, a function of the amount X_i allocated to project i. The total reward is denoted by $R = \sum_{i=1}^{n} r_i(X_i)$. The utility function $u : [0, \infty) \rightarrow [0, \infty)$ gives us a return risk ordering for the random return of the chosen projects. The portfolio problem may be thought of as that of finding $X_i\,, i = 1, ...n$ to maximize the expected utility of the portfolio return generated, subject to the budget constraint $\sum_{i=1}^{n} X_i \leq C$.

This can be formulated as a mathematical programming problem:

$$\begin{array}{ll} \text{Maximize} & \mathrm{E}\{u(R)\} \quad \text{over } X_1, X_2, \ldots X_n \\ \textit{subject to} & R = \sum_{i=1}^{n} r_i(X_i), \\ & \sum_{i=1}^{n} X_i \leq C, \\ & X_i \geq 0, \quad \text{for all } i = 1, \cdots n. \end{array}$$

Markowitz (1952) started by considering the returns on securities (any sort of investments such as shares or capital projects) over some given time period as random variables with known finite means and variances. In this kind of analysis, it is also assumed that investors like high expected return, but dislike uncertainty about the amount of that return, as expressed by the variance. Therefore, portfolios with high expected return and low variance seem to be particular favorable.

The variance of the returns can enter the analysis in two ways. In the first case the variance is the objective function, and one desires to minimize the variance of the return for given level of the expected reward. In the other case the optimal portfolio is that achieving the highest expected reward for a given level of variance. Portfolios which have a criterion exclusively based on the mean and the variance of the gain generated by the portfolio are called *mean-variance portfolios.*

Portfolio diversification

Suppose that there are n asset categories and that the portfolio has proportions $\mathbf{y} = (y_1, \ldots, y_n)$ in each, with $\sum_i y_i = 1$. This constraint will be relaxed later.

Let the expected rewards for these categories be $\mathbf{r} = (r_1, \ldots, r_n)$. The expected reward of portfolio given by $\mathbf{y}$ can then be defined as $r(\mathbf{y}) = \mathbf{y} \cdot \mathbf{r} = \sum y_i r_i$. Let s_i be the standard deviation of the return yielded by asset i. Thus the vector $\mathrm{s} = (s_1, \ldots, s_n)$ describes the variability of the returns. The correlation between the returns of assets y_i and y_j is denoted by c_{ij}. The variance of the return from the portfolio $\mathbf{y}$, S^2 is:

$$S^2 = \sum_{i=1}^{n} y_i^2 s_i^2 + \sum_{i \neq j} \sum y_i y_j s_i s_j c_{ij},$$

where $i, j \in \{1, 2, \dots, n\}$.

Suppose, that the minimum risk portfolio is $\mathbf{m} = (m_1, \dots, m_n)$. Assume that there are some non-investment related features which create some risk denoted by e and this risk is uncorrelated with investment policy. The relative risk of portfolio $\mathbf{y}$ relative to portfolio $\mathbf{m}$ is the variance difference of two portfolios denoted by $S(\mathbf{y}, \mathbf{m})$ where:

$$S^2 = e^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - m_i) s_i c_{ij} s_j (y_j - m_j).$$

The disadvantage of this approach is that it is difficult to implement relative risk measure for non-symmetric distributions. In this example, lower covariances between the returns from investments will reduce the risk of the portfolio.

The non-linear resource allocation problem

In the non-linear resource allocation problem (Bretthauer & Shetty, 1995), resource should be allocated to each of a set of different activities. The objective is to maximize the financial return, which is a non-linear function of the resources allocated to each activity.

Suppose that, there are n different activities and denote the general activity by j, for $j = 1, \dots, n$. Let the total amount of the available resource be b. The variable x_j represents the amount of resource allocated to activity j, that is the project j. The variable x_j is supposed to have lower bound l_j and upper bound u_j. If an amount x_j of the resource is allocated to the activity j, then a non-linear return $f_j(x_j)$ is gained. The total return is $\sum_{j=1}^{n} f_j(x_j)$. The single constraint is the sum of n functions $\sum_{j=1}^{n} g_j(x_j)$ where $g_j(x_j)$ is the constraint for the activity j dependent on x_j for all j.

The non-linear resource allocation problem can be formulated as follows:

$$\begin{aligned} \textit{Maximize} \quad & \sum_{j=1}^{n} f_j(x_j) \\ \textit{subject to} \quad & \sum_{j=1}^{n} g_j(x_j) \leq b, \\ & l_j \leq x_j \leq u_i, \\ & x_j \quad \text{integer}, \qquad \text{for all } j = 1, \cdots n. \end{aligned}$$

The Fractional resource allocation problem

In the fractional resource allocation problem the objective function has ratio form (Ibaraki & Katoh, 1988). The criterion is the ratio of the expected utility to the standard deviation. The portfolio problem is:

$$
\begin{aligned}
&\textit{Maximise} \quad && \frac{\mathrm{E}\{u(R)\}}{\sqrt{V\{u(R)\}}} \quad \text{over } X_1, X_2, \ldots X_n \\
&\textit{subject to} \quad && R = \sum_{i=1}^{n} r_i(X_i), \\
& && \sum_{i=1}^{n} X_i \leq C, \\
& && X_i \geq 0, \quad \text{for all } i = 1, \ldots, n.
\end{aligned}
$$

Maximize the Utility of Terminal Wealth

So far the criterion to choose a portfolio was to maximize the expected utility of the random gain of the portfolio or a non-linear function of the resource allocated. A different criterion is to maximize the expected utility of terminal wealth (Huang & Litzenberger, 1988).

Consider an investor with initial wealth W_0. Suppose that there are n different risky investments and the individual invests y_j pounds in the j^{th} investment. The riskless interest rate is equal to r_f. The j^{th} risky investment yields return $y_j(1+r_j)$ where r_j is the random rate of return on the j^{th} risky investment. The terminal wealth is thus

$$
\begin{aligned}
W &= (W_0 - \sum_j y_j)(1 + r_f) + \sum_j y_j(1 + r_j) \\
&= W_0(1 + r_f) + \sum_j y_j(r_j - r_f).
\end{aligned}
$$

The investor's problem is to achieve

$$
\max_{y_j} \quad \mathrm{E}\left[u\left\{W_0(1 + r_f) + \sum_j y_j(r_j - r_f)\right\}\right].
$$

The concept of maximizing the expected utility of the terminal wealth is a criterion of multiperiod selection models. In this example the terminal wealth is expressed as a proportion of the initial wealth. In the next section the terminal wealth is just a summation of the cash flow in and out.

3.2.1 Multiperiod selection models

Imagine an investor who wants to maximize his terminal wealth W_T at a certain future point in time, T. The time between the present and the horizon is divided into n periods. At the end of each period the investor can change the structure of his portfolio in a way which he will maximize the expected utility of his terminal wealth W_T.

Let W_t (the state variable) be the value of the investor's portfolio at the beginning of the period $(t, t+1)$. Then,

$$W_{t+1} = W_t + \sum_{i=1}^{n} y_{it}(I_{it} - c_{it}), \qquad \text{for} \quad t = 0, 1, \ldots, T-1,$$

where I_{it} is the random return from asset i for $i = 1, \ldots, n$ at time t, c_{it} is the cost to hold a unit of asset i for period $(t, t+1)$ and y_{it} is the proportion of asset i chosen at period $(t, t+1)$, which has lower bound 0. The wealth available at time $t+1$ is equal to the value of the n risky investments $\sum_{i=1}^{n} y_{it} I_{it}$ minus their cost $\sum_{i=1}^{n} y_{it} c_{it}$ plus the wealth W_t at the beginning of the period $(t, t+1)$.

Let $f_t(W_t)$ be the expected utility of following an optimal policy from period t to the horizon T given the wealth W_t available at time t. By definition for $t = T$ we have (Elton & Gruber, 1975)

$$f_T(W_T) = \mathrm{E}[u(W_T) \mid W_{T-1}, y_{i(T-1)}]. \tag{3.1}$$

For $t = T-1$ we have,

$$f_{T-1}(W_{T-1}) = \max_{y_{i(T-1)}} \mathrm{E}[f_T(W_T) \mid W_{T-1}, y_{i(T-1)}], \tag{3.2}$$

where

$$W_T = W_{T-1} + \sum_{i=1}^{n} y_{i(T-1)} \{I_{i(T-1)} - c_{i(T-1)}\}. \tag{3.3}$$

Consider equation (3.2). The terminal wealth at time T, W_T, is a random variable. Equation (3.2) states that the value of W_{T-1} pounds is equal to the expected utility of the investment which has the maximum expected utility of its outcomes. It is easy to generalize this relationship for any time t:

$$f_t(W_t) = \max_{y_{it}} \mathrm{E}[f_{t+1}(W_{t+1}) \mid W_t, y_{it}], \qquad \text{for} \quad t = 0, 1, \ldots, T-1. \tag{3.4}$$

This equation represents the general dynamic programming formulation that can be used to solve all portfolio problems which have as their criterion the maximization of the expected utility of terminal wealth.

3.3 Selection of an Efficient Portfolio

In this section, a model is presented in the framework of consumer theory, where an individual chooses his consumption pattern by optimizing his utility subject to a budget constraint.

The target is to maximize the utility of the terminal wealth associated with the portfolio chosen. In an uncertain environment, the individual maximises the expectation of the utility at the next point in time. Denote the initial wealth by C, and suppose the individual implements project j at rate y_j $(0 \leq y_j \leq 1)$ for all $j = 1, \cdots n$. Project j requires an investment of c_j and yields random reward I_j. The terminal wealth $Z = C + \sum_{j=1}^{n} y_j(I_j - c_j)$ and the objective function is the mean value of the utility of the random terminal wealth $\mathrm{E}[u(Z)]$. The stochastic optimization problem is:

Problem 3

$$
\begin{array}{lll}
\textit{Maximise} & \mathrm{E}u(Z) = \mathrm{E}\left\{ u\left(C + \sum_{j=1}^{n} y_j(I_j - c_j) \right) \right\} & \\
\textit{subject to} & \sum_{j=1}^{n} y_j c_j \leq C, & \\
 & 0 \leq y_j \leq 1, & \text{for all } j = 1, \cdots n.
\end{array}
$$

The nature of utility function determines the level of difficulty of getting solutions for these stochastic optimization programs. Simple cases of non-linear utility functions are the negative exponential, exponential and quadratic function. These are, respectively,

$$
\begin{aligned}
u(z) &= 1 - \exp\{\lambda z\}, \\
u(z) &= \exp\{\lambda z\}, \\
u(z) &= z - mz^2.
\end{aligned}
$$

where λ and m are constants. The above problems yield a quadratic programming problem, with a closed form solution, when the utility is an exponential function of the terminal wealth and Normal distribution of the random rewards is assumed.

In the next sections, we first discuss different forms of utility functions. Then we investigate what happens if the utility function is negative exponential and the random reward has a Normal distribution.

3.4 Different approaches to risk

Utility functions represent the attitude of the decision maker towards risk.

One way to characterize an agent's attitude to risk is to examine whether or not the decision maker prefers a probability distribution of wealth to its expected value. In this approach, an agent is said to be:

1. Risk-averse if, for any probability distribution, he prefers the expected value of his distribution to the distribution itself;

2. Risk-neutral if, for any probability distribution, he is indifferent between the expected value of the distribution and the distribution itself; and

3. Risk-loving if, for any probability distribution, he prefers the distribution to its expected value.

Denote the utility function by $u(z)$ where z is wealth. The agent's attitude to risk is directly related to the curvature of his utility function $u(z)$ as follows:

$$\left.\begin{array}{l} Risk-averse \\ Risk-neutral \\ Risk-loving \end{array}\right\} \quad \text{if } u(z) \text{ is} \quad \left\{\begin{array}{l} concave \\ linear \\ convex. \end{array}\right.$$

One way to measure risk aversion is by means of the absolute risk aversion R_A, given by

$$R_A = -\frac{u''(z)}{u'(z)}.$$

Suppose that a decision maker is asked to state his preference between changing his initial wealth x_0 by random amount $\tilde{x}$ which has mean zero or paying a fixed amount π instead which may be dependent on his initial wealth x_0 and $\tilde{x}$. Consider a small fair bet $\tilde{x}$, *i.e.,* $\tilde{x}$ takes values in a very small interval and $E(\tilde{x}) = 0$. Now consider to add $\tilde{x}$ to the initial wealth x_0. Let $\pi(x_0, \tilde{x})$ be the decision maker's premium for $x_0 + \tilde{x}$. Then, the decision maker is indifferent between certain wealth $x_0 - \pi$ and a random terminal wealth which has mean utility $\mathrm{E}\{u(x_0 + \tilde{x})\}$, that is,

$$u(x_0 - \pi) = \mathrm{E}\{u(x_0 + \tilde{x})\}.$$

One can show that by using Taylor's expansion that

$$\pi(x_0, \tilde{x}) \approx \frac{1}{2}\sigma_x^2 R_A(x_0), \tag{3.5}$$

where σ_x^2 is the variance of the lottery $\tilde{x}$ which is equal to $E(\tilde{x}^2)$. By (3.5) the risk premium increases with the degree of absolute risk aversion and also the variance of the bet $\tilde{x}$. It is interesting to study what happens to $\pi(x_0, \tilde{x})$ as x_0 increases. We are interested in the following result:

Theorem 3.1 (see Keeney & Raiffa, 1976, p.167)
The risk aversion R_A is constant if only if $\pi(x_0, \tilde{x})$ is a constant function of x_0 for all $\tilde{x}$.

Constant absolute risk aversion is obtained from a negative exponential utility function.

3.5 Negative exponential utility

Consider the negative exponential utility function $u(z) = -\exp(-\lambda z)$ where λ is constant.

The problem becomes:

$$\begin{array}{ll} \textit{Maximise} & \mathrm{E}\left[-\exp^{-\lambda\left(C+\sum_{j=1}^{n} y_j(I_j - c_j)\right)}\right] \\ \textit{subject to} & \displaystyle\sum_{j=1}^{n} y_j c_j \leq C, \\ & 0 \leq y_j \leq 1, \qquad j = 1, \cdots n. \end{array}$$

This is equivalent to minimising a function of the moment generating function $\psi_i(y_i)$ of the distribution of the random reward I_j subject to the above constraints as follows:

$$\begin{array}{ll} Minimize & \prod_{j=1}^{n} \mathrm{E}\left[exp^{-\lambda y_j(I_j - c_j)}\right] = \prod_{j=1}^{n} \psi_j(-\lambda y_j)\left[exp^{(\lambda y_j c_j)}\right] \\ subject\ to & \sum_{j=1}^{n} y_j c_j \leq C, \\ & 0 \leq y_j \leq 1, \qquad \text{for all } j = 1, \cdots n. \end{array}$$

One case where an analytical approach would be used is when I_j has a Normal distribution with mean μ_j and variance σ_j^2. Its moment generating function is $\psi_i(y_i) = e^{\frac{1}{2}\sigma^2 y_i^2 + \mu_i y_i}$ and the minimising problem takes the following form:

Problem 4

$$\begin{array}{ll} Minimise & \mathcal{Q} = \sum_{j=1}^{n} y_j \left[\frac{1}{2}\sigma_j^2 \lambda^2 y_j - \lambda(\mu_j - c_j)\right] \\ subject\ to & \sum_{j=1}^{n} y_j c_j \leq C, \\ & 0 \leq y_j \leq 1, \qquad \text{for all } j = 1, \cdots n. \end{array}$$

The objective function Q requires the minimization of the variance of the selected projects as well as the maximization of their expected return. The objective function is convex since $\sigma_j^2 \lambda^2 > 0$ for $j = 1, \ldots, n$. The feasible set L consists of all $\boldsymbol{y} = (y_1, \ldots, y_n)$ such that $L = \{\mathbf{y} | 0 \leq y_j \leq 1 \, \forall j, \, \sum_{j=1}^{n} y_j c_j \leq C\}$ is nonempty and closed. Since the objective function is convex everywhere and the feasible set is a closed and bounded set in $\mathbb{R}^n$, then there must be a solution.

The above quadratic mathematical problem is non-trivial but one can note two interesting things. First, the above problem is a special kind of quadratic mathematical programming problem, known as a quadratic knapsack problem (Bretthauer & Shetty, 1995). It has been studied before in a different context and has several applications, for example, in promotion models, hydrological studies, determining if a graph possess a clique of order k (see, Gallo, Hammer, & Simeone, 1980). Below an explanation and demonstration is given how to solve a continuous version of integer quadratic Knapsack problem as presented by Bretthauer, Shetty, and Syam (1995). The second observation is that this problem can be thought of as a Mean-Variance Portfolio problem since the variance of the chosen portfolio will be $\sigma_j^2 \sum_{j=1}^n y_j^2$ and the expected value of the net gain is equal to $\sum_{j=1}^n y_j(\mu_j - c_j)$.

3.5.1 Quadratic Knapsack problem

In this section the continuous version of integer quadratic knapsack problem is introduced, with extension to inequality constraints and a solution method is presented. The integer quadratic knapsack problem (Bretthauer et al., 1995; Bretthauer & Shetty, 1995) is written as follows:

Problem 5

$$\begin{aligned}
&\textit{Minimize} \quad && f(\mathbf{y}) = \sum_{j=1}^{n} y_j \left\{ \frac{1}{2} d_j y_j - \alpha_j \right\} && (3.6)\\
&\textit{subject to} \quad && \sum_{j=1}^{n} b_j y_j \le b, && (3.7)\\
& && l_j \le y_j \le u_j, \qquad j = 1, \cdots n, && (3.8)\\
& && y_j \text{ an is integer}, \qquad j = 1, \cdots n, && (3.9)
\end{aligned}$$

where $d_j > 0$ and $b_j \ge 0$ for $j = 1, \ldots, n$ and b is a real constant. The continuous version of the quadratic knapsack problem is defined by the conditions (3.6), (3.7) and (3.8) only. Condition (3.8) can be written as two inequality constraints

$$l_j - y_j \le 0, \quad y_j - u_j \le 0, \qquad j = 1, \cdots n.$$

Kuhn-Tucker conditions

Non-linear optimization problems with inequality constraints can be solved by applying Kuhn-Tucker theory (Hadley, 1964), which gives of a set of conditions that need to be satisfied. The case we are interested in is:

$$\begin{aligned}
&\textit{Minimize} \quad && f(\mathbf{x})\\
&\textit{subject to} \quad && g_1(\mathbf{x}) \le 0, \ldots, g_m(\mathbf{x}) \le 0,\\
& && \mathbf{x} \in C,
\end{aligned}$$

where $C \subset \mathbb{R}^n$ is convex set and $f(\mathbf{x}), g_1(\mathbf{x}) \le 0, \ldots, g_m(\mathbf{x}) \le 0$ are convex functions defined on C. It is assumed that the objective function $f(\mathbf{x})$ and the constraints $g_i(\mathbf{x})$ for $i = 1 \ldots , m$ have continuous first partial derivatives. Kuhn-Tucker theory defines an unconstrained optimization problem which has the same solution as the original one, as follows:

The objective function of the unconstrained problem is called the Langragian function $L(\mathbf{x}, \boldsymbol{\mu})$, and is defined as

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mu_i g_i(\mathbf{x}),$$

where μ_i is positive constant, called a Langragian multiplier, which refers to the constraint $g_i(\mathbf{x})$, and $\boldsymbol{\mu} = (\mu_1, \ldots , \mu_m)$. If $\mathbf{x}^*$ is a feasible point for this program and an interior point of C, then $\mathbf{x}^*$ is a solution to the program if and only if there is a $\boldsymbol{\mu}^* = (\mu_1^*, \ldots , \mu_m^*)$ such that:

$$\mu_i^* \geq 0, \qquad \text{for} \quad i = 1, \ldots , m, \tag{3.10}$$

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \qquad \text{for} \quad i = 1, \ldots , m, \tag{3.11}$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0}. \tag{3.12}$$

where $\nabla f(\mathbf{x}^*)$ is the derivative at point $\mathbf{x}^*$. Condition (3.11) says that $\mu_i^* = 0$ unless the constraint $g_i(\mathbf{x})$ is active at $\mathbf{x}^*$. Condition (3.12) is necessary to identify the stationary point. The idea is that the existence of these multipliers sets up a transfer from the given constrained problem to the corresponding unconstrained problem. These multipliers provide information about the sensitivity of the solution to the constraints.

For the quadratic program, $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{Q}$ is a positive definite $n \times n$ matrix, $\mathbf{x} \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, and A is an $m \times n$ matrix of rank m. We have:

$$\text{Minimize} \qquad f(\mathbf{x}) = \alpha + \mathbf{c}^{\mathbf{T}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{Q}\mathbf{x} \tag{3.13}$$

$$\text{subject to} \qquad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \tag{3.14}$$

$$\mathbf{x} \geq \mathbf{0}. \tag{3.15}$$

The theory says that if a regular point $\mathbf{x}^*$ minimizes this quadratic program, then there are $\boldsymbol{\mu}^* \in \mathbb{R}^m$, $\boldsymbol{\nu}^* \in \mathbb{R}^n$ such that:

$$\mu^* \geq 0, \quad \text{for} \quad i = 1, \dots, k, \quad \nu^* \geq 0 \quad \text{for} \quad j = 1, \dots, n,$$
$$\mathbf{c} + \mathbf{Q}\mathbf{x} + \mathbf{A}^{\mathbf{T}}\boldsymbol{\mu}^* - \boldsymbol{\nu}^* = \mathbf{0},$$
$$\mu_i^*[(Ax)_i - b_i] = 0, \quad \text{for} \quad i = 1, \dots, m,$$
$$x_j^* \nu_j^* = 0, \quad \text{for} \quad j = 1, \dots, n.$$

However, in the continuous version of the Knapsack problem, when one applies the Kuhn-Tucker conditions the multipliers of the conditions are required to satisfy condition (3.7) and an algorithm is required to be applied in order to yield values which will satisfy all the conditions.
One can get closed form solutions as a function of the multiplier μ of the knapsack inequality constraint $\sum_{j=1}^n b_j y_j \leq b$. Let w_j be the multiplier for the constraint $l_j \leq y_j$, and v_j the multiplier for the constraint $y_j \leq u_j$. The Kuhn-Tucker conditions are

$$\frac{\partial}{\partial y_j} f(\mathbf{y}) + \mu \frac{\partial}{\partial y_j} \left[\sum_{j=1}^n b_j y_j - b \right] + w_j \frac{\partial}{\partial y_j}(y_j - u_j) +$$
$$+ u_j \frac{\partial}{\partial y_j}(l_j - y_j) = d_j y_j - \alpha_j + w_j - u_j + \mu b_j = 0, \qquad \text{for all} \quad j, \tag{3.16}$$
$$w_j(y_j - u_j) = 0, \qquad \text{for all} \quad j, \tag{3.17}$$
$$u_j(l_j - y_j) = 0, \qquad \text{for all} \quad j, \tag{3.18}$$
$$w_j \geq 0, \, u_j \geq 0, \qquad \text{for all} \quad j, \tag{3.19}$$

together with (3.7) and (3.8). For the problem with the inequality budget constraint (3.7), the extra conditions are

$$\left(\sum_{j=1}^n b_j y_j - b \right) \mu = 0, \tag{3.20}$$
$$\mu \geq 0. \tag{3.21}$$

The solution to the continuous version of the problem with constraint (3.7) requiring equality instead of inequality, is given by the following system of equations, in terms of μ:

$$\begin{aligned} y_j(\mu) &= Max\left\{Min\left\{\frac{(\alpha_j - \mu b_j)}{d_j}, u_j\right\}, l_j\right\}, \quad \text{for all } j, & (3.22)\\ w_j(\mu) &= Max\{\alpha_j - \mu b_j - d_j u_j, 0\}, \quad \text{for all } j, & (3.23)\\ v_j(\mu) &= Max\{\mu b_j - \alpha_j + d_j l_j, 0\}, \quad \text{for all } j. & (3.24)\end{aligned}$$

It has been proved by Nielsen and Zenios ((1992), Bretthauer et al., 1995) that for any μ the solution of the above equations satisfies the Kuhn-Tucker conditions, except for the knapsack constraint (3.7). This yields the additional equation $g(\mu) = b$, with

$$\begin{aligned} g(\mu) &= \sum_{j=1}^{n} b_j y_j(\mu) & (3.25)\\ &= \sum_{j=1}^{n} b_j Max\left\{Min\left\{\frac{(\alpha_j - \mu b_j)}{d_j}, u_j\right\}, l_j\right\}. & (3.26)\end{aligned}$$

The function $g(\mu)$ is a piecewise linear non-increasing function of μ. Denote by μ^* the value of μ that satisfies $g(\mu^*) = b$.

To solve problem 4, note that $d_j = (\sigma_j\lambda)^2$, $\alpha_j = \lambda(\mu_i - c_j)$, $b_j = c_j$ and $b = C$. The equations (3.22)-(3.23) become

$$\begin{aligned} y_j(\mu) &= Max\left\{Min\left\{\frac{\lambda\,(\mu_j - c_j) - \mu c_j}{\sigma_j^2\lambda^2}, 1\right\}, 0\right\},\\ w_j(\mu) &= Max\{\lambda(\mu_j - c_j) - \mu c_j - \sigma_j^2\lambda^2, 0\}, \qquad \text{for all} \quad j = 1, \cdots n\\ v_j(\mu) &= Max\{\mu c_j - \lambda(\mu_j - c_j), 0\}.\end{aligned}$$

In this case the optimal solution to the problem is given by

$$y_j(\mu) = \begin{cases} 1 & \text{if} \quad \mu^* \le \frac{\lambda(\mu_j - c_j) - (\lambda\sigma_j)^2}{c_j},\\ \frac{\lambda(\mu_j - c_j) - \mu^* c_j}{(\lambda\sigma_j)^2} & \text{if} \quad \frac{\lambda(\mu_j - c_j) - (\lambda\sigma_j)^2}{c_j} \le \mu^* \le \frac{\lambda(\mu_j - c_j)}{c_j},\\ 0 & \text{if} \quad \mu^* \ge \frac{\lambda(\mu_j - c_j)}{c_j}.\end{cases} \tag{3.27}$$

An example

Suppose that there are two projects 1 and 2. Let the total cost for project 1 be $c_1 = 3$ and that for the project 2 be $c_2 = 2$. The amount available to be spent is $C = 1.8$. Suppose that the random rewards for project 1 and 2 come from identical and independent Normal distributions, $I_j \sim N(5, 1)$ for $j = 1, 2$.

The target is the maximisation of the terminal wealth z. The utility function is $u(z) = \exp(-\lambda z)$.

$$\begin{aligned} &\mathit{Maximize} && \mathrm{E}u(Z) = \mathit{Max}\,\mathrm{E}\{-\exp[-\lambda(1.8 + y_1(I_1 - 3) + y_2(I_2 - 2))]\} \\ &\mathit{subject\ to} && 3y_1 + 2y_2 \le 1.8, \\ & && 0 \le y_1 \le 1 \quad \text{and} \quad 0 \le y_2 \le 1. \end{aligned}$$

The equivalent minimisation problem (see section, 3.5) is

$$\begin{aligned} &\mathit{Minimize} && \frac{1}{2}\lambda^2(y_1 + y_2) - 2\lambda y_1 - 3\lambda y_2 \\ &\mathit{subject\ to} && 3y_1 + 2y_2 \le 1.8, \\ & && 0 \le y_j \le 1, \quad j = 1, 2. \end{aligned}$$

Then

$$\begin{aligned} g(\mu) &= \sum_{j=1}^{n} c_j y_j(\mu) \\ &= 3\mathit{Max}\left\{\mathit{Min}\left\{(2 - 3\mu), 1\right\}, 0\right\} + 2\mathit{Max}\left\{\mathit{Min}\left\{(3 - 2\mu), 1\right\}, 0\right\} \end{aligned}$$

and setting $g(\mu) = 1.8$ yields $\mu = 1.05$, which gives the solution $y_1 = 0$, $y_2 = 0.9$.

3.6 The exponential utility function

Now suppose that the utility function is a positive exponential utility function $u(z) = e^{\lambda z}$.

The problem becomes:

$$\begin{aligned} &\mathit{Maximize} && \mathrm{E}u(Z) = \mathrm{E}\left[e^{\lambda c}\prod_{j=1}^{n} e^{\lambda y_j(I_j - c_j)}\right] && (3.28) \\ &\mathit{subject\ to} && \sum_{j=1}^{n} y_j c_j \le C, && (3.29) \\ & && 0 \le y_j \le 1, \qquad \text{for all } j = 1, \cdots n. \end{aligned}$$

One can view the objective function (3.28) as a function of the random profit $P_j = I_j - c_j$ yielded by each project j, for $j = 1, 2, \ldots, n$.

We want to investigate whether it is always better to spend more in the sense that the mean value of the utility of the terminal profit increases as y_j increases.

Consider

$$\frac{\partial}{\partial y_j}\mathrm{E}[\exp\{\lambda y_j(I_j - c_j)\}] = \mathrm{E}\left[\lambda(I_j - c_j)\exp\{\lambda y_j(I_j - c_j)\}\right] = \mathrm{E}[g_y(P)],$$

where $g_y(P) = \lambda P e^{\lambda y P}$.

In general, we try to understand how the slope of the objective function changes with respect to y. The unconstrained optimization problem is to choose y_j to minimize

$$\mathrm{E}\left[\exp\{\lambda y_j(I_j - c_j)\}\right],$$

and the necessary condition is to equate the first derivative with zero.

It is interesting to show that $\mathrm{E}\left[g_y(P)\right] \geq 0$, by showing that,

- $\mathrm{E}\left[g_y(P)\right] \geq g\left[\mathrm{E}g_y(P)\right]$,
- $g(\alpha) \geq 0$.

Let

$$\frac{\partial}{\partial y_j}\mathrm{E}\left[\exp\{\lambda y_j(I_j - c_j)\}\right] = \begin{cases} \mathrm{E}\left[\lambda(I_j - c_j)\right] = \alpha & \text{for} \quad y_i = 0 \\ \mathrm{E}\left[\lambda(I_j - c_j)\exp\lambda(I_j - c_j)\right] = b & \text{for} \quad y_i = 1. \end{cases}$$

If $\alpha, b > 0$ one can expect a maximum at $y_j = 1$. If both $\alpha, b < 0$ the maximum occurs at $y_j = 0$. If $\alpha < 0, b > 0$ then we expect a maximimum either at $y_j = 0$ or $y_j = 1$. The slope of the objective function for $y_j = 0$ is equal to the expected profit $\lambda\mathrm{E}(I - c)$. To check whether the slope of the objective function is positive, one can consider whether the $\mathrm{E}[g_y(P)]$ is a convex function in some range of profit. The first derivative of $g_y(P)$ is

$$g' = e^{\lambda y P} + \lambda y P e^{\lambda y P},$$

which is positive if the profit $P = I - c$ is positive. The second derivative is

$$g'' = \lambda e^{\lambda y P}(2y + \lambda y^2 P).$$

So $g_y(P)$ is convex in P for $P \geq -2/\lambda$.

Also by Jensen's inequality

$$\mathrm{E}[g_y(P)] \geq g_y\left[\mathrm{E}(P)\right] = g_y(\alpha) = \alpha e^{\alpha y},$$

where $\mathrm{E}[g_y(P)] > 0$ if only if $\alpha > 0$.

Therefore, if the expected profit $\mathrm{E}[g_y(P)] \geq 0$ and the profit $P \geq -2/\lambda$ always, then it is better to spend more.

Normal distribution

Using the moment generating function of the Normal distribution with mean μ_j and variance σ_j^2, the problem becomes

$$\begin{array}{ll} \textit{Maximize} & \frac{1}{2}\lambda^2 \sum_{j=1}^{n} \sigma_j^2 y_j^2 + \lambda \sum_{j=1}^{n} y_j(\mu_j - c_j) \\ \textit{subject to} & \sum_{j=1}^{n} y_j c_j \leq C, \\ & 0 \leq y_j \leq 1, \qquad \text{for all } j = 1, \cdots n. \end{array}$$

This is a quadratic programming problem, but this time it requires maximization of quadratic convex function because all terms $\sigma_j^2 > 0$ for $j = 1, \ldots, n$. Also all terms $c_j > 0$ and since y_j are positive the budget constraint $\sum_{j=1}^{n} y_j c_j > 0$. The feasible region is a closed convex polytope. The decision variables y_i are restricted between zero and one. The problem has multiple maxima.

The Kuhn-Tucker conditions for y, every global maximum, are

$$\sum_{j=1}^{n} y_j c_j \leq C, \tag{3.30}$$

$$1 - y_j \geq 0, \quad j = 1, \ldots, n, \tag{3.31}$$

$$y_j \geq 0, \quad j = 1, \ldots, n, \tag{3.32}$$

$$\lambda^2 \sigma_j^2 y_j + \lambda(\mu_j - c_j) + \mu c_j + \nu_j - \mu_j = 0, \quad j = 1, \ldots, n, \tag{3.33}$$

$$\mu \left(C - \sum_{j=1}^{n} y_j c_j \right) = 0, \quad j = 1, \ldots, n, \tag{3.34}$$

$$\nu_j y_j - \mu_j y_j = 0, \quad j = 1, \ldots, n, \tag{3.35}$$

$$\nu_j \geq 0, \quad \mu_j \geq 0, \quad j = 1, \ldots, n. \tag{3.36}$$

An example

Consider the following example.

$$\begin{aligned} &\textit{Maximize} && f(y_1, y_2) = y_1^2 + y_2^2 + y_1 + y_2 \\ &\textit{subject to} && 3y_1 + 2y_2 \leq 1.8, \\ & && 0 \leq y_1 \leq 1, \quad \text{and} \quad 0 \leq y_2 \leq 1. \end{aligned}$$

In terms of utility, each unit of asset contributes to utility in exactly the same way. However project 1 costs more per unit than project 2. The optimal solution is given by $y_1 = 0$ and $y_2 = 0.9$ and we have $f(0, 0.9) = 1.71$.

3.7 Discussion

In chapter 2, we concluded that the Pearson index solves the problem of maximizing the expected net reward of the selected project subject to a budget constraint. The Pearson index plays the role of Lagrangian multiplier in problem 2 in section (2.5.2). In chapter 3 we extended problem 2, to the problem 3 which is a maximization of non-linear utility function. We showed that when this utility is negative exponential the optimization problem is equivalent to Quadratic Knapsack problem. We suggest that the solution of the quadratic problem can be used as prioritisation index to select projects when the decision maker is risk averse with utility function $u(z) = -\exp(-\lambda z)$ where λ is a constant and the random profit comes from Normal distribution . The index has the following form as explained by equation (3.27):

$$\lambda \frac{\text{Expected net reward -}\lambda\text{(variance)}}{\text{Expected cost}}$$

where the term variance refers to the variance of the random profit.

Chapter 4

Forecasting System Modelling

4.1 The use of forecasting model

A forecasting model is developed to predict for a given future point in time, the value of Pharmaceutical share. This model might be useful to some Pharmaceutical managers who like to think in terms of the market potential of a drug in development, as a function of the value of the Pharmaceutical share. The forecast value of the random variable could be used for the evaluation of the portfolio of shares too.
The marketing man could also use these forecasts of the market value over time, in order to decide whether or not to enter a given segment of the market for classes of drugs related to a group of diseases such as osteoarthritis, rheumatoid arthritis and dismenohrea. A forecasting system can be used to predict future data values from those values that have already been observed, in order to base these decisions. We assume that information arrives continuously and that forecasts need to be modified to take account of it.

More technically, it can be assumed that the data process is modelled as a continuous time stochastic process, typically a diffusion type process with known drift and diffusion coefficient. This process is observed, usually at some points in time which are equally spaced, and observations are used to derive the forecasting system rule. If the observations are not equally spaced, the derived forecast rule will be different. In the case where coefficients of the data process are not known because they may depend on an unknown parameters and an unbiased predictor is required, then the forecast is not a martingale, but it is a reversed martingale as pointed out by Björk and Johansson (1992). Examples of this type are not the theme of this chapter.

Another interesting issue is that when the data process is observed randomly, that is, observations are collected as determined by a counting process (as originally suggested by Clark (1973)) the observation process might exhibit characteristics which do not belong to the original process. This is called a "Time Deformed Process". It is important to understand that there may be discrepancy between the data and the observation process.

The work in this chapter is about how to derive these streams of forecasts based on statistical laws. The forecaster considers as a "best" prediction method between two random variables their conditional expectation. The conditional variance of successive forecasts is calculated for the examples under consideration.

In the next section a framework for an adaptive prediction model is developed. In section (4.3) various generating data processes for the forecasting system are studied. Then in section (4.4) we develop the forecasting rule for each of these data generating processes. A forecasting rule for a stochastic variance model is given in section (4.5). The last section (4.6) is about some remarks of the theory of pricing of financial options in a Black and Scholes market and how these are relevant to the adaptive forecasting model.

4.2 Adaptive prediction model

Let V_1, V_2,... be a sequence of random variables from a specified joint distribution. The forecaster has observed the first n random quantities and noted n observations v_1, v_2, ... v_n $(n = 1, 2, \dots)$. His task is to issue a forecast at any time $t < T$ for the random variable $V = V_T$ by using the vector of the observed data $(v_0, v_2, \dots, v_t)$.

Let F_t be the forecast $F_t = E(V|v_t, v_{t-1}, \dots, v_0)$ that is, the conditional expectation of the random variable V given the present information v_t and the past information v_{t-1}, v_{t-2}, ... v_0 of the observed data. The conditional expectation F_t is the best approximation of V_T in the mean-squared error sense by a function of $(V_0, V_1, \ldots, V_t)$, that is to say that the mean square-error $E[(Y - V_T)^2]$ is minimised over all choices of functions of $Y = Y(V_0,\ V_1,\ \ldots\ V_t)$ by the choice $Y = F_t = E(V|V_0, V_1, \ldots, V_t)$. The problem is to deduce information about the value $V = F_T$ from the observed data up to the present time t, $(v_0, v_1, \ldots, v_t)$, and derive a rule for updating the forecast $F_t(\forall t < T)$ whenever a new observation is made.

Consider a forecast of the form $Y = F_t = E(V|v_t, v_{t-1}, \ldots, v_0)$. One could think about the distribution of these sequential forecasts $F_0, F_1, \ldots, F_t, \ldots, (\forall t < T)$. One should notice that the kind of prediction in the set up of the problem is a multi-step prediction, and not only for the next observation. Usually the predictive distribution several steps ahead is not easy to derive. In principle, the j-step ahead predictive distribution $P(V_T|V_{T-j})$ given the information $V_{T-j} = \{v_0, \ldots, v_{T-j}\}$ up to time $T - j$ is given by

$$P(V_T|V_{T-j}) = \int \quad \ldots \int \prod_{i=1}^{j} P(V_{T-j+i}|V_{T-j+i-1}) dV_{T-j+1} \ldots dV_{T-1}.$$

However, we can study the conditional mean value and the variance of forecasts. For example, the conditional expected value of F_T forecast can be shown to be equal to the previous forecast by using the properties of conditional expectation as follows:

We have

$$F_t = E(V|v_t, v_{t-1}, \ldots, v_0).$$

Then

$$\begin{aligned} E(F_t|v_{t-1}, v_{t-2}, \ldots, v_0) &= E\{E(V|v_t, v_{t-1}, \ldots, v_0)|v_{t-1}, v_{t-2}, \ldots, v_0\} \\ &= E(V|v_{t-1}, v_{t-2}, \ldots, v_0) \\ &= F_{t-1}. \end{aligned} \tag{4.1}$$

As a result the expected value of the next forecast is equal to its most recent forecast. Thus, F_t is a martingale process.
The variance of the forecast $Var(F_t) = \sum_{i=1}^{t} Var[(F_i - F_{i-1})]$ because the martingale differences $(F_i - F_{i-1})$ are uncorrelated. The more variance' terms are added

to the summation $\sum_{i=1}^{t} Var[(F_i - F_{i-1})]$, the variance increases. Therefore, the further into the future the forecast is made, the bigger its variance $Var(F_t)$ will be.

Consider the prediction error y_t, that is,

$$y_t = V_T - E(V_T | v_t, v_{t-1}, \cdots, v_0),$$

where it can easily be shown that its variance $\mathrm{Var}(y_t)$ is equal to the conditional variance of the forecast $V = F_T$ conditional on the information at time t, that is,

$$\begin{aligned} Var(y_t | v_t, v_{t-1}, \cdots, v_0) &= Var(V_T - F_t | v_t, v_{t-1}, \cdots, v_0) & (4.2) \\ &= Var(V_T | v_t, v_{t-1}, \cdots, v_0), & (4.3) \end{aligned}$$

which happens to be the conditional mean square error.

The behaviour of the conditional variance of the prediction error is dependent on the assumptions of the data generating process and this might imply that the conditional variance does not decrease.

However, the variance of the random variable V_T is equal to

$$\begin{aligned} \mathrm{Var}(V_T) &= Var[E(V_T | v_t, v_{t-1}, \cdots, v_0)] + E\left[Var(V_T | v_t, v_{t-1}, \cdots, v_0)\right] \\ &= Var[E(V_T | v_{t+1}, v_t, \cdots, v_0)] + E[Var(V_T | v_{t+1}, v_t, \cdots, v_0)] \\ &= Var[F_t] + E[\phi_t] \\ &= Var[F_{t+1}] + E[\phi_{t+1}], & (4.4) \end{aligned}$$

where $\phi_t = Var[V_T | v_t, v_{t-1}, \cdots, v_0]$.

Generally, the conditional distribution $\{V_T | v_t, v_{t-1}, \cdots, v_0\}$ has mean F_t and variance ϕ_t. So, at the time 0 the variance of the forecast $\mathrm{Var}(F_0)$ is small relative to $\mathrm{Var}(F_t)$ (for any $t > 0$) and because $\mathrm{Var}(V_t)$ is constant, it is concluded that $E\{\phi_t\} > E\{\phi_{t+1}\}$ for any t. Now consider the expectation of conditional variance:

$$\begin{aligned} \mathrm{E}(\phi_{t+1} | \mathcal{F}_t) &= E\left[E\{(F_T - F_{t+1})^2 | \mathcal{F}_{t+1}\} | \mathcal{F}_t\right] \\ &= E\{(F_T - F_{t+1})^2 | \mathcal{F}_t\} \\ &= E\{(F_T - F_t)^2 + (F_t - F_{t+1})^2 | \mathcal{F}_t\} \\ &= \phi_t + E[(F_{t+1} - F_t)^2 | \mathcal{F}_t], & (4.5) \end{aligned}$$

where $\mathrm{E}\{(F_T - F_{t+1})(F_{t+1} - F_t)|\mathcal{F}_t\} = \mathrm{E}\{(F_T - F_{t+1})|\mathcal{F}_t\}\{E(F_{t+1}|\mathcal{F}_t) - F_t\} = 0$.
From (4.5), we get

$$\mathrm{E}(\phi_{t+1}|\mathcal{F}_t) \leq \phi_t.$$

Therefore, the conditional variance is a supermartingale process.
In the case where the conditional variance $\phi_t = Var[V_T|v_t, v_{t-1}, \cdots, v_0]$ increases as $t \to T$, it implies that the more data here is the greater is the uncertainty.
One could also calculate the variance of the difference between the predicted variable V and the current forecast F_t.
Note that,

$$V_T - F_t = F_t + (F_{t+1} - F_t) + \cdots + (F_T - F_{T-1}) - F_t = \sum_{i=1}^{T-t}(F_{t+i} - F_{t+i-1})$$

then,

$$\begin{aligned}
Var(V - F_t) &= Var\left[\sum_{i=1}^{T-t}(F_{t+i} - F_{t+i-1})\right] \\
&= \sum_{i=1}^{T-t} Var(F_{t+i} - F_{t+i-1}) + \\
&\qquad + \sum\sum Cov(F_{t+i} - F_{t+i-1}, F_{t+j} - F_{t+j-1}) \\
&= \sum_{i=1}^{T-t} E[(F_T - F_t)^2],
\end{aligned}$$

since $E(F_{t+1} - F_{t+1-1}) = 0$ for every i because of martingale property of Forecasts F_t.
By assuming that the $E(F_t^2) < \infty$, $\forall t$.

$$Var(V_T - F_t) = E(F_T - F_t)^2 = \sum_{i=1}^{T-t} E[(F_{t+1} - F_{t+i-1})^2]$$

For the conditional variance $Var(V_T - F_T|v_t, v_{t-1}, \cdots, v_0) = Var(V_T - F_T|V_t)$

$$\begin{aligned}
Var(V - F_t|v_t) &= E((F_T - F_t)^2|V_t) \\
&= E(F_T^2 - F_t^2 - 2F_t(F_T - F_t)|V_t) \\
&= E(F_T^2 - F_t^2|V_t)
\end{aligned}$$

4.3 Data Generating Processes

For the development of the forecasting system, a process which will generate the set of observations needs to be specified. Name this process, "Observation process". This observation process may be a continuous or discrete stochastic process. A continuous time model can be represented in terms of a stochastic differential equation whereas in a discrete time framework the observations follow a stochastic difference equation.

Whether one should have a preference for either a continuous or discrete version of the observations process, is not clear, but since the random variable is the value of a share which can be measured at any particular point in time and not only in some time intervals, a continuous time framework is acceptable.

In this section an account is given of possible models for the dynamics of share prices both in continuous and discrete time. Continuous time modeling might be more useful for the development of theoretical models, as opposed to the forecaster job who receives the data which are sampled at discrete time intervals.

Geometric Brownian Motion

Black and Scholes (1973) assumed that stock prices $S = (S_t)_{t \geq 0}$ follow a Geometric Brownian motion (Øksendal, 1995), that is,

$$dlogS_t = \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t, \tag{4.6}$$

where $W = (W_t)_{t \geq 0}$ is a Brownian motion (the Wiener process), r is some constant interest rate and $\sigma > 0$ is some volatility coefficient. The solution (Björk, 1998 p. 55) of the stochastic differential equation (4.6) is,

$$S_t = S_0 e^{(\sigma W_t - \frac{\sigma^2}{2} t)} e^{(rt)}. \tag{4.7}$$

The discrete analogue of the process (4.6) can be explained as follows: Divide the time horizon of the process $[0, T]$ into a partition $\{0 \equiv t_0 < t_1 < \cdots < t_N \equiv T\}$ such that for each $h = 1, \cdots, N$, the difference $t_h - t_{h-1} = \Delta t \equiv \frac{T}{N}$. Then the logarithmic increments $\Delta_t s \equiv \log S_t - \log S_{t-\Delta t}$ of the stock price over each of these time steps are i.i.d normal, with mean $(r - \frac{1}{2}\sigma^2)\Delta t$ and variance $\sigma^2 \Delta t$. The logarithmic process is $\log S_T = \log S_0 + \sum_{h=1}^{N} \Delta_t s$. The random variable $\log S_T$ has Normal distribution,

$$N\left[\log S_0 + \left(r - \frac{1}{2}\sigma^2\right)(T - t_0), \sigma^2(T - t_0)\right], \tag{4.8}$$

where its mean and variance are not dependent on the intermediate values of $\log S_i$ for $i = 1, \ldots, T-1$.

Now, for the forecasting problem, we assume that we observe a stochastic process V_t which follows a Normal distribution with constant mean $\delta = r - \frac{1}{2}\sigma^2$ and variance σ^2 that is, $\log \frac{V_{t+1}}{V_t} \sim N(\delta, \sigma^2)$.
Generally, it could be assumed that $U_{t+1} = \log V_{t+1} - \log V_t$ are i.i.d and that the moment generating function of the underlying distribution U exists and denoted by $M_U(h) = E[e^{hU}]$.

Mean Reverting Ornstein-Uhlenbeck Process

Instead of the above version of the discrete type Geometric Brownian motion model, where the log process is an arithmetic random walk with i.i.d normal increments, one could use a model for the successive logarithms of the observed values of $\log V_t$ as a mean-reverting process, and also known as an Ornstein-Unlenbeck process (Arnold, 1971, p.134) which is a continuous representation of an AR(1) model. In continuous time,

$$\begin{aligned} d\log V_t &= -\alpha(\log V_t - \mu)dt + \sigma dW_t \\ \log V_0 &= x, \end{aligned} \tag{4.9}$$

where $\alpha \geq 0$ is the speed of reversion, μ is the level to which $\log V_t$ tends to revert, and x is the known value of the process at its starting point.

Proposition 1 Equation (4.9) has the following solution

$$\begin{aligned} \log V_t &= xe^{-\alpha t} + \mu\alpha \int_0^t e^{-\alpha(t-s)} ds + \sigma \int_0^t e^{-\alpha(t-s)} dW_s \\ &= xe^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma \int_0^t e^{-\alpha(t-s)} dW_s. \end{aligned} \tag{4.10}$$

Proof.

Set $Y_t = \log V_t e^{\alpha t}$. Then

$$\begin{aligned} dY_t &= d(\log V_t)e^{\alpha t} + \alpha(\log V_t)e^{\alpha t}dt \\ &= -\alpha(\log V_t - \mu)e^{\alpha t}dt + \sigma e^{\alpha t}dW_t + \alpha(\log V_t)e^{\alpha t}dt \\ &= \mu\alpha e^{\alpha t}dt + \sigma e^{\alpha t}dW_t. \end{aligned}$$

Integrating

$$\log V_t e^{\alpha t} - \log V_0 = \mu\alpha \left[\frac{e^{\alpha s}}{\alpha}\right]_0^t + \sigma \int_0^t e^{\alpha s} dW_s$$

Therefore

$$\log V_t = \log V_0 e^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma \int_0^t e^{-\alpha(t-s)} dW_s. \qquad \square$$

A discrete version of OU process is:

$$\log V_{t+\Delta t} - \log V_t = -\alpha(\Delta t)\log V_t + \mu\Delta t + \epsilon_t \tag{4.11}$$

where $\epsilon \sim N(0, \sigma^2\Delta t)$. The log-price process is the sum of a zero-mean stationary Gaussian process. The successive logarithm values for $\Delta t = 1$ have conditionals moments

$$\mathrm{E}[\log V_{t+1} - \log V_t | \log V_t] = -\alpha \log V_t + \mu, \tag{4.12}$$

$$Var[\log V_{t+1} - \log V_t | \log V_t] = \sigma^2. \tag{4.13}$$

The distribution of the sequence of observations $\log V_t, \log V_{t+1}, \log V_{t+2}, \cdots$ can be thought of as a Conditionally Gaussian Model, that is, the moments of the Normal distribution are dependent on the past information

$$\{\log V_{t+1} - \log V_t | \ \log V_t, \dots, \} \sim N\left(\mu_t\{\log V_t\}, \sigma_t^2\{\log V_t\}\right) \tag{4.14}$$

and from (4.11) follows that $\log V_{t+\Delta t}$ is distributed as

$$N\left\{-\alpha(\Delta t)\log V_t + \mu(\Delta t), \sigma^2\right\},$$

given V_t, where $\mu_t\{\log V_t\}$ is the mean and $\sigma_t^2\{\log V_t\}$ is the variance.

Trending Ornstein-Uhlenbeck Process

An alternative process is the Trending OU Process (see, Lo & Wang, 1995, p.93) as follows

$$\begin{aligned} d\log S_t &= [-\alpha(\log S_t - \mu t) + \mu]\,dt + \sigma dW_t \\ \log S_0 &= x. \end{aligned} \tag{4.15}$$

Rewriting equation (4.15) as

$$d[\log S_t - \mu t] = -\alpha(\log S_t - \mu t)dt + \sigma dW_t, \tag{4.16}$$

and we get the explicit solution:

$$\begin{aligned} \log S_t &= xe^{-\alpha t} + \mu \int_0^t s ds + \sigma \int_0^t e^{-\alpha(t-s)} dW_s \\ &= xe^{-\alpha t} + \mu t + \sigma \int_0^t e^{-\alpha(t-s)} dW_s. \end{aligned} \tag{4.17}$$

4.4 Forecasting Rules

Consider a forecaster who is required to predict the future behaviour of a given one dimensional stochastic process $V = \{V(t); t \geq 0\}$ at a certain point in time T where $t \leq T$. The process has been observed and some values noted. In this section we derive the forecasting rules for $\log V_T$ and V_T given the information up to the present. Examples for different stochastic processes are presented.

Forecasting the value of log V_T

Assume that the observation process is Geometric Brownian motion (see section 4.3). Denote $\log V_t$ by W_t. Now if one wants to forecast $logV_T$ process then, denote its forecast by Φ_t

$$\begin{aligned} \Phi_t &= E(logV_T|W_t, W_{t-1}, \ldots, W_0) = E(W_T|W_t, W_{t-1}, \cdots, W_0) \\ &= \mathrm{E}\left\{ W_t + \sum_{j=t+1}^{T} (logV_j - logV_{j-1})|W_t, W_{t-1}, \cdots, W_0 \right\} \\ &= W_t + (T-t)\delta \\ &= logV_t + (T-t)\delta. \end{aligned} \tag{4.18}$$

Consider the conditional mean square error which is defined to be

$$\begin{aligned}
\mathrm{E}[(\Phi_t - logV_T)^2|W_t,\cdots,W_0] &= \mathrm{E}[(logV_t + (T-t)\delta - logV_T)^2|W_t,\cdots,W_0] \\
&= \mathrm{Var}[logV_T - logV_t|W_t,\cdots,W_0] \\
&= (T-t)\sigma^2.
\end{aligned}$$

It is concluded that the conditional mean square error decreases as $t \to T$.

Consider the difference between two successive forecasts $\Delta\Phi_{t+1} = \Phi_{t+1} - \Phi_t$.

$$\Delta\Phi_{t+1} = \Phi_{t+1} - \Phi_t = logV_{t+1} - logV_t - \delta. \tag{4.19}$$

It can be seen that

$$\begin{aligned}
\mathrm{E}(\Delta\Phi_{t+1}|W_t, W_{t-1},\cdots,W_0) &= \mathrm{E}(\Phi_{t+1} - \Phi_t|W_t, W_{t-1},\cdots,W_0) \\
&= \mathrm{E}(logV_{t+1} - logV_t - \delta|W_t, W_{t-1},\cdots,W_0) \\
&= 0.
\end{aligned}$$

This is not suprising since the forecasts form a martingale process, and the expected value of a martingale difference is zero.

For the variance of successive forecasts, we have:

$$\begin{aligned}
\mathrm{E}((\Delta\Phi_{t+1})^2|W_t, W_{t-1},\cdots,W_0) &= \mathrm{E}((logV_{t+1} - logV_t - \delta)^2|W_t, W_{t-1},\cdots,W_0) \\
&= \mathrm{Var}(logV_{t+1} - logV_t - \delta|W_t, W_{t-1},\cdots,W_0) \\
&= \mathrm{Var}(logV_{t+1} - logV_t|W_t, W_{t-1},\cdots,W_0) \\
&= \sigma^2.
\end{aligned}$$

Similarly, we derive the forecasting rule for the value V_T.

Forecasting the value of V_T

Denote $U_{t+1} = log\frac{V_{t+1}}{V_t}$.

Let $W_t = logV_t \Leftrightarrow V_t = e^{W_t}$.

So, $U_{t+1} = W_{t+1} - W_t$ where independent increments for W_t are assumed to be distributed with any arbitrary distribution with moment generating function denoted by $M_{U_t}(h) = E[e^{hU_t}]$. We express $V = W_T = W_t + \sum_{j=t+1}^{T} U_j$ and if the forecast F_t takes the form $F_t = E(V|v_t, v_{t-1},\ldots,v_0)$, then it becomes

$$\begin{aligned}
F_t &= e^{W_t} E\left(e^{\sum_{j=t+1}^{T} U_j}|v_t, v_{t-1},\ldots,v_0\right) && (4.20)\\
&= e^{W_t} E\left(e^{\sum_{j=t+1}^{T} U_j}\right). && (4.21)
\end{aligned}$$

But $E(e^{\sum_{j=t+1}^{T} U_j})$ is the moment generating function of a summation of independent and identically distributed random variables which is equal to

$$\begin{aligned} E(e^{\sum_{j=t+1}^{T} U_j}) &= \prod_{j=t+1}^{T} E(e^{U_j}) \\ &= \{M_U(1)\}^{(T-t)} \\ &= \{E(e^{U_t})\}^{(T-t)}. \end{aligned}$$

The forecast F_t takes the form

$$F_t = e^{W_t}\{E(e^{U_t})\}^{(T-t)} = V_t\{E(e^{U_t})\}^{(T-t)}. \tag{4.22}$$

Example 1: Geometric Brownian motion

If the stock prices follow Geometric Brownian motion, then the moment generating function of the increments is a Normal distributions with mean δ and variance σ^2 and the forecast F_t becomes

$$F_t = V_t\left(e^{\delta+\frac{1}{2}\sigma^2}\right)^{(T-t)} = V_t\lambda^{(T-t)}. \tag{4.23}$$

where $\lambda = \left(e^{\delta+\frac{1}{2}\sigma^2}\right)$. The forecast F_t is equal to the value of the process V_t if $\lambda = 1$, that is, $F_t = V_t$.

To study the behaviour of the successive forecast, one can calculate the conditional variance of the successive forecasts. Let ΔF_{t+1} denote the difference of successive forecasts

$$\begin{aligned} \Delta F_{t+1} &= F_{t+1} - F_t = V_{t+1}\lambda^{(T-t-1)} - V_t\lambda^{(T-t)} \\ &= \lambda^{(T-t-1)}(V_{t+1} - \lambda V_t) \\ &= \lambda^{(T-t-1)}(e^{W_{t+1}} - \lambda e^{W_t}) \\ &= \lambda^{(T-t-1)}e^{W_t}(e^{U_{t+1}} - \lambda). \end{aligned} \tag{4.24}$$

The conditional expectation $\mathrm{E}(\Delta F_{t+1}|v_t)$ is equal to zero due to martingale property. The conditional variance of successive forecasts can be calculated as follows

$$\begin{aligned}
\mathrm{E}(\Delta F_{t+1}^2|v_t, v_{t-1}, \cdots v_0) &= \mathrm{E}(\Delta F_{t+1}^2|\mathcal{F}_t) \\
&= \mathrm{E}\left\{\lambda^{2(T-t-1)}e^{2W_t}(e^{U_{t+1}} - \lambda)^2|\mathcal{F}_t\right\} \\
&= \lambda^{2(T-t-1)}V_t^2\mathrm{E}\left\{(e^{U_{t+1}} - \lambda)^2|\mathcal{F}_t\right\} \\
&= \lambda^{2(T-t-1)}V_t^2\mathrm{E}\left[\{e^{U_{t+1}} - \mathrm{E}(e^{U_{t+1}}|\mathcal{F}_t)\}^2|\mathcal{F}_t\right] \\
&= \lambda^{2(T-t-1)}V_t^2\left[\mathrm{E}\{(e^{U_{t+1}})^2|\mathcal{F}_t\} - \mathrm{E}^2(e^{U_{t+1}}|\mathcal{F}_t)\right] \\
&= \lambda^{2(T-t-1)}V_t^2\mathrm{E}\left\{(e^{U_{t+1}})^2|\mathcal{F}_t)\right\} \\
&= \lambda^{2(T-t-1)}V_t^2\mathrm{E}\left\{(V_{t+1} - V_t)^2|\mathcal{F}_t)\right\} \\
&= \lambda^{2(T-t-1)}{v_t}^2\sigma^2 \\
&= \left(e^{\delta+\frac{1}{2}\sigma^2}\right)^{2(T-t-1)}{v_t}^2\sigma^2. \qquad (4.25)
\end{aligned}$$

The relation between the two forecasts, F_t and Φ_t can found as follows:
Take the logarithm of (4.23),

$$logF_t = logV_t + (T-t)\left(\delta + \frac{1}{2}\sigma^2\right). \qquad (4.26)$$

Compare (4.18) and (4.26)

$$\Phi_t = logF_t - \frac{1}{2}\sigma^2(T-t), \qquad (4.27)$$

where $T \geq t$. Therefore, the forecast of the $logV_T = \phi_t$ is always smaller in magnitude than the logarithm of the forecasted value of V_t, denoted by $logF_t$. This conclusion can also be justified by using Jensen's inequality, as follows:

Suppose that $g(x)$ is concave function, and X is a random variable. Jensen's inequality says that $E[g(x)] \leq g(E[x])$ provided the expectations exists.
Let the forecast $F_t = x = E(V|v_t, v_{t-1}, \ldots, v_0)$ and let $g(X) = logF_t$.
Therefore,

$$\begin{aligned}
E[g(x)] = \Phi_t = E[log\Phi_{t+1}] &= E\left[logF_{t+1} - \frac{1}{2}\sigma^2(T-t-1)\right] \\
&\leq log[E(F_{t+1})] = logF_t.
\end{aligned}$$

So,

$$\Phi_t \leq logF_t, \quad \forall t.$$

Clearly this result is consistent with (4.27).

Example 2: Process with Gamma increments

Suppose that the increments U_t follow Gamma(α, β). Then the forecast F_t

$$F_t = e^{W_t}\{E(e^{U_t})\}^{(T-t)} = V_t\left\{\frac{\beta}{\beta-1}\right\}^{\alpha(T-t)}. \tag{4.28}$$

Therefore,

$$F_t \neq V_t \qquad \text{for any } \beta \text{ and } t > 0$$

Denote the difference of successive forecasts with ΔF_{t+1}

$$\begin{aligned}
\Delta F_{t+1} &= F_{t+1} - F_t \\
&= V_{t+1}\left\{\frac{\beta}{\beta-1}\right\}^{\alpha(T-t-1)} - V_t\left\{\frac{\beta}{\beta-1}\right\}^{\alpha(T-t)} \\
&= (V_{t+1} - V_t)\left\{\frac{\beta}{\beta-1}\right\}^{\alpha(T-t-1)} - V_t\left\{\frac{\beta}{\beta-1}\right\} \\
&= (V_{t+1} - V_t)\psi^{\alpha(T-t-1)} - V_t\psi.
\end{aligned}$$

where $\psi = \frac{\beta}{\beta-1}$. The conditional variance is equal to

$$\begin{aligned}
Var\{\Delta F_{t+1}|v_t, v_{t-1}, \cdots v_0\} &= Var\{F_{t+1} - F_t|v_t, v_{t-1}, \cdots v_0\} \\
&= Var\left\{(V_{t+1} - v_t)\psi^{\alpha(T-t-1)} - v_t\psi \mid v_t, v_{t-1}, \cdots v_0\right\} \\
&= \psi^{2\alpha(T-t-1)}Var\{V_{t+1} - v_t|v_t, v_{t-1}, \cdots v_0\}' \\
&= \frac{\alpha}{\beta^2}\left\{\frac{\beta}{\beta-1}\right\}^{2\alpha(T-t-1)}.
\end{aligned} \tag{4.29}$$

The variance of successive forecasts decreases as the number of steps ahead $T-t$ diminishes.

Example 3: Mean Reverting Ornstein-Uhlenbeck Process

Recall that if V_t follows a mean reverting OU process its solution is given by

$$V_t = V_0 e^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma \int_0^t e^{-\alpha(t-s)} dW_s. \tag{4.30}$$

Choose $\delta t = 1$ and given the availability of a set of observations up to and including time t, the forecast one step ahead is,

$$\mathrm{E}(V_{t+1}|v_t, v_{t-1}, \cdots v_0) = v_t e^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma \mathrm{E}\left(\int_t^{t+1} e^{-\alpha(t+1-s)} dW_s\right).$$

The expectation of the stochastic integral is zero and therefore the forecast is

$$\mathrm{E}(V_{t+1}|v_t, v_{t-1}, \cdots v_0) = \mu + (v_t - \mu)e^{-\alpha}. \tag{4.31}$$

Optimal k-step-ahead forecasts for $k > 1$ can be evaluated in two different ways. The first one is to use appropriate forecasts in place of future values of the process of V_t. The second approach is to back-substitute from the defining equation of the process (4.30) so as to eliminate future values of V_t. For example, in order to get the forecast $\mathrm{E}(V_{t+2}|v_t, \cdots v_0)$ with the first method, we need an estimate $\hat{V}_{t+1}$ for the future value of V_{t+1}. Let the estimate be $\hat{V}_{t+1} = \mathrm{E}(V_{t+1}|v_t, v_{t-1}, \cdots v_0) = \mu + (v_t - \mu)e^{-\alpha t}$ which is the optimal forecast. Applying the first method, we get:

$$\begin{aligned}
\mathrm{E}(V_{t+2}|v_t, v_{t-1}, \cdots v_0) &= \mathrm{E}\left[\mathrm{E}(V_{t+2}|V_{t+1}, v_t)|v_t, \cdots v_0\right] \\
&= \mathrm{E}\left[\mathrm{E}(V_{t+2}|V_{t+1})|v_t, \cdots v_0\right] \\
&= \mathrm{E}\left[\mu + (\hat{V}_{t+1} - \mu)e^{-\alpha}|v_t, \cdots v_0\right] \\
&= \mu + (v_t - \mu)e^{-2\alpha}. \qquad (4.32)
\end{aligned}$$

Similarly, for the forecast $(T-t)$ steps ahead, we have:

$$\begin{aligned}
F_t = \mathrm{E}(V_T|v_t, v_{t-1}, \cdots v_0) &= \mathrm{E}\cdots\left[\mathrm{E}(V_T|V_{T-1}, \cdots V_{t+1}, v_t)|v_t, \cdots v_0\right] \\
&= \mu + (v_t - \mu)e^{-\alpha(T-t)}. \qquad (4.33)
\end{aligned}$$

For the second approach, the process at time $t = T$ is

$$\begin{aligned}
V_T &= V_{T-1}e^{-\alpha} + \mu(1 - e^{-\alpha}) + \sigma e^{-\alpha T}\int_{T-1}^{T} e^{\alpha s} dW_s \\
&= \left\{V_{T-2}e^{-\alpha} + \mu(1 - e^{-\alpha}) + \sigma e^{-\alpha(T-1)}\int_{T-2}^{T-1} e^{\alpha s} dW_s\right\} e^{-\alpha} \\
&\qquad + \mu(1 - e^{-\alpha}) + \sigma e^{-\alpha T}\int_{T-1}^{T} e^{\alpha s} dW'_s,
\end{aligned}$$

where $W(s)$ and $W'(s)$ are independent Wiener processes for every s. We have:

$$V_T = V_{T-2}e^{-2\alpha} + \mu(1 - e^{-2\alpha}) + \sigma e^{-\alpha T}\int_{T-1}^{T} e^{\alpha s} dW_s + \sigma e^{-\alpha T}\int_{T-2}^{T-1} e^{\alpha s} dW'_s.$$

Similarly, one can show that

$$\begin{aligned}
V_T &= V_t e^{-(T-t)\alpha} + \mu(1 - e^{-\alpha})\sum_{k=1}^{T-t} e^{-(k-1)\alpha} + \sigma^{(T-t)}\left(\int_{T-1}^{T} e^{-\alpha(T-s)} dW_s\right) + \\
&\quad + \cdots + \sigma^{(T-t)}\left(\int_{t}^{t+1} e^{-\alpha(T-s)} dW'_s\right). \qquad (4.34)
\end{aligned}$$

Taking expectations, we have:

$$
\begin{aligned}
F_t &= \mathrm{E}(V_T|v_t, v_{t-1}, \cdots v_0) \\
&= \mathrm{E}\{V_t e^{-(T-t)\alpha}|v_t, v_{t-1}, \cdots v_0\} + \mu(1-e^{-\alpha})\sum_{k=1}^{T-t} e^{-(k-1)\alpha} + 0 \\
&= v_t e^{-(T-t)\alpha} + \mu(1-e^{-\alpha(T-t)}) \\
&= \mu + (v_t - \mu)e^{-(T-t)\alpha}. \qquad (4.35)
\end{aligned}
$$

since the expectation of the stochastic integrals are zero and $V_t = v_t$. Equation (4.35) is the same as (4.33).

To estimate the conditional variance $Var\{V_T|V_{T-2}\}$

$$
\begin{aligned}
Var\{V_T|V_{T-2}\} &= Var[\sigma\left(\int_{T-1}^{T} e^{-\alpha(T-s)}dW_s\right) + \sigma\left(\int_{T-2}^{T-1} e^{-\alpha(T-s)}dW'_s\right)] \\
&= \sigma^2 e^{-2\alpha T}[Var\left(\int_{T-1}^{T} e^{\alpha s}dW_s\right) + Var\left(\int_{T-2}^{T-1} e^{\alpha s}dW'_s\right)] \\
&= \sigma^2 e^{-2\alpha T}[\mathrm{E}\left(\int_{T-1}^{T} e^{\alpha s}dW_s\right)^2 + \mathrm{E}\left(\int_{T-2}^{T-1} e^{\alpha s}dW'_s\right)^2].
\end{aligned}
$$

From the properties of the stochastic integrals ((Øksendal, 1995) p. 26) we write,

$$
\mathrm{E}\left\{\int_S^T e^{-\alpha(T-s)}dW_s\right\}^2 = \int_S^T \{e^{-\alpha(T-s)}\}^2 ds.
$$

Then the variance $Var\{V_T|V_2\}$ is equal to:

$$
\begin{aligned}
Var\{V_T|V_{T-2}\} &= \sigma^2 e^{-2\alpha T}[\int_{T-1}^{T} e^{2\alpha s}ds + \int_{T-2}^{T-1} e^{2\alpha s}ds] \\
&= \sigma^2\frac{1-e^{-2\alpha}}{2\alpha}.
\end{aligned}
$$

and $Var\{V_T|V_t\}$.

Consider the difference of successive forecast $\Delta F_{t+1} = F_{t+1} - F_t$. We calculate the conditional variance $\mathrm{Var}(F_{t+1}|V^t)$ where $V^t = \{v_t, v_{t-1}, \dots\}$ are the past data.

$$
\begin{aligned}
\mathrm{Var}(\Delta F_{t+1}|V^t) &= \mathrm{Var}(F_{t+1} - F_t|V^t) \\
&= \mathrm{E}\{F_{t+1} - F_t - \mathrm{E}(F_{t+1} - F_t|V^t)|V^t\}^2 \\
&= \mathrm{E}\{F_{t+1} - F_t|V^t\}^2 \\
&= \mathrm{E}\{F_{t+1} - \mathrm{E}(F_{t+1}|V^t)|V^t\}^2 \\
&= \mathrm{Var}\{F_{t+1}|V^t\} \qquad (4.36)
\end{aligned}
$$

The variance is equal to

$$\begin{aligned}
Var\{\Delta F_{t+1}|V^t\} &= Var\left\{\mu + e^{-\alpha(T-t-1)}(V_{t+1} - V_t) \mid V^t\right\} \\
&= e^{-2\alpha(T-t-1)} Var\{(V_{t+1} - v_t) \mid V^t\} \\
&= e^{-2\alpha(T-t-1)} Var\{V_{t+1} \mid V^t\} \\
&= e^{-2\alpha(T-t-1)} \sigma^2 \frac{1 - e^{-2\alpha}}{2\alpha}.
\end{aligned}$$

Therefore the variance of successive forecasts increases as the length of the time interval $T - t$ decreases.

Example 4: Trending Ornstein-Uhlenbeck Process

An alternative process is

$$\begin{aligned}
dV_t &= [-\alpha(V_t - \mu t) + \mu]\, dt + \sigma dW_t \qquad (4.37) \\
V_0 &= x.
\end{aligned}$$

$$d[V_t - \mu t] = -\alpha(V_t - \mu t)dt + \sigma dW_t, \qquad (4.38)$$

and get explicit solution:

$$\begin{aligned}
V_t &= xe^{-\alpha t} + \mu \int_0^t s ds + \sigma \int_0^t e^{-\alpha(t-s)} dW_s \qquad (4.39) \\
&= xe^{-\alpha t} + \mu t + \sigma \int_0^t e^{-\alpha(t-s)} dW_s.
\end{aligned}$$

If we compare the data process (4.30) in example 3 with data process (4.39) they differ only in the deterministic part. As a result their forecast are different but not their conditional variances. For the forecast $(T - t)$ steps ahead, we have:

$$\begin{aligned}
F_t = \mathrm{E}(V_T|v_t, v_{t-1}, \cdots v_0) &= \mathrm{E} \cdots [\mathrm{E}(V_T|V_{T-1}, \cdots V_{t+1}, v_t)|v_t, \cdots v_0] \\
&= u_t e^{-\alpha(T-t)} + \mu(T - t). \qquad (4.40)
\end{aligned}$$

4.5 Stochastic Variance Models

An Autoregressive Conditional Heteroskedastic (ARCH) process

Suppose that a financial agent needs to make a decision based on the distribution of a random variable at some future point in time. Apart from the mean of conditional distribution, the agent finds it useful to know about the variance of this conditional distribution. An example would be that he is trying to maximise his mean-variance utility function, and therefore needs to model the conditional variance as a function of the observed data and past variances and since both conditional moments are functions of time, he uses the model to rebalance his portfolio.

Suppose that the observation process is governed by the following process

$$Y_t = \mu + \varphi Y_{t-1} + \epsilon_t, \qquad \forall t, \quad |\varphi| \leq 1 \tag{4.41}$$

where ϵ=(ϵ_t) is a weak white noise (*i.e.*, sequence of uncorrelated random variables with constant mean and variance) satisfying the martingale difference condition:

$$E(\epsilon_t|\epsilon_{t-1}) = 0, \qquad \forall t.$$

The conditional variance $Var(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$ of the process is time dependent through an autoregressive equation as follows:

$$\epsilon_t^2 = c + \alpha\epsilon_{t-1}^2 + u_t \tag{4.42}$$

where $u = (u_t)$ is a strong white noise (*i.e.*, indepedent random variables with constant mean and variance). To ensure the existence of the process some conditions must be imposed (see Gourieroux, 1997, p. 30) these are:

1. The process (ϵ_t^2) must be positive. The sufficient conditions are $\alpha > 0$ and $c + u_t \geq 0$ for any admissible value of u_t

2. The mean of squared innovations is

$$m_t = \mathrm{E}(\epsilon_t^2) = c + \alpha m_{t-1},$$

   where m_0 is given. When $\alpha < 1$, the initial condition can be set up to the equilibrium value $m_0 = c/(1-\alpha)$.

The target is to get a forecast for the value of process at time $t = T$ given the past values $Y^t = \{y_t, y_{t-1}, \ldots\}$. The conditional expectation $F_t = E(Y_T|Y^t)$ is the best approximation of Y_T in the mean square error sense by a function of the past values $Y^t = \{y_t, y_{t-1}, \dots\}$. By writing the Y_T in terms of the most recent innovations, we get:

$$Y_T = \mu\frac{1-\varphi^{(T-t)}}{1-\varphi} + \varphi^{(T-t)}y_t + \sum_{j=0}^{T-t-1} \varphi^j \epsilon_{T-j}. \tag{4.43}$$

Therefore, the forecast F_t is

$$F_t = \mathrm{E}(Y_T|Y^t) = \mu\frac{1-\varphi^{(T-t)}}{1-\varphi} + \varphi^{(T-t)}y_t.$$

Similarly, by using (4.43) and if we recall (Gourieroux (1997) p. 31) that,

$$Var(\epsilon_t|Y^{t-h}) = c\frac{1-\alpha^h}{1-\alpha} + \alpha^h\epsilon_{t-h}^2,$$

then, the conditional variance $Var(Y_T|Y^t)$ is calculated as follows:

$$\begin{aligned}
Var\{Y_T|Y^t\} &= \sum_{j=0}^{T-t-1} \varphi^{2j} Var(\epsilon_{T-j} \mid Y^t) \\
&= \sum_{j=0}^{T-t-1} \varphi^{2j}\left\{c\frac{1-\alpha^{(T-j-t)}}{1-\alpha} + \alpha^{(T-j-t)}\epsilon_t^2\right\} \\
&= \frac{c}{1-\alpha}\frac{1-\varphi^{2(T-t)}}{1-\varphi^2} - \left(\frac{c\alpha}{1-\alpha} - \alpha\epsilon_t^2\right)\left(\frac{\alpha^{(T-t)} - \varphi^{2(T-t)}}{\alpha-\varphi^2}\right).
\end{aligned}$$

Now consider the case $\varphi = 0$. The forecast is $F_t = \mu$, and therefore there is no variability in the forecast. As a consequence of no variability in the forecast,

$$\begin{aligned}
Var(Y_T|Y^t) &= E_t\{Var(Y_T|Y^{t+1})\} + Var_t\{E(Y_T|Y^{t+1})\} \\
&= E_t\{Var(Y_T|Y^{t+1})\}.
\end{aligned}$$

Thus the conditional variance process is a martingale (instead of a supermartingale). This is not suprising since the conditional variance $Var(F_t) = 0$.
Also, the conditional variance of forecast can be compared with the unconditional variance of the forecast as follows:

$$\begin{aligned}
Var(Y_T|Y^t) - Var(Y_T) &= c\sum_{j=1}^{T-t}\alpha^{j-1} + \alpha^{T-t}\epsilon_t^2 - Var(Y_T) \\
&= c\sum_{j=1}^{T-t}\alpha^{j-1} + \alpha^{T-t}\epsilon_t^2 - Var(\epsilon_T) \\
&= c\frac{1-\alpha^{T-t}}{1-\alpha} + \alpha^{T-t}\epsilon_t^2 - \frac{c}{1-\alpha} \\
&= \alpha^{(T-t)}\left[\epsilon_t^2 - E(\epsilon_t^2)\right].
\end{aligned}$$

Comparing the conditional and unconditional variances, the conditional variance will be greater than the unconditional one every time the squared error term ϵ_t^2 is greater than its expected value.

So far, the forecast was considered to be $F_t = E(Y_T|Y^t)$ which was always equal to zero ($F_t = 0$). Now, redefine the forecast as $F_t^* = E(Y_T^2|Y^t)$ which is the variance of ARCH(1) process and is time dependent. So, the forecast is

$$F_t^* = E(Y_T^2|Y^t) = c\sum_{j=1}^{T-t}\alpha^{j-1} + \alpha^{T-t}Y_t^2.$$

We can write the model as a non-Gaussian autoregression

$$Y_t^2 = \sigma_t^2 + (Y_t^2 - \sigma_t^2) = c + \alpha Y_{t-1}^2 + u_t,$$

where $u_t = \sigma_t^2(\epsilon_t^2 - 1)$ is a martingale difference.

The simplest linear ARCH model, ARCH(1) is defined as:

$$Y_t = \epsilon_t\,\sigma_t, \quad \sigma_t^2 = c + \alpha\,Y_{t-1}^2, \quad t = 1, \ldots T,$$

with data $Y^T = (Y_1, \ldots, Y_T)$, where $\epsilon_t \sim NID(0,1)$, with constraints on parameters $c > 0$ and $\alpha > 0$, to ensure that the variance σ_t^2 remains positive for all t. It has a conditional Gaussian representation, that is, $Y_t|Y^{t-1} \sim N(0, \sigma_t^2)$ which means that Y_t is a martingale difference. Therefore, the mean $E(Y_T|Y^t) = 0$ and the predictive distribution is $Y_T|Y^t \sim N\{0, \mathrm{E}(Y_T^2|Y^t)\}$. The forecast $F_t = \mathrm{E}(Y_T|Y^t)$ is zero. The variance of the predictive distribution is $\mathrm{E}(Y_T^2|Y^t)$ which is the forecast of Y_T^2 at time t and for notational convenience denote this with $\mathrm{E}_t(Y_T^2)$. Then, $\mathrm{E}_t(Y_T^2) = E_t\,E_{t+1}\cdots E_{T-1}(Y_T^2)$ and note that $\mathrm{E}_{T-1}(Y_T^2) = c + \alpha Y_{T-1}^2$. Repeating this operation yields

$$\mathrm{E}(Y_T^2|Y^t) = c\,(1 + \alpha + \cdots + \alpha^{T-t-1}) + \alpha^{T-t}Y_t^2 = c\sum_{j=1}^{T-t}\alpha^{j-1} + \alpha^{T-t}Y_t^2.$$

Several properties can be noted here. Consider the conditional variance process $Var(Y_T|Y^t) = \mathrm{E}(Y_T^2|Y^t)$. Also the expected value of the conditional variance $\mathrm{E}_t\{Var(Y_T|Y^{t+1})\}$ is positive. We also know that:

$$E_t\{Var(Y_T|Y^{t+1})\} = Var(Y_T|Y^t) - Var_t\{E(Y_T|Y^{t+1})\}.$$

Consequently,

$$Var(Y_T|Y^t) > Var_t\{E(Y_T|Y^{t+1})\}$$

since $E_t\{Var(Y_T|Y^{t+1})\} \geq 0$. It is concluded that the conditional expected value of the conditional variance of the value of the process at time T is equal to the difference of two variances, that is, the unconditional variance of the process minus the conditional variance of the forecast on the next time $t+1$.

$$\begin{aligned}
Var(Y_T) &= E\{Var(Y_T|Y^t)\} + Var\{E(Y_T|Y^t)\} \\
&= E\{Var(Y_T|Y^t)\} \\
&= E\{Var(Y_T|Y^{t+1})\} \\
&= c\sum_{j=1}^{T} \alpha^{j-1} + \alpha^T Var(Y_0) \quad \text{for } \alpha < 1
\end{aligned}$$

It is concluded that the expected value of the conditional variance process is constant.

4.6 Option Pricing: The Martingale approach

4.6.1 Model Specification for Securities Market

Consider an economy over the time interval $[0, T]$. There are $T+1$ trading dates, these are, $t = 0, 1, \cdots, T$. In this economy, there are only two kind of "assets". Bonds of fixed interest rate r, which will be represented by bank account process and stock(s), the value of which fluctuates randomly.

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a given stochastic process $S_n = \{S_n(t); t = 0, 1, \cdots T\}$ for the n^{th} security $(n = 1, \ldots, N)$ which represents the time t value of the n^{th} security, where points $\omega \in \Omega$ represent states of the world, given a finite sample space Ω with $K < \infty$ elements, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$. The number of risky assets in the Market is denoted by n. For each time $t = 0, 1, \ldots, T$, the investor learns the current state of the world, and therefore the true value of the random variable, namely, the value of risky security at time t. More specifically, the discrete time financial model needs to be built on:

1. A probability measure P on Ω with $P(\omega) > 0$ for all $\omega \in \Omega$.
2. A filtration $\mathbb{F} = \{\mathcal{F}_t; t = 0, 1, \ldots, T\}$, which is a sub-model describing how information is revealed to the investors.
3. A bank account process $B = \{B_t; t = 0, 1, \ldots, T\}$, where B is a stochastic process with $B_0 = 1$ and $B_t(\omega) > 0$ for all t and ω.
4. N risky security processes $S_n = \{S_n(t); t = 0, 1, \ldots, T\}$, where S_n is a non-negative stochastic process for each $n = 1, 2, \ldots, N$.
5. The discounted price process $S_n^* = \{S_n^*(t); t = 0, 1, \ldots T\}$ is defined by $S_n^*(t) \equiv S_n(t)/B_t$, for $t = 0, 1, \ldots, T$ and $n = 1, 2, \ldots, N$.

4.6.2 The general pricing problem and its solution

Contingent Claims

Definition 4.1 A contingent claim or financial derivative is any random-cash flow X of the form $X = \Phi(S_t)$, where S_t is a stochastic variable representing the stock price process and the contract function Φ is some given real valued function.

The holder of the contract receives the stochastic amount X at time $t = T$. The simplest contract is one in which the value of the claim only depends on the value S_T. The alternative is to consider a contract with stochastic payoff which depends on a random path $(S_0, S_1, \ldots, S_T)$ of stock prices as determined by function $f_T = f_T(S_0, S_1, \ldots, S_T)$ at time $t = T$.

The question under consideration is how much an investor should pay at time $t = 0$ for this contingent claim which yields random reward $\Phi(S_T)$ payable at time $t = T$. More generally, the question is, what is a "fair" price at any time $t \leq T$ for the investor to pay so as to receive a random amount $\Phi(S_T)$ at time T. Denote this price process by $\{\Pi(t; X);\ t = 0, \ldots, T\}$.

Trading Strategy (Portfolio Strategy)

Definition 4.2 A trading strategy $H = (H_0, H_1, \ldots, H_N)$ is a vector of stochastic processes $H_n = \{H_n(t); t = 1, 2, \ldots, T\}$, for $n = 0, 1, \ldots, N$ where H_n denotes the number of shares that the investor owns from time $t-1$ to time t. H is a predictable stochastic process, that is $H_n(0)$ is $\mathcal{F}_0$ – measurable and for $n > 1$, $H_n(t)$ is $\mathcal{F}_{n-1}$-measurable.

Portfolio Process

Definition 4.3 The value process $V_t = \{V_n(t); t = 0, 1, \ldots, T\}$ corresponding to portfolio H is a stochastic process defined by

$$V_t = \begin{cases} H_0(1)B_0 + \sum_{n=1}^{N} H_n(1)S_n(0), & t = 0 \\ H_0(t)B_t + \sum_{n=1}^{N} H_n(t)S_n(t), & t \geq 1. \end{cases}$$

The initial value of the portfolio at $t = 0$ is V_0, and V_t is the t time value portfolio just after the"announcement" of new prices at the market is made. Note that $H(t) = (H_0(t), H_1(t), \ldots, H_N(t))$ is already known since $H(t)$ is $\mathcal{F}_{t-1}$-measurable, *i.e.*, $H(t)$ is only dependent to the prices recorded by the investor up to the time $t-1$.

Self-Financing Trading Strategies

Definition 4.4 A strategy is called self-financing if

$$V_t = H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t), \quad t = 1, \dots, T-1.$$

This represents the time t value of the portfolio just after any t transactions, that is just before the portfolio is carried forward to the time $t+1$, *i.e.,* after t time prices are observed.

This implies that,

$$H_n(t)S_n(t) = H_{n+1}(t)S_n(t), \quad n = 0, 1, \dots, N.$$

To understand this restriction consider the portfolio value V_t at an arbitrary time t

$$V_t = H_0(t)B_t + \sum_{n=1}^{N} H_n(t)S_n(t).$$

The value of the portfolio at time $t-1$ is:

$$V_{t-1} = H_0(t)B_{t-1} + \sum_{n=1}^{N} H_n(t)S_{n-1}(t),$$

which represents the market value of the portfolio strategy H_t just after it has been established at time $t-1$. Now, consider the change in the portfolio

$$\Delta V_t = V_t - V_{t-1} = H_0(t)[B_t - B_{t-1}] + \sum_{n=1}^{N} H_n(t)\Delta S_n(t), \qquad (4.44)$$

which means no money is added to or withdrawn from the portfolio between times $t-1$ and t, any change in the value of portfolio process is due to random changes of stocks $\sum_{n=1}^{N} \Delta S_n(t)$ and the change of the bond values ΔB_t only.

Admissible strategies and Arbitrage

Definition 4.5 A strategy H is said to be admissible if it is self-financing and if $V_t \geq 0$ for any $t \in \{0, 1, \dots, T\}$.

This is a restriction which requires the investor to have positive initial wealth. However, a restriction is needed to stop the investor to follow a strategy, if one exists, to making any profit at any time $t > 0$ from zero or negative initial wealth and without investing any funds.

Definition 4.6 An arbitrage strategy is an admissible strategy with zero initial value and non-zero final value.

The market is viable if there are no arbitrage opportunities. To ensure that the market is viable one uses the following theorem.

Theorem 4.1 (Lamberton and Lapeyre (1996))
The market is viable if and only if there exists a probability measure $\mathcal{Q}$ equivalent to $\mathcal{P}$ such that the discounted prices of assets are $\mathcal{Q}$ martingale.

Pricing Principles

Consider a discrete time stochastic model of a frictionless[1] security market as in (4.6.1). The market is assumed to be viable and complete. A contingent claim defined by $X = \Phi(S_t)$ is attainable if there exists an admissible strategy worth X at time T. A market is defined to be complete if every contingent claim is attainable. There is a unique probability measure $\mathcal{Q}$ under which the discounted prices of financial assets are martingales.

Harrison and Kreps (1979) have shown that the absence of arbritrage opportunities implies the existence of a probability measure $\mathcal{Q}$, such that the current price of any basic security is equal to the $\mathcal{Q}$-expectation of its discounted future payments. This result summarises in the following theorem which relates the absence of arbitrage possibilities with the $\mathcal{Q}$-martingale probability measure as is presented by Kabanov and Kramkov (1994) (see Harrison & Pliska, 1981):

Theorem 4.2 (Kabanov and Kramkov (1994))
The following conditions are equivalent:

(i) there is no arbitrage;

(ii) there is probability measure $\mathcal{Q} \sim \mathcal{P}$ with bounded density $d\mathcal{Q}/d\mathcal{P}$ such that the price process S is (a vector-valued) martingale with respect to $(\mathbb{F}, \mathcal{Q})$, *i.e.*, $\boldsymbol{E}_{\mathcal{Q}}|S_n| < \infty$ and $\boldsymbol{E}_{\mathcal{Q}}(\Delta S_n|\mathcal{F}_{n-1}) = 0$ for all $n \leq T$.

The idea is that the change in financial value (gain or loss) of the replicating portfolio during the time period $(t-1, t)$ as shown by (4.44) is the product

[1] This means that there is no trading cost.

$H(t)\Delta S(t)$. So the accumulated gain from time $t = 0$ up to time t is given by the "martingale transform":

$$\sum_t H(t)\Delta S(t).$$

Therefore we have the definition of a gain process is:

Definition 4.7 A Gain process $G = \{G_t; t = 1, 2, \dots, T\}$ is the stochastic integral of the trading strategy H with respect to the price, that is

$$G_t \equiv \sum_{u=1}^{t} H_0(u)\Delta B_u + \sum_{n=1}^{N}\sum_{u=1}^{t} H_n(u)\Delta S_n(u), \quad t \geq 1.$$

Now, returning to the question what is a "fair" price of the contingent claim with payment function f_T, its solution is summarized by the following theorem:

Theorem 4.3 (Shiryaev, Kabanov, Kramkov, and Melnikov (1994))

1. Under conditions of a Black-Scholes market (Black & Scholes, 1973), the fair price C_T of a contingent claim with expiration time T, payment function $f_T = f_T(S_0, \dots, S_T)$, and use of self-financing strategies is given by $C_T = E^*\left[(1+r)^{-T} f_T\right]$ where E^* is the expectation with respect to the probability measure $\mathcal{Q}$ which makes the discounted price process $\mathcal{Q}$ martingale.

2. There exists a self-financing strategy $H^* = (H_t^*)_{0 \leq t \leq T}$ such that evolution of the corresponding capital is given by the formulas $X_t^H = E^*[(1 + r)^{-(T-t)} f_T | \mathcal{F}_t]$.

4.7 Option pricing and the Forecasting Systems

In the previous section, we discussed the fundamental theorem of Asset pricing as initiated by Harrisson and Kreps (1979) and Harrisson and Pliska (1981). The absence of arbitrage opportunities implies the existence of a probability measure $\mathcal{Q}$. The current price of any basic security is equal to the the $\mathcal{Q}$-expectation of the discounted future payments. With the restriction that markets are complete, $\mathcal{Q}$ is unique.

The reason why the martingale method should be of interest is that the current price of any derivative asset (European option) is given by the discounted expected future pay-off under the "appropriate" probability measure. In other words the option price is a sequence of random variables, which are conditional expectations.

Alternatively, a forecaster can predict the exact value of payoff $X = \Phi(S_T)$, if he can produce a series of forecasts about the value of the random variable S_T given the information up to time t, and then use this forecast to estimate $X = \Phi(S_T)$.

For example a European call option has payoff equal to $h = (X_T - K)^+$ where h is its fair price at time $t = T$ if $V_T = X_T$ and K its exercise price. So if the forecaster can predict and knows the payoff function h, then he can give a series of predictions about the future price of the contingent claim.

However, a financial Statistician who would like to price an option adopts the so-called martingale procedure, where he knows the stochastic process followed by share prices. The key issue is that he is looking for the unique probability measure $\mathcal{Q}$ under which the discounted expected stock price is a martingale. The idea of the martingale approach is that the value of the derivative is estimated by calculating the expected value of the pay-off with respect to the risk -neutral probability measure, that is the unique measure that makes the discounted share process martingale.

Having found the risk-neutral probability measure $\mathcal{Q}$, the Statistician can calculate the price of his derivative denoted by Y_t at time t, as conditional expectation of the terminal payoff X

$$Y_t = B_t E_{\mathcal{Q}}(B_T^{-1} X | \mathcal{F}_t)$$

where X is the payoff function of the contingent claim, $\mathcal{Q}$ is the martingale measure for the discounted stock price, B_t is the discounted factor at time t and $\mathcal{F}_t$ is the information structure of the share prices available up to time t. The forecaster is able to predict the actual V_T, of the share and then he only needs the payoff function of the derivative to do his pricing.

Chapter 5

Real Options and R&D projects

5.1 Introduction

R&D projects are characterised by long planning horizon, large investments usually in lumpy and sequential outlays, high uncertainty and high rewards when they are completed successfully. One can think of two types of uncertainty. The first type relates to the effort and time needed to complete an R&D project. The second type of uncertainty is due to the dynamic structure of the Market and this affects the total development cost and the reward of a project.

The purpose of this chapter is to study the impact of Market uncertainty in the value of R&D projects, which is contingent on future market events with unknown effect on the value of the project, when the investment initiates. While the uncertainty will be resolved, managerial decision might need to be taken with regard to its development before its completion that is not in the original plan and becomes available to the manager just after Market uncertainty is partially resolved. An example is to postpone either temporarily or permanently the development after phase II for a certain drug, if stopping proved more economically viable than the continuation of the drug development.

The methodology used to evaluate the impact of Market uncertainty is the application of the theory of Financial option pricing as applied to real investments such as to undertake an R&D project. This area is known as Real options. Recall that in financial option pricing, the option value is derived by the ability to choose in the future whether to exercise the option or not, depending on the stock price which will then be observed. One can say that real options arise due to uncertainty in costs, benefit, and opportunities to favorably alter the project course contingent on future information.

Marketing uncertainty modelling may be crucial, since this decision flexibility may be important for the manager and therefore has a value in itself which needs to be identified, quantified and added to the already known value of a given project as calculated by any Productivity Index evaluation method. One example of such index is the Pearson Index, as will be explained later in this introduction. This managerial flexibility might give an additional value to the project, which is derived by the fact that manager can revise operating decision in response to market conditions.

If the manager does not make risky decisions to develop drugs with uncertain future payoff and development costs in order to gain competitive advantage in different markets, the company might not be able to survive in the long run. The importance of evaluating the impact of Market uncertainty and the associated decisions is to provide a better aid to the manager who is required to make decisions how to build a portfolio of research projects.

Initially a conceptual framework of Real options is given and then an explanation of how the theory of option pricing could be applied to a valuation of risky R&D projects is presented. Thus, Option pricing models are developed. This analysis will be done with connection to the existing valuation tools, namely Pearson index.

5.2 R&D Project as a Real Option

A key characteristic of R&D projects is their multistage structure. The implications of the multistage structure make an R&D project a highly risky investment. Each stage of an R&D project must be completed successfully and once a Pharmaceutical company makes the decision to start an R&D project, this decision is associated with future liability (additional funds for production capacity, Marketing research etc.). However, the information that will gradually resolve the uncertainty of an R&D project is only available at discrete points in time. The development of an R&D project can be thought of as Investment expenditures with irreversible cost, that is there is a sunk cost associated with this investment (Dixit & Pindyck, 1995).

In real options theory sequential investments are analogous to compound options, that is an option written on another option. The reasoning is that each stage completed of the project gives the firm the option to complete the next stage. If one had no choice but to complete the project once it had been started, investing would involve only a single decision.

Many examples in Dixit and Pindyck (1994) as well as in Smith and Nau (1995) are related to the option to defer an R&D investment (timing option), with the intention that this investment will start next year or sometime in the future, the option pricing technique remains the same for any kind of option such as abandon, growth, temporarily shut down etc (see, Trigeorgis, 1996, Chapter 1).

Historically, Kester (1984) and Myers (1984) introduced the theory of real options which takes market uncertainty and managerial flexibility of projects into account in their evaluation procedure. Managerial flexibility is the future actions or decisions management will apply to project management that are contingent on future events, these future events become known at the discrete points in time where the uncertainty is resolved. For example if a manager realises that the actual cost is significantly higher than the anticipated cost so far, he might want to abandon the project or put it on hold. The option to abandon the project if new information is unfavourable is created due to the multistage structure of an R&D project and each stage can be view as an option to the value of the next stage.

Pearson Index

The Pearson index evaluation method is dependent on the expected reward and expected development cost of a project. These expectations are related to the probabilities of success of each stage of the R&D project and not to the uncertainty related to the magnitude of the final rewards (and costs). It means that there is no variability to final reward. It is a well-established fact in option pricing theory that the option value is created due to variability.

To illustrate better the variability issue, suppose that there are two projects where the final rewards have the same distribution with equal mean but different variance. Also, available information suggests that their costs, which are random variables, have the same distribution with equal means and different variance. The Pearson index would give the same value for both projects provided equal probabilities of success for each equivalent stage for both projects.

Suppose that a project consists of two stages. We assume that the final reward R is equal to the sum of two independent random variables R_1 and R_2 where R_1 is observed at the end of the first stage and R_2 is known at the completion of the second stage. R_1 and R_2 take the values,

$$R_1 = \begin{cases} \bar{R}_1 + \delta_1 & \text{with probability } 1/2 \\ \bar{R}_1 - \delta_1 & \text{with probability } 1/2 \end{cases}$$

$$R_2 = \begin{cases} \bar{R}_2 + \delta_2 & \text{with probability } 1/2 \\ \bar{R}_2 - \delta_2 & \text{with probability } 1/2, \end{cases}$$

where $\bar{R}_1, \bar{R}_2, \delta_1, \delta_2$ are constants and $\delta_1 > \delta_2$. Thus, the final reward R is a random variable with mean value $\bar{R}_1 + \bar{R}_2$ and variance $\delta_1^2 + \delta_2^2$. Stage one costs C_1 and stage two costs C_2. One criterion is to say, that we do not accept the project when mean reward R is less than the cost, that is

$$E[R] = \bar{R}_1 + \bar{R}_2 < C_1 + C_2. \tag{5.1}$$

This standard cost-benefit criterion is used if R&D is not seen as a sequential decision process. For example suppose that it is optimal to pay C_1 and proceed, then if R_1 turns out to be $\bar{R}_1 + \delta_1$, we proceed to the second stage and stop otherwise. In this case the criterion has the form:

$$-C_1 + 1/2\{-C_2 + 1/2(\bar{R}_1 + \bar{R}_2 + \delta_1 - \delta_2) + 1/2(\bar{R}_1 + \bar{R}_2 + \delta_1 + \delta_2)\} > 0,$$

which can be written as

$$\bar{R}_1 + \bar{R}_2 + \delta_1 > 2C_1 + C_2. \tag{5.2}$$

For large value of $\delta_1 > C_1$, it is still optimal to start this R&D project even though the standard criterion (5.1) does not suggest so. It is the possibility of termination at the end of the first stage if random reward R_1 turns out to be in its low level which makes the decision to start the project valuable.

Define *strategic options* as opportunities that are made available in the future by undertaking a project but are not part of the initial project plan. Uncertainty creates these strategic options. The Pearson index averages all possible outcomes (success and failures) without gaining any insights into the managerial flexibility. It presupposes the two assumptions of discounted cash flow techniques, such as, future cash flow are replaced by their expected value and treated as given at outset. The second assumption is that how risky a project appears to be, is determined by the assumed discounting rate that is constant throughout the whole R&D period. This is unrealistic because the level of risk associated with projects needs not remain constant over time. A given project does not have the same level of riskiness on its first stage as in its last stage.

The valuation procedure of an R&D project is based on the analogy between financial option and strategic or operating options. Option valuation procedure is considered to be a good choice if the analysis is intended to estimate the market value of a project or decision if the underlying asset value known accurately.

Real options investment analysis uses the insights of Black and Scholes (1973) formula of pricing a European option. The starting point to the financial option pricing is the asset on which the option is contingent, usually called the underlying asset (Share). In the real option case, the present value of expected cash flow is equivalent to the current value of stock. A consequence of this equivalence of share and project is that stock value uncertainty represents project value uncertainty. A portfolio of assets is constructed that has exactly the same payoff as the option (investment opportunity) in all states of the world. This portfolio consists of an underlying asset, that is a perfectly correlated share with the undeveloped project or with any other drug that is in market already and it is believed that the undeveloped project would behave exactly as the project already in the market.

5.2.1 Option based models

Option based models may be used to value investment opportunities when future market conditions are uncertain. The option pricing technique does not predict future values of the underlying asset, but future values are assumed to follow a well defined stochastic process. For example, the market value of a project at time t, $V(t)$ is assumed to be uncertain, and to evolve over time according to an Ito process (Arnold, 1971)

$$dV(t) = A\{t, V(t)\}dt + B\{t, V(t)\}dW(t), \qquad A\{t, V(t)\}), B\{t, V(t)\} \in \mathbb{R},$$

where $A\{t, V(t)\}$ and $B\{t, V(t)\}$ represent the instantaneous rate of return and standard deviation of the project value $V(t)$ process, respectively. $W(t)$ is one dimensional Wiener process.

In financial option pricing, some assumptions should be met, these are: frictionless markets for stocks, bonds and options which means there is no transactions costs, participants can take out long and short positions without any constraint and tax can be ignored and that markets are arbitrage-free. We also impose these assumptions in the real option valuation method.

One of the early real option models which has an analytic solution is the optimal timing problem of undertaking an investment which is also known as the option to defer.

Optimal timing of investment in an irreversible project

McDonald and Siegel (1986) considered the following problem. They assumed that the value of a project follows a Geometric Brownian motion which implies that the current value of the project is known, and the futures values are log-normally distributed with a instantaneous standard deviation σ and α as the instantaneous expected return on the project. The project value $V(t)$ process is given by:

$$dV(t) = \alpha V(t)dt + \sigma V(t)dW(t),$$

and represents the present value of expected future cash flow conditional on undertaking the project with present value V_t. The firm is allowed to undertake such an investment opportunity up to any time $t \leq T$, where T is the expiration time of the opportunity. Given a fixed cost C necessary to take the project, they calculate the values of V_t^* that exceeds C such that when $V_t \geq V_t^*$ it would be optimal to make the investment , and otherwise defer. The investment opportunity value is defined to be the expected present value of the payoff at the first passage time $\acute{t}$, that is, the first time at which $V_t \geq V_t^*$ first reaches the boundary:

$$X(T) = E_0\{e^{-\mu\acute{t}}[V_{\acute{t}} - C]\}$$

where $X(T)$ is the time zero value of an investment opportunity that expires at time T with discount rate μ. They also study the special case of a perpetual investment opportunity, *i.e.*, $T = \infty$ which means that the opportunity to invest is infinitely lived. In this case the boundary is not dependent on time. The value of the opportunity is given by

$$X = (V^* - C)\left(\frac{V}{V^*}\right)^b$$

where

$$V^* = C\left(\frac{b}{b-1}\right), \qquad b \equiv (1/2 - \hat{\alpha}/\sigma^2) + \sqrt{(\hat{\alpha}/\sigma^2 - 1/2)^2 + 2r/\sigma^2}$$

and $\hat{\alpha} = r - \delta$ which is the difference of risk-neutral rate r and δ which may represent the opportunity cost of delaying completion of the project. Based on this analysis and a simulation related to this specific valuation, they concluded that it may be optimal to wait until benefits are twice the investment cost and then invest.

5.2.2 An example

Consider two projects which are alternatives. It is assumed that each project will be completed successfully at the end of the time period $[0, T]$, that is, if a project is started at time $t = 0$ will be completed at time $t = T$ and reward $R(T)$ will be gained. Denote a sequence of numbers $R(t)$ for time $0 < t < T$ where $R(t)$ is the forecast for the reward to be gained at time $t = T$ as can be predicted at time t. The total development cost of each project is fixed and equal to C. The only difference between the two projects is that for the project 1 the investor pays the total cost C immediately whereas for the project 2 the total cost could be paid in any arbitrary number n instalments. The instalments could be paid at any time in the time interval $[0, T]$. The discounting factor could be excluded from the analysis only for the cases where the total cost is paid as a series of finite instalments. (Otherwise the total cost C is paid as continuously payable annuity for finite time period in such way its present value is equal to C.) The reward to be gained is a random variable (unknown yet) and only its expected value is known. In the evaluation procedure the expected value of the random gain is taken into account. It is also expected that more information with regard to reward process will be available to the manager sometime in the future and the opportunity for the manager to change its strategy plan will still be accessible.

The project manager should recognize that operating flexibility arises when the project is developed in the second way, that is, the development can be stopped, so saving the costs that otherwise would be spent to the completion of an unsuccessful project, provided this information will be available before the payment of the last instalment. Managers should appreciate the role of uncertainty in a random environment, which is due to unknown information, not available yet to the manager, with regard to the parameters needed to be known for the evaluation of a project. These parameters might be the development cost, benefits, time needed to complete a project. In such an uncertain environment a manager often faces difficulties in assessing a project because of the incomplete knowledge of these parameters and the decision whether the project is still worth continuing can not be proved to be correct or false before the uncertainty is resolved. When the uncertainty is resolved, the already assumed scenario might turn to be more (or less) valuable than its initial estimation. To explain this, suppose that the expected present value of cash flow from a similar completed project at time t is $R(t)$. It is known that after certain time h the forecast $R(t+h)$ can either take value $uR(t)$ or $dR(t)$ with probabilities p and $1-p$ respectively in such way that $dR(t) < R(t) < uR(t)$, where u and d are positive constants ($d < 1 < u$). If the actual forecast for the cash flow at time $t+h$ is $uR(t)$ the project is considered as a successful one and continuation of this project would be profitable action. Otherwise the project is a failure (not profitable) and its continuation is not justified. When a manager uses the expected value at time t, $R(t)$, as an estimate of the future cash flow, the project value will be an underestimate of the actual value of a project at the completion time given success. Similarly, when the project turns out to be a failing one, $R(t)$ will give a false estimation again.

Therefore, the expected scenario of cash flow technique is inadequate to value a project with the above characteristics.

The Option pricing method could in principle value a project of this form. Denote the value of an option (could be any option) at time t by $V(R(t), t)$. In financial option pricing the current value of the underlying traded asset (share) is known in contrast to the case of an undeveloped project where its present value is stochastic as it may be related to the estimate of future profits. Otherwise, $R(t)$ can also be thought of as the expected present value of a cash flow from a similar completed project at time t which is the market value of the equivalent asset. The value of the option at time $t+h$, $V(R(t+h), t+h)$ will be either equal to:

$$f_{\text{up}} = V(uR(t), t+h)$$

or

$$f_{\text{down}} = V(dR(t), t+h)$$

where the words "up" and "down" indicate whether the market conditions is favourable or not respectively. f denotes the option value or the claim value time-process (Baxter & Rennie, 1994, p. 28). In other words the uncertainty of the cash flow is resolved and this uncertainty creates the option value which can be calculated by constructing a hedge portfolio at time t which is guaranteed to have the same value at time $t+h$ independently of the value of the underlying asset. The assumption that there is no transaction cost and that the pricing is done in an arbitrage-free world is necessary (Baxter & Rennie, 1994, Chapter 1&2). The option pricing theory says that there must be a risk free security, which earns the risk free interest rate denoted by r. To replicate the value of the option one constructs a portfolio with ϕ units of the underlying asset and ψ units in the risk free asset:

$$\phi uR(t) - \psi R(t)\exp(rh) = V(uR(t), t+h) = f_{\text{up}}$$

$$\phi dR(t) - \psi R(t)\exp(rh) = V(dR(t), t+h) = f_{\text{down}}$$

with ϕ and ψ to be unknown. Solving the system of the two equations, gives

$$\begin{aligned}
\phi R(t) &= \frac{V(uR(t), t+h) - V(dR(t), t+h)}{u-d} & (5.3)\\
\psi R(t) &= \frac{dV(uR(t), t+h) - uV(dR(t), t+h)}{u-d}\exp(-rh). & (5.4)
\end{aligned}$$

The value of portfolio at time t is:

$$\begin{aligned}
V(R,t) &= \phi uR(t) - \psi dR(t)\\
&= \frac{1-de^{-rh}}{u-d}V(uR(t), t+h) + \frac{ue^{-rh}-1}{u-d}V(dR(t), t+h)\\
&= p_u e^{-rh}V(uR(t), t+h) + p_d e^{-rh}V(dR(t), t+h),
\end{aligned}$$

where:

$$p_u = \frac{e^{rh}-d}{u-d} \quad \text{and} \quad p_d = \frac{u-e^{rh}}{u-d},$$

are the so called "risk-neutral" probabilities.

In the decision theory approach, the probabilities represents the decision maker's belief about the uncertainties of the project and the future cash flows are assessed in terms of the utility function. Let $x(s)$ denote an individual's wealth level as a function of state s. This individual has subjective probability $p(s)$ for state s and utility function $u(x)$ for wealth x. Let $u'(x)$ denote individual's marginal utility function evaluated at x which is the derivative of individual's utility function at x. One can multiply the probability density function $p(s)$ by the function $u'(x)$ which yields a positive function which can be normalized to yield another probability density function. Denote this probability by $\pi(s)$, we have:

$$\pi(s) \propto p(s)u'(x(s)). \qquad (5.5)$$

The distribution π is the individual's risk neutral probability distribution, because the decision maker evaluates small gambles as though he was risk neutral with probability distribution π, when in fact he is risk averse with probability p.

5.2.3 The no-arbitrage condition

We mentioned the term arbitrage in section (4.6.2). In this section, we discuss the basic duality theorem from linear algebra which connects the "no-arbitrage" principle with the existence of "risk-neutral" distributions. We assume an enviroment with a set of agents, a set of events (states of nature), money, a contingent claim market and an outside observer. An *arbitrage opportunity* is a collection of acceptable contracts that wins money for the observer and loses money for the agents in the aggregate in at least one possible outcome of events, with no risk or loss for the observer in any outcome.

Consider the following theorem which enables us to explain the use of the risk-neutral probabilities in asset pricing. In linear algebra this theorem is known as duality theorem (Ostaszewski, 1990):

Theorem 5.1 Let $\mathbf{X}$ be an $m \times n$, let $\boldsymbol{\alpha}$ be an m-vector, let $\boldsymbol{\pi}$ be an n-vector, and let $[\alpha' X](s)$ and $\pi(s)$ denotes the s^{th} elements of $\boldsymbol{\alpha' X}$ and $\boldsymbol{\pi}$, respectively. Then exactly one of the following systems of inequalities has a solution:

1. $\boldsymbol{\alpha} \geq \mathbf{0},\ \boldsymbol{\alpha' X} \leq \mathbf{0}, [\alpha' X](s) < 0$

2. $\boldsymbol{\pi} \geq \mathbf{0}, \boldsymbol{X\pi} \geq \mathbf{0}, \pi(s) > 0.$

This theorem says that either the system of equations $\boldsymbol{\alpha' X} \leq \mathbf{0}$ has non-negative solution or else the system $\boldsymbol{X\pi} \geq \mathbf{0}$ has a semi-positive solution. For example, a column of the matrix $\mathbf{X}$ corresponds to a state-contingent asset. The rows of $\mathbf{X}$ may associated with gambles or trades that agents have offered to accept. If $\boldsymbol{\alpha' X} \leq \mathbf{0}$ has a solution, then one can construct an arbitrage opportunity in state s, by choosing the weighted sum of gambles or trades in $\mathbf{X}$, that is, $\boldsymbol{\alpha' X}$. If $\boldsymbol{X\pi} \geq \mathbf{0}$ has a solution, then there is a vector $\boldsymbol{\pi}$ of probabilities (or prices states) which assigns to all the gambles or trades that have been accepted, non-negative expected value or profit, and in which state s has strictly positive probability or asset.

Suppose that we have a market for assets that pay off in different states of nature. Denote by $s = 1, 2, \ldots, S$ the finite number of available states and assume that there are A different assets. Let's denote the payoff of asset α in state s by $R_{s\alpha}$. An asset is described by a vector giving its payoff in each of the

S states of the nature. The payoff matrix is

$$R = \begin{pmatrix} R_{11} & \cdots & R_{1A} \\ \vdots & \ddots & \vdots \\ R_{S1} & \cdots & R_{SA} \end{pmatrix}.$$

Thus the i^{th} security is described by the vector $(R_{1i} \cdots R_{Si})$. If we hold amount x_α of asset α and we choose a portfolio $x = (x_1, \ldots, x_A)$ then one receives wealth w_s in states s equal to $\sum_{\alpha=1}^{A} x_\alpha R_{s\alpha}$. Let the price of asset α be denoted by p_α and let $p = (p_1, \ldots, p_A)$ be a vector of asset prices. The cost of a portfolio $x = (x_1, \ldots, x_A)$ is given by $px = \sum_{\alpha=1}^{A} p_\alpha x_\alpha$.

One important kind of assets are Arrow-Debreu securities which promise to deliver one unit of purchasing power in the specified state. Thus the payoff pattern is of the form $(0, \ldots, 1, \ldots, 0)$, where the 1 occurs in the location s. Let π_s be the price of the Arrow-Debreu securities that pays 1 if state s occurs and let R_{sa} be the value of asset α in state s. Then the equilibrium price of asset α must be given by

$$p_\alpha = \sum_{s=1}^{S} \pi_s R_{sa}.$$

The no arbitrage condition (Hodges, 1998) can be expressed in terms of structuring the portfolio with minimum cost in order to obtain a portfolio with non-negative payment:

$$\begin{array}{ll} \text{Minimize} & \mathbf{p}'\mathbf{x} \\ \text{such that} & \mathbf{Rx} \geq \mathbf{0} \\ & \mathbf{x} \text{ unrestricted,} \end{array}$$

where the components of $\mathbf{x}$ are unconstrainted in sign. The dual Linear program problem is

$$\begin{array}{ll} \text{Maximize} & \mathbf{0}'\boldsymbol{\pi} \\ \text{such that} & \mathbf{R}'\boldsymbol{\pi} = \mathbf{p} \\ & \boldsymbol{\pi} \geq \mathbf{0}. \end{array}$$

The dual program suggests that the primal problem is equivalent to the existence of non-negative pricing vector $\boldsymbol{\pi} \geq 0$ which explains the price of each security according to its state-contingent cash flow $\mathbf{p}' = \boldsymbol{\pi}'\mathbf{R}$. As a consequence, a pricing vector $\boldsymbol{\pi} \geq \mathbf{0}$ exists if and only if there is no arbitrage.

So far, we have shown that the concept of risk-neutral probabilities is used in option pricing process. In the next section, we study an option pricing example and we calculate the expected value of different investment opportunities using both option pricing "tools" and decision analysis method.

5.3 R&D funding as a sequential decision

Option to contract the scale of an R&D project

In this section, we consider the decision problem to change the scale of production of an R&D project. Suppose that the market conditions are less favorable than originally expected. In this case, it may be possible and desirable for the company to reduce the scale of operation by $x\%$, and as result part of the initially planned investment outlays will not be spent and presumably the investment will become less unprofitable. This flexibility to mitigate loss is an analogous to an American put option on $x\%$ part of the base-scale project. The exercise price equals the potential cost savings I_c. Because the contraction option gives the project manager the right to reduce the operating scale, if market conditions turn out to be unfavorable, a project that can be contracted is worth more than the same project without the flexibility to contract. Thus we can say that the expected profit is equal to the sum of expected profit without the option value plus the option value:

$$\text{Exp. Profit} = \text{Exp. Profit without Flexibility} + \text{Value of Flexibility}. \tag{5.6}$$

Consider the following investment problem where there are two states of the world, these are "favorable" and "unfavorable". The project manager has an opportunity to invest in an R&D project in a new plant whose future cash flow depends on an uncertain state of the world. The decision maker assess that the favorable state occurs with probability p, and the unfavorable state with probability $1 - p$. Once a project is undertaken, the project manager may have the flexibility to alter it in a various ways at different times during its life. Let us suppose the decision maker wants to have right to contract the scale of the project's operation by factor $x\%$. Let $x\% = 50\%$. This means that they reduce the scale production to the half and they pay a cost $I_1^{''}$ which is less than the 50% of the original cost. Thus the reduction in cost is $I_1^{'} = I_1 - I_1^{''}$. Let the gross value of the project be denoted by V. Then one can compare the net reward $V - I_1$ of the decision without the option with the reward received given the option to contract has been exercised as follows:

$$\max(V - I_1, 0.5V - I_1^{''}) = (V - I_1) + \max(0, I_1^{'} - 0.5V). \tag{5.7}$$

Consider the capital budgeting problem which is represented by the decision tree below

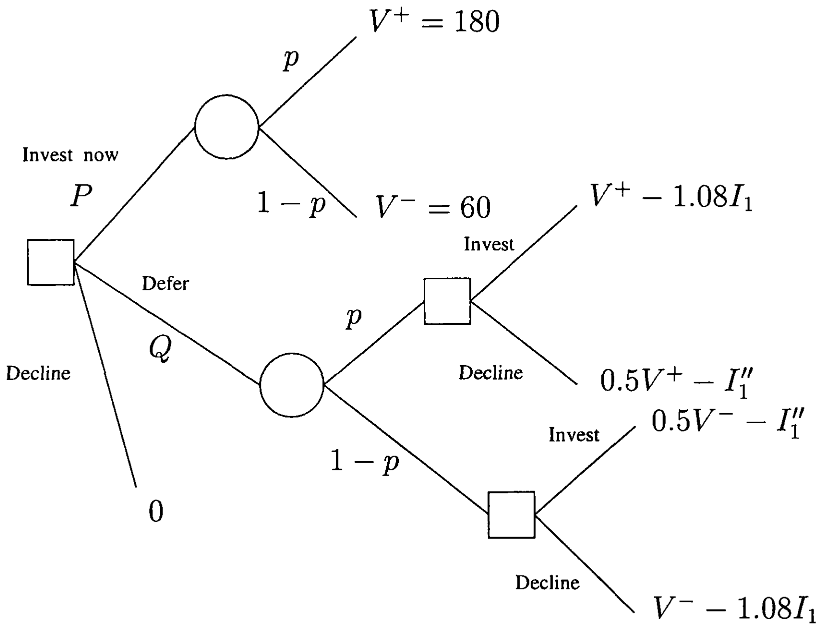

Figure 5.1: Decision tree with an option to contract

The firm can pay P for the first year and get return 180 in the favorable state or 60 in the unfavorable state. Alternatively, the firm can pay Q for a license which allows them to defer their decision to reduce the scale of production until the state of nature becomes known. If the firm chooses this option, it can invest $1.08P$ and receive the known reward, where the factor of 1.08 reflects a risk-free interest rate of 8%.

In decision analysis based on stochastic dynamic programming, the discount rate accounts for the time value of money and a premium for risk. The difficulty is to choose the appropriate discount rate and usually is adapted by the policy of the company. Otherwise it can be chosen by using the capital asset pricing model (CAMP) (Trigeorgis, 1996, Chapter 2). The other option is not to invest and reject the opportunity to develop the R&D project which yields zero cash flow.

The Option approach

To apply the technique of contingent claim analysis, we assume that there is an asset which is usually called the "twin" security whose returns are believed to be perfectly correlated with the returns of the project under consideration. More specifically, there must be some linear combination of traded assets that has proportional cash flow in all states of the world to the hypothetical state dependent cash flow of the investment opportunity under evaluation.

Suppose that one share of the security of the correlated asset worth 36 in the good state and 12 in the bad state. If p is the probability of the good state, then the market rate of return on this security is r, which is equal to $r = (36(p) + 12(1-p))/S$ where S is the current price of share. If the current price $S = 20$ and $p = 0.5$, the return is equal to 1.2, meaning that the project with this level of risk should earn a 20% return. The option of investing immediately yields expected net present value of $(180(p)+60(1-p))/r - P = 5S - P = 100 - P$. If $P = 104$ then the expected net present value is equal to -4.

For the option to contract scenario, the total cost P is split up into an initial installment Q and another installment with present value equal to I_1, *i.e.,* $Q + I_1 = P$. We continue to assume that $p = 0.5, S = 20$ and $P = 104$. The question is for what value of Q the option to contract will be preferred to the no option situation. Given the favorable state the decision maker will invest at the same level of production with return $V^+ - 1.08 I_1 = 180 - 1.08 \cdot 54 = 121.68$ and thus he will not exercise the option to contract . In the bad state, the optimal decision is to exercise the option for a return $0.5V^- - I_1'' = 0.5(60) - 25 = 5$ which is greater than the value of the decision not to contract. If we discount these values back to the present at 20% yields an expected net present value of $((121.68(0.5) + 5(0.5))/1.2) - Q = 52.78 - Q$. The option to contract should be chosen if $Q < 52.78$.

The contingent claim analysis approach would discount cash flow at the risk-free rate and taking expectations with respect to risk-neutral probabilities that can be estimated from market prices. For our example, the risk neutral probabilities are π_1 for favorable state and π_2 for non-favorable state. These probabilities can be found if we equate the price of twin security with its expected return discounted at the risk free discount rate:

$$\begin{aligned} \frac{(36(\pi_1) + 12(1 - \pi_1))}{1.08} &= 20, && (5.8) \\ \pi_1 + \pi_2 &= 1. && (5.9) \end{aligned}$$

Therefore, $\pi_1 = 0.4$ and $\pi_2 = 0.6$. If we use the risk-neutral probability to value the opportunity to invest without the option to contract we get its expected net present value

$$\frac{180(0.4) + 60(0.6)}{1.08} - P = 100 - P \qquad (5.10)$$

which is equal to -4 if P is chosen to be 104. The opportunity to invest with the option to contract and initial cost Q has expected net present value:

$$\frac{121.68(0.4) + 5(0.6)}{1.08} - Q = 47.84 - Q. \qquad (5.11)$$

The value of the option is positive if $Q < 47.84$. Set $Q = 50$ as the initial outlay. Then the value of the option to contract is

$$\begin{aligned} \text{Option value} &= \text{Exp. Profit} - \text{Exp. Profit without Flexibility} \\ &= -2.16 - (-4) = 1.84. \end{aligned}$$

The option pricing approach proceeds to solve the problem by creating a portfolio of observable securities whose prices and required rates of return are known and whose payouts exactly mimic the payouts of the decision tree. Using a no-arbitrage argument we can construct a replicating portfolio of n shares of the twin security and selling bonds of value B. We have the following two equations

$$\begin{aligned} \text{Good state:} \quad 36n - (1+r)B &= 121.68 && (5.12) \\ \text{Bad state:} \quad 12n - (1+r)B &= 5, && (5.13) \end{aligned}$$

where $r = 0.08$. The solution to the system of equations is $n = 4.86$ and $B = 49.39$ *i.e.,* the replicating portfolio consists of borrowing 49.39 and buying 4.86 shares of the twin security. If we buy 4.86 shares of the twin security and sell 49.39 worth of risk-free 8% bonds will exactly replicate the payoffs of the contract option in the last period and the cost of this transaction is $4.86(20) - 49.38 \approx 47.81$ which is less than the value 47.84 given by simple decision analysis. In order to get the answer 47.81 using decision analysis we should have used r^* as a discount rate given by

$$\frac{121.68(0.5) + 5(0.5)}{1 + r^*} = 47.81$$

which is equal to 32% and 20% for the invest without the option to contract alternative.

To apply decision analysis, the decision maker's utility can be considered as function of the terminal wealth at the end of the second year. Suppose that it is of the exponential negative form *i.e.,* $u(x) \propto 1 - \exp(-x/200)$. This utility function has the constant absolute risk aversion property, and thus the optimal investment decision is independent of the initial wealth level. To determine the number of shares s we use (5.5) and the data as follows and we solve:

$$\frac{\pi}{1-\pi} = \frac{p}{1-p} \frac{u'((36 - 1.08(20)s))}{u'((12 - 1.08(20)s))} \tag{5.14}$$

where $u'(x) \propto \exp(-x/200)$ is the marginal utility of money at wealth level x. Using the values $\pi = 0.4$ and $p = 0.5$, we get $2/3 = \exp(-24s/200)$, $s = -(200/24)\ln(2/3) = 3.38$ shares.

Now, consider an option to contract the scale of a project for an initial payment Q, assuming that $P = 104$. The payments for this option is either 121.68 or 5 which can be replicated by buying 4.86 shares of twin security and selling 49.39 worth of risk free bonds paying 8%, for a year outlay 47.81. If $Q > 47.81$ the option to contract is less attractive and the decision maker will prefer the investment decision without the option to contract. If $Q < 47.81$ an arbitrage opportunity exists for the decision maker by choosing the option to contract and divesting 4.86 shares of the twin securing, ending up with $3.38 - 4.86 = -1.48$ shares and as result keeping the same terminal wealth while the initial wealth increases.

5.4 Discussion

The previous analysis allow us to compare the dynamic programming approach and the option pricing approach. Dynamic programming makes use of the subjective probability of state to occur as the individual assess it whereas the option pricing is based on the risk-neutral probabilities.

In option pricing method, one seeks market-based valuations methods and policies that maximise market values. This does not mean that the personal probabilities of each agent do not matter. A consequence of the no-arbitrage condition is that the market requires all the agents to have unique price for each investment opportunity or security. Also, every asset must yield the same expected return r per pound invested, that is known as the riskless rate of return. The expected rate of return depends on the initial price which is given and since the condition of no-arbitrage is imposed the expected discounted value of the security must be a martingale process. Expectations are taken with respect to risk-neutral probabilities which should be understood as state prices, that is the price of the contract which pays one pound only when the specific state occurs. Each state has unique price as was suggested by the linear dual program.

It has been claimed that decision tree analysis using discounted cash flows is not appropriate technique to value managerial flexibility and as an alternative method the option pricing technique has been suggested. We can claim that this statement has probably been originated by misuse of decision analysis, since every agent in the market can adapt his or her own subjective probability and utility function as required by the market and then apply the decision analysis and derive the exact value for his or her investments opportunities as would be yielded in option pricing analysis. In decision analysis, the decision maker's risk preferences are confounded with time preferences and the future decision opportunities are misvalued. This is reason why it is required to use two different interest rates (20% and 35%) in order to get the correct option values using decision analysis.

The assumption of perfectly correlated twin securities may not be met in the case where there is no market with traded assets with the same uncertainty as the investment opportunity. In this case, it may be feasible to use option pricing method to create lower and upper bounds on the market valuation of the decision options. Since the market is not complete, there exist a set of probability measures, which are considered to be risk-neutral and one can use them to get the expected discounted cash flow. Thus the decision maker has upper and lower bounds for the project or investment opportunity value.

Chapter 6

Multi-Armed Bandit Problem

6.1 Introduction

In chapter 2, we studied various sequential decision processes. One of them was the discrete time version of the multi-armed bandit problem. We also mentioned that the Gittins index (see section 2.5.2) is the solution to the bandit problem. The optimality of the Gittins index is based on the principle of the forward induction policy, that is, projects are scheduled in the decreasing order of the expected reward per unit time which they yield. Recall the following version of the bandit problem:

Consider a situation in which several projects are candidates simultaneously for the attention of a single investigator. Suppose we have N independent projects, indexed $i = 1, 2, \ldots, N$. Project i can be in one of a finite number of states $x_i \in N_i$. We are allowed to choose only one project at each instant of discrete time $t = 0, 1, 2, \ldots$. If project i, which is in state $x_i(t)$ at time t is chosen, *i.e.*, $i = i(t)$, then an immediate reward of $R_{i(t)}\{x_{i(t)}(t)\}$ is earned. Rewards are additive and are discounted in time by a discount factor $0 < \alpha < 1$. The state $j = x_i(t)$ changes to $k = x_i(t+1)$ according to a homogeneous Markov transition rule, with transition matrix $P^i = (p^i_{jk})_{j,k \in N_i}$, *i.e.*, the distribution of k to occuring conditional on all previous state values of all project values is in fact dependent only on the current state j. It is also assumed that the current state of each project is known. The states of the projects that have not been chosen remain unchanged. The problem is to find a scheduling policy π that determines which project to choose at each time in order to maximize the total discounted reward earned over an infinite horizon as follows:

$$V_\pi(x) = E_\pi\left[\sum_{t=0}^{\infty} \alpha^t R_{i(t)}\{x_{i(t)}(t)\}\middle| x_1(0), x_2(0), \ldots, x_N(0)\right].$$

The above problem is known as the Multi-Armed Bandit problem and it was first solved by Gittins and Jones (1974) as follows:

Optimal strategies for such problems are known to be determined by a collection of dynamic allocation indices (DAI). This is an index associated to each project involved in the problem, and it depends on characteristics of that project (states, rewards and transition probabilities), and not of those of the other projects. A formula for this index was presented in section (2.5.2) to be:

$$\nu_i(x_i) = \max_{\tau > 1} \frac{\mathrm{E}\left[\sum_{t=0}^{\tau-1} \alpha^t R_i\{x_i(t)\} \;\middle|\; x_i(0) = x_i\right]}{\mathrm{E}\left[\sum_{t=0}^{\tau-1} \alpha^t \;\middle|\; x_i(0) = x_i\right]}, \tag{6.1}$$

where the maximization is over the set of all stopping times $\tau > 1$.

They established by the DAI theorem that the optimal action at each time is to work on a project with the largest current index. Such a policy is referred to as an *Index policy.*

Gittins (1979) defined a bandit process as a Markov decision process on state space, Θ, which is a subset of some space, together with $\sigma -$ algebra $\mathcal{X}$ of subsets of Θ which includes every subset consisting of just one element of Θ, and with set of controls $\Omega(x) = \{0, 1\}$, $\forall x$. Application of control $u \in \Omega(x)$ at time t with the process in state x yields a reward $R(x, u)$ which has discounted value $\alpha^t R(x, u)$ $(0 < \alpha < 1)$. When control 0 is applied the bandit process remains in the same state with probability one and this yields no reward, that is, $P(\{x\}|x, 0) = 1$ and $R(x, 0) = 0$, $\forall x$. The alternative action is control 1 which is the *continuation* control, where no restriction is placed on the transition probabilities and the rewards. Also, the number of times control 1 has been applied to a bandit process is termed the *process time.* The state of the process at time t is denoted by $x(t)$. If control 1 has been applied to all decision times up to now, then the process time coincides with real time, because the bandit process changed state at each decision time t and thus yields reward $\alpha^t R(x(t), 1)$. Also, Gittins introduced the concept of a standard bandit process which yield constant reward λ, every time the control 1 is applied, that is, $R(x, 1) = \lambda$, $\forall x$.

The DAI theorem refers to a simple family of alternative bandit process. A family of alternative bandit processes is a collection of n bandit processes, with the constraint that the control 1 is applied to only one bandit process and the control 0 to the other $n-1$ available bandit process. The reward gained at time t comes from the bandit process which the control 1 applied at time t.

Whittle (1980) provided a proof of the DAI result based on dynamic programming. He introduced the concept of the "retirement" reward (or option) in the following sense:

Consider the situation in which we have just project i and two alternative actions, these are, either to operate project i or to stop operation and receive a 'retirement' reward of M. If the operation of the project is terminated, its state will not change and it is assumed that the operation will not start again in the future. The reward at time t will be $R_i\{(x_i(t))\}$ if $i(t)$ is the project engaged at time t and for simplicity can be denoted by $R(t)$. In this case the total discounted reward is equal to $\sum_{t=0}^{\infty} \alpha^t R(t)$. He also supposed that rewards are uniformly bounded:

$$-\infty < A(1-\alpha) \leq R_i(x) \leq B(1-\alpha) < \infty$$

where A and B are lower and upper bounds on the total discounted reward respectively. Denote by $V(x_i, M)$ the value function of this optimal stopping problem which is dependent on the retirement reward M and on the project state x_i as explicitly,

$$V(x_i, M) = \max\left[M, R_i(x_i) + \alpha \mathrm{E}\left\{V(x_{i+1}, M)|\, x_i\right\}\right]. \tag{6.2}$$

The continuation option yields a total reward which consists of the sum of immediate reward $R_i(x_i)$ and present value of the expected future reward denoted by $\alpha \mathrm{E}\left[V(x_{i+1}, M)|\, x_i\right]$, that is the reward that will be gained if one proceeds optimally after the gain of $R_i(x_i)$ and the option of retirement is still available. The other option yields reward M.

The dynamic allocation index $M_i(x_i)$ of project i in state x_i is the infimum value of M which makes the decision maker indifferent between retirement option and the policy to continue operation for the project i because the M is just large enough that the options of continuing or terminating are equally attractive; or alternatively it is the supreme value over all terminal rewards, such that the decision maker still prefers to continue with the random stream of rewards *i.e.,*

$$M_i(x_i) = \inf\{M \in \mathbb{R} | V(x_i, M) = M\}, \tag{6.3}$$

$$= \sup\{M \in \mathbb{R} | V(x_i, M) > M\}. \tag{6.4}$$

However, what Gittins characterizes as a dynamic index is not the $M_i(x_i)$ but a multiple of it, that is,

$$\nu_i(x_i) = (1 - \alpha)M_i(x_i). \tag{6.5}$$

This difference arises because Gittins initially used the concept of a standard project with fixed reward rate $R = \nu$ whereas Whittle introduced the concept of a retirement reward M.

In this chapter we present the classical Multi-Armed Bandit problem. In section (6.2) we prove that it is optimal to add two Gittins Indices and the resulting number is an index for choosing two projects simultaneously only when the two projects have identical operating times. We also give an example to show that this rule is not optimal for projects with different operating times. Modified Bandit problems are presented in section (6.3). Option pricing and Bandit process are reconciled in section (6.4).

6.2 Multi -Armed Bandits with Two Servers

A deterministic two-armed bandit problem

Consider two bandit machines, X and Y. Their rewards are known sequences of non-negative numbers. If machine X is operated continuously, it yields rewards $X(1), X(2), \cdots$ respectively to the $1^{\text{st}}, 2^{\text{st}}, \cdots$ operation of X. Similarly the machine Y yields rewards $Y(1), Y(2), \cdots$ in its $1^{\text{st}}, 2^{\text{st}}, \cdots$ operation respectively. Suppose that the rewards are discounted with discount factor α where

$0 < \alpha < 1$. The problem is to carry out a series of trials on X and Y in such a way that only one bandit is selected each time and the total reward received is maximized.

The Gittins index for X is defined to be:

$$\lambda = \max_{\tau > 1} \frac{\sum_{t=1}^{\tau-1} \beta^t X(t)}{\sum_{t=1}^{\tau-1} \beta^t}, \tag{6.6}$$

where the maximization is over all the integers τ larger than 1. The interpretation is that it gives the maximum discounted reward per unit discounted time that can be obtained from the period which starts at $t = 1$ and ends at $t = \tau - 1$. The maximization is over the time t and the maximum value is realized for $t = \tau - 1$, meaning that a decision maker would be indifferent to choose between the bandit with sequence of rewards $\{X(t)\}_{t=0}^{t=\tau-1}$ and the one which yields constant reward λ at any time $t\,(0 \leq t < \tau)$.

Suppose that we are looking for a criterion for making a decision whether it is best to work on project X first, with the option of switching later to a project with constant reward λ, or to start with the project λ. Then one would start with X if and only if

$$\sup_{\tau} \left\{ \sum_{t=1}^{\tau-1} \beta^t X(t) + \lambda \sum_{t=\tau}^{\infty} \beta^t \right\} > \lambda \sum_{t=0}^{\infty} \beta^t \tag{6.7}$$

i.e., for some τ

$$\sum_{t=1}^{\tau-1} \beta^t X(t) + \lambda \sum_{t=\tau}^{\infty} \beta^t > \lambda \sum_{t=0}^{\infty} \beta^t \tag{6.8}$$

which yields

$$\lambda < \frac{\sum_{t=1}^{\tau-1} \beta^t X(t)}{\sum_{t=0}^{\tau-1} \beta^t} \tag{6.9}$$

and therefore

$$\lambda < \sup_{\tau} \frac{\sum_{t=1}^{\tau-1} \beta^t X(t)}{\sum_{t=0}^{\tau-1} \beta^t} \tag{6.10}$$

which means it is best to start work on project X if and only if (6.7) holds for some τ, *i.e.,* if and only if (6.10) is valid.

6.2.1 Identical operating times of both projects in a given pair

Consider two pairs of projects with deterministic rewards:

$$\text{Pair 1:} \qquad X^i(s),\ s = 1, 2, \ldots \qquad \text{for i=1,2.}$$

$$\text{Pair 2:} \qquad Y^i(t),\ t = 1, 2, \ldots \qquad \text{for i=1,2.}$$

By the term pair, we mean that if the X pair is operated for t^{th} time, an immediate reward $X^1(t) + X^2(t)$ is earned. Alternatively, one can choose to operate Y pair for s^{th} time and gain an immediate reward $Y^1(s) + Y^2(s)$. We denote by ν_{X_i} the Gittins index for project with reward sequence $\{X^i(t)\}_{t=1}^{t=\infty}$. The sum of Gittins indices for the pair of projects X^i for $i = 1, 2$ is denoted by $\nu_X = \sum_{i=1}^{2} \nu_{X_i}$. The Gittins index for Y^i is denoted by ν_{Y_i} and their sum is $\nu_Y = \sum_{i=1}^{2} \nu_{Y_i}$.

In this section we shall show that, in the case where the Gittins indices for projects which belong to the same pair are always maximized for the same t (t may differ for different groups), and if we add the Gittins indices for one pair of projects, let's say ν_X, and compare their sum with the sum of Gittins indices of the other pair of projects, that is ν_Y and then select the pair of projects having greatest sum equal to $\max[\nu_X, \nu_X]$, then we end up with an index policy *i.e.,* it is optimal to choose the pair of X projects instead of Y pair, if and only if,

$$\nu_X > \nu_Y. \tag{6.11}$$

In order to prove (6.11), we define two policies for choosing projects and then we use an interchange argument to find the policy which yields the highest reward.

Proof: Suppose that $\nu_X > \nu_Y$ and that ν_X is realised at τ. Define two policies π and μ as follows:

Policy π: first operate the Y pair machines for k_1 times, then play X pair machines for one time, then play Y pair for $(k_2 - k_1)$ times, etc where k_t denotes the number of times Y pair projects used preceding the t^{th} use of X pair projects.

Policy π has rewards

$$\sum_{i=1}^{2} Y^i(1), \sum_{i=1}^{2} Y^i(2), \cdots, \sum_{i=1}^{2} Y^i(k_1), \sum_{i=1}^{2} X^i(1), \sum_{i=1}^{2} Y^i(k_1+1), \cdots, \sum_{i=1}^{2} Y^i(k_2),$$
$$\sum_{i=1}^{2} X^i(2), \cdots, \sum_{i=1}^{2} Y^i(k_\tau - 1), \sum_{i=1}^{2} X(\tau-1), \sum_{i=1}^{2} X(\tau), \sum_{i=1}^{2} X(t+1), \cdots.$$

Policy μ : first operates the X pair projects for $(\tau-1)$ times, Y pair projects for $k_{\tau-1}$ times and then follows policy π.

Policy μ has rewards

$$\sum_{i=1}^{2} X^i(1), \sum_{i=1}^{2} X^i(2), \cdots, \sum_{i=1}^{2} X^i(\tau-1), \sum_{i=1}^{2} Y^i(1), \sum_{i=1}^{2} Y^i(2), \cdots,$$
$$\sum_{i=1}^{2} Y^i(k_{\tau-1}), \cdots \text{ as policy } \pi.$$

The present values of these policies are

$$\begin{aligned} V(\pi) = \sum_{i=1}^{2} \sum_{t=1}^{k_1} \alpha^t Y^i(t) + \ldots + \alpha^{\tau-2} \sum_{i=1}^{2} \sum_{t=k_{\tau-2}+1}^{k_{\tau-1}} \alpha^t Y^i(t) \\ + \sum_{i=1}^{2} \sum_{t=1}^{\tau-1} \alpha^{k_t+t} Y^i(t) + \sum_{i=1}^{2} \sum_{t=T+1}^{\infty} \alpha^t Z^i(t). \end{aligned}$$
$$V(\mu) = \sum_{i=1}^{2} \sum_{t=1}^{\tau-1} \alpha^t X^i(t) + \alpha^{\tau-1} \sum_{i=1}^{2} \sum_{t=1}^{k_{\tau-1}} \alpha^t Y^i(t) + \sum_{i=1}^{2} \sum_{t=T+1}^{\infty} \alpha^t Z^i(t).$$

Denote $V(\mu) - V(\pi) = \Delta_X - \Delta_Y$. Now, we need to calculate Δ_X. Before we attempt to calculate Δ_X, we note the following inequality:

$$\sum_{n=t}^{\tau-1} \alpha^n \sum_{i=1}^{2} X^i(n) \geq \left(\sum_{i=1}^{2} \nu_X^i \right) \sum_{n=t}^{\tau-1} \alpha^n. \tag{6.12}$$

To verify (6.12) by definition of the Gittins index we assume that the Gittins indices ν_{X_1} and ν_{X_2} for projects X^1 and X^2 are always maximised for same t and let the best t be τ.

For the best τ, and $i = 1, 2$,

$$\begin{aligned} \nu_X^i \sum_{t=1}^{\tau-1} \alpha^t &= \sum_{t=1}^{\tau-1} \alpha^t X^i(t) \\ &= \sum_{t=1}^{\sigma-1} \alpha^t X^i(t) + \sum_{t=\sigma}^{\tau-1} \alpha^t X^i(t) \quad \text{where } (\sigma \leq \tau - 1). \end{aligned} \tag{6.13}$$

Also,

$$\nu_X^i \sum_{t=1}^{\sigma-1} \alpha^t \geq \sum_{t=1}^{\sigma-1} \alpha^t X^i(t). \tag{6.14}$$

Subtracting (6.13) $-$ (6.14), we get

$$\nu_X^i \sum_{t=\sigma}^{\tau-1} \alpha^t \leq \sum_{t=\sigma}^{\tau-1} \alpha^t X^i(t). \tag{6.15}$$

Writing (6.15) for $i = 1, 2$ and adding them we get (6.12).

Now, we need to calculate Δ_X as follows:

$$\begin{aligned}
\Delta_X &= \sum_{i=1}^{2}\sum_{t=1}^{\tau-1} \alpha^t X^i(t) - \sum_{i=1}^{2}\sum_{t=1}^{\tau-1} \alpha^{k_t+t} X^i(t) \\
&= \sum_{i=1}^{2}\sum_{t=1}^{\tau-1} (1-\alpha^{k_t})\alpha^t X^i(t) \\
&= \sum_{t=1}^{\tau-1} (1-\alpha^{k_t})\alpha^t \sum_{i=1}^{2} X^i(t) \\
&= \sum_{t=1}^{\tau-1} (\alpha^{K_{t-1}} - \alpha^{K_t}) \sum_{n=t}^{\tau-1} \alpha^n \sum_{i=1}^{2} X^i(t) \\
&\geq \sum_{t=1}^{\tau-1} (\alpha^{K_{t-1}} - \alpha^{K_t}) \sum_{n=t}^{\tau-1} \alpha^n \sum_{i=1}^{2} \nu_X^i \qquad \text{by (6.12)} \\
&= \left(\sum_{i=1}^{2} \nu_X^i\right) \sum_{i=1}^{\tau-1} (1-\alpha^{K_t})\alpha^t.
\end{aligned}$$

Similarly, we calculate Δ_Y as follows:

$$\begin{aligned}
\Delta_Y &= -\alpha^{\tau-1}\sum_{i=1}^{2}\sum_{t=1}^{k_{\tau-1}} \alpha^t Y^i(t) + \sum_{i=1}^{2}\sum_{t=1}^{\tau-1} \alpha^{k_1} Y^i(t) + \ldots + \alpha^{\tau-2}\sum_{i=1}^{2}\sum_{t=k_{\tau-2}+1}^{k_{\tau-1}} \alpha^t Y^i(t) \\
&= (1-\alpha^{\tau-1})\sum_{t=1}^{k_1} \alpha^t \sum_{i=1}^{2} Y^i(t) + (\alpha - \alpha^{\tau-1}) \sum_{t=k_1+1}^{k_2} \alpha^t \sum_{i=1}^{2} Y^i(t) + \ldots \\
&\quad \ldots + (\alpha^{\tau-2} - \alpha^{\tau-1}) \sum_{t=k_1+1}^{k_2} \alpha^t \sum_{i=1}^{2} Y^i(t) \\
&= \sum_{t=1}^{\tau-1} (\alpha^{t-1} - \alpha^t) \sum_{n=1}^{k_t} \alpha^n \left(\sum_{i=1}^{2} Y^i(n)\right).
\end{aligned} \tag{6.16}$$

From (6.16), and using (6.12) and that $\sum_1^\sigma \alpha^t Y_t^i \leq \nu_Y^i \sum_1^\sigma \alpha^t$, we have:

$$\begin{aligned}
\Delta_Y &\leq \left[\sum_{t=1}^{\tau-1}(\alpha^{t-1}-\alpha^t)\sum_{n=1}^{k_t}\alpha^n\right]\sum_{i=1}^{2}\nu_X^i & (6.17)\\
&= \left(\sum_{i=1}^{2}\nu_X^i\right)\sum_{t=1}^{\tau-1}(1-\alpha^{k_t})\alpha^t. & (6.18)
\end{aligned}$$

So,

$$\Delta_Y \leq \left(\sum_{i=1}^{2}\nu_X^i\right)\sum_{i=1}^{\tau-1}(1-\alpha^{K_t})\alpha^t \leq \Delta_X,$$

and,

$$\nu(\mu) - \nu(\pi) = \Delta_X - \Delta_Y$$

$$\Rightarrow \nu(\mu) \geq \nu(\pi).$$

Now, we have proved that it is better to use the rule $\sum_{i=1}^2 \nu_X^i \geq \sum_{i=1}^2 \nu_Y^i$ until the time $\tau - 1$. This argument can be repeated starting at time τ. This prove the optimality of the following rule:

Optimal rule is based on

$$\sum_{i=1}^{2}\nu_X^i \geq \sum_{i=1}^{2}\nu_Y^i. \tag{6.19}$$

To derive this result we assumed that both X pair project have indices which are maximised for the same best τ. If we relax this assumption there is no optimal rule as shown in the following example:

Example

Consider the following two pairs of projects; Let

$$\begin{aligned}
X_1(s) &= \{3,0,3,0,3,\ldots\}\\
X_2(s) &= \{0,3,0,3,0\ldots\}
\end{aligned}$$

The Y pair is:

$$\begin{aligned}
Y_1(s) &= \{1,2,1,2,1,\ldots\}\\
Y_2(s) &= \{2,1,2,1,2\ldots\}
\end{aligned}$$

The Gittins index is equal to

$$\gamma_n^j = \max_{\tau > n} \frac{\sum_{t=n}^{\tau-1} \beta^t X^j(t)}{\sum_{t=n}^{\tau-1} \beta^t}$$

Set the discounting factor $\beta = 0.5$. One can calculate the sequence of Gittins indices of each project as follows: Consider the sequence of rewards $\{3, 0, 3, 0, 3, \dots\}$ then

$$\gamma_0^1 = \max\left\{3, \frac{3 + 0(0.5)}{1 + 0.5}, \frac{3 + 0(0.5)}{1 + 0.5 + 0.5^2}, \frac{3 + 0(0.5) + 3(0.5)^2}{1 + 0.5 + 0.5^2 + 0.5^3}, \dots\right\} = 3$$

Similarly, one can show that

$$\gamma_1^1 = \max\left\{0, \frac{0(0.5) + 3(0.5)^2}{0.5 + 0.5^2}, \dots\right\} = 1$$

One can easily show that the index sequence for project $\{X_j(s)\}$ and $\{Y_j(s)\}$ for $j = 1, 2$ are

Sequence of Gittins index values

Project	γ_0	γ_1	γ_2	γ_3	γ_4	γ_n
X_1	3	1	3	1	3	...
X_2	1	3	1	3	1	...
Y_1	$\frac{4}{3}$	2	$\frac{4}{3}$	2	$\frac{4}{3}$	...
Y_2	2	$\frac{4}{3}$	2	$\frac{4}{3}$	2	...

Table 6.1: Gittins Index values

For any given time t, if you add the Gittins indices for pair X, we get the value 4. For the pair of Y projects, the sum of indices is equal to $10/3$. Although, the sum of indices for the two pairs of projects are not equal, the rewards of the two pairs are the same. This is a contradiction, and therefore the index rule is not valid.

When the problem is formulated in a finite horizon time T, for example $T = 2$, the Gittins index may yield suboptimal results as follows:

Let the problem be to maximise the expected discounted reward received up to time T, with respect to a policy π, that is

$$\begin{array}{ll} \underset{\pi}{\text{Maximize}} & E_\pi \left[\sum_{t=0}^{2} \alpha^t \{\pi_t^X X(t) + \pi_t^Y Y(t) + \pi_t^Z Z(t)\} \middle| X(0), Y(0), Z(0) \right] \\ \textit{subject to} & \pi_t^X + \pi_t^Y + \pi_t^Z \leq 1, \end{array}$$

where the set of all policies is determined by the values of $\pi = (\pi_t^X, \pi_t^Y, \pi_t^Z)$ and $\pi_t^X, \pi_t^Y, \pi_t^Z \in \{0,1\}\, \forall t$, meaning that, for example, $\pi_t^X = 1$ if the project X is chosen at time t and $\pi_t^X = 0$ otherwise. Let the projects X, Y give identical rewards $(2,0)$ and project Z has reward $(2,1)$. The Gittins index for the projects are equal to 2. The optimal policy is to choose project Z and one of X or Y project which yields total discounted profit $4+\beta$, whereas if projects X and Y are chosen the profit is 4.

6.3 Modified Bandit Problems

In this section, we point out that the *Gittins Index policy* is not an optimal solution for the following modified Bandit problems:

1. When an strategic option such as switching cost or switching delay is incorporated in the allocation process,
2. When more than one machine operates simultaneously,
3. When the time horizon is finite.

The nature of optimal strategies for the general multi-armed bandit problem when some strategic options are incorporated is not generally easy to discover. The Gittins index will not give the optimum allocation strategy and therefore a different index must be derived for each case. It is difficult to know in advance what the optimum allocation strategy is, but it can be derived in the same way as in the classical multi-armed bandit problem

All three above cases are interesting problems because of their importance in practice. It is not unrealistic for the project manager to have to pay a switching cost or penalty for each switch made from one project to another or to delay the development of a specific project. However this changes the problem drastically.

This is so because the objective is to find the optimal allocation strategy which maximises the sum of discounted net reward where its value reduces every time a switching cost is paid. For example, suppose that, there are two projects with deterministic sequences of rewards and whenever the manager switches to the other project, a cost of amount C is incurred. During the switching period, no project is operated and therefore no reward is gained and this cost can be understood as a switching cost. Denote by $m(t)$ the project chosen at time t. A switching cost is paid every time t for which $m(t) \neq m(t-1)$. Let I denote the indicator function. The task is to determine the order of operation of the projects, one at each time, such that

$$\underset{m(t)}{\text{Maximize}} \sum_{t=0}^{\infty} \alpha^t \left[R(t) - I\{m(t) \neq m(t-1)\}C\right] \tag{6.20}$$

The solution to the optimization problem (6.20) is given by the following index (see, Asawa & Teneketzis, 1996)

$$\max_{\tau>1} \frac{\sum_{t=0}^{\tau-1} \alpha^t R^j(t) - C\alpha^t}{\sum_{t=0}^{\tau-1} \alpha^t}$$

and represents the maximum discounted reward rate given project j which has been operated from time period $t = 0$ to $\tau - 1$ and switching cost C is incurred at the next time interval at instant t.

More generally, one can associate with each project a switching cost C_j dependent of the current project. The index becomes

$$\max_{\tau>1} \frac{\sum_{t=0}^{\tau-1} \alpha^t R^j(t) - C_j\alpha^t}{\sum_{t=0}^{\tau-1} \alpha^t}.$$

6.3.1 Bandits with switching delay

Assume the set up scenario of Multi-Armed Bandits with switching penalty. Suppose that instead of the project manager paying a penalty, a switching delay D

occurs when the operator changes projects.An appropriate index is:

$$\max_{\tau>t} \frac{\alpha^D \sum_{l=t}^{\tau-1} \alpha^t R^j(t)}{\sum_{l=t}^{\tau+D-1} \alpha^t}$$

where the maximization is over all stopping times $t < \tau < \infty$.

6.3.2 Allocation in finite horizon case

The issue of a finite horizon time is of great importance since allocation of effort or resources will not be done forever because of its limited nature. The question is how the Gittins index can give us a solution to the finite horizon allocation problem.

A Bandit problem with finite horizon T seems to be an interesting case since in practice allocation problems may have a deadline. The problem is to maximise the expected discounted reward received up to time T, with respect to a policy π, that is

$$V_\pi^T(x) = \sup_\pi E_\pi \left[\sum_{t=0}^{T} \alpha^t R_i\{x_i(t)\} \middle| x(0) = x \right].$$

It is known that the Gittins index policy does not maximize the expected discounted reward obtained by finite time T. For example, set $T = 2$ and consider project X with rewards $(0, 4, 0)$ and project Y with rewards $(0, 0, 9)$. Set the discount factor $\alpha = 0.5$. The sequence of the Gittins index for the these projects are: The Index rule says to choose projects in the sequence X, X, Y since $\frac{4}{3} > \frac{9}{7}$

Gittins index			
Project	γ_0	γ_1	γ_2
X	$\frac{4}{3}$	4	0
Y	$\frac{9}{7}$	3	9

Table 6.2: Sequential choice of projects based on the Gittins index

then we compare $4 > \frac{9}{7}$ and choose project X and the third choice is the project

Y because $\frac{9}{7} > 0$. The index rule choice yields a return of 2. It is easy to see that the optimal choice is to select the project Y which yields discounted profit $0 + 0 + (\frac{1}{2})^2 7 = 2.25$. Therefore the index rule gives us a suboptimal solution.

In practice, sometimes it is not so obvious what is the value of T. However, we know that the allocation will be terminated in finite time as in the case of a clinical trial. Consider N patients who are in a clinical trial. Assume also two treatments, A and B, each treatment is allocated to a patient, one at each time and the result of each treatment may classified as a success or failure. One can denote the response of patient i by R_i which is a Bernoulli variable with unknown parameter θ. The random variable R takes value 1 for successful treatment and 0 otherwise. We also need to discount reward received, this indicates that it is important how early a success is observed and that without discounting the order would not matter. The objective will be to maximize the expected number of discounted successes in the trial, that is $E(\alpha^t \sum R_i)$.

When one decides to terminate the allocation of treatment a reasonably large value of T is needed to get a realistic value of the index since the value of T affects its value. However, one can specify a target level which may represent a significant level of potential therapeutic activity in screening trial.

In this case we can think in terms of the *bandit sampling process*, that is, a sequence of independent random variables $X_1, X_2, \ldots$, which have an unknown probability distribution from a parametric distribution $\mathcal{D}$ (Gittins, 1989, Chapters 6 and 7). A sampling process is a reward process if the observations themselves constitute a sequence of rewards.

The Gittins index $\nu(\Pi)$ for a bandit process in state Π is defined to be the maximum expected reward per unit of discounted time up to stopping time. Denote by $R_\tau(\Pi)$ the total expected discounted reward up to time τ, where τ is a stopping time defined in terms of the filtration $\{\mathcal{F}_t : t \geq 1\}$, where $\mathcal{F}_t$ denotes the $\sigma-$fields $\sigma\{X_i : i \leq t\}$ and the total expected discounted time up to time τ is denoted by $W_\tau(\Pi)$. Then, we have

$$\nu_\tau(\Pi) = \frac{R_\tau(\Pi)}{W_\tau(\Pi)} = \frac{E\sum_{t=0}^{\tau-1} \alpha^t r(X_t)}{E\sum_{t=0}^{\tau-1} \alpha^t} \qquad \text{and} \qquad \nu(\Pi) = \sup_\tau \nu_\tau(\Pi).$$

One can look at the supremum which is over the class of stopping times for which

$P(\tau \leq T) = 1$, that is,

$$\nu^T(\Pi) = \sup_{\tau \leq T} \nu_\tau(\Pi)$$

Gittins and Wang (1992) defined the difference of $\nu(\Pi) - \nu^1(\Pi)$ to be learning component of index, that is the difference between the index and the expected immediate reward. So, when one has two bandits with same expected immediate reward, the learning component should be larger for the bandit with the more uncertain reward rate.

Now consider a bandit sampling process in which a unit of reward occurs only when the observations is greater than some target value L for first time, and 0 for all other X_i's. Also, there is an additional state C, which is the completion state, and the process reaches this state when the target level L is achieved and then no further reward is possible. An important characteristic of an index policy for a set of target processes is that it minimizes the expected flow time to complete all of them and also minimizes the expected time to minimize just one of them (Gittins, 1989, page 131).

For a Bernoulli reward process, one can set $P(X_i = 1) = 1 - P(X_i = 0) = \theta$. We assume that θ has prior density of the conjugate form Beta(α, β) with probability density function of the form

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad 0 \leq \theta \leq 1,$$

where $\alpha > 0$ and $\beta > 0$. The state of the bandit process can be represented by parameters α and β and is denoted by (α, n) where $n = \alpha + \beta$ is number of patients have been allocated to a treatment and α is the number of success out of n. The Gittins index is defined as $\nu(\alpha, n) = \sup_{\tau>0} \nu_\tau(\alpha, n)$. The state of bandit process changes every time a new observation is taken and n becomes $n+1$ and α increases by 1 when a "success" has been observed and by 0 otherwise. The expected value of the next observation is given by

$$r(\alpha, \beta) = \int_0^1 \theta \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \frac{\alpha}{\alpha+\beta} = \frac{\alpha}{n}$$

which is the expected value of θ. This is equal to the expected immediate reward.

A related sequential decision optimization problem is the decision a financial analyst faces when his financial option must be exercised by specific time t. He has to decide when it is optimum to do so. This must only happen if the current price of share exceeds a threshold level (as in the target process) which is a function of the number of days to go before its expiry of the option. In the next section, we study option pricing and bandit processes together.

6.4 Option Pricing and Bandit process

The purpose of this section is to integrate the Multi-armed bandit problem with the financial option pricing method. The reconciliation of these two problems can be made through consideration of the two questions addressed by each problem and their link with the methodology used to get a solution to them. We present the theorems which justify the optimality of Gittins Index, and the pricing rules in Option Pricing method. The reasons why they hold are discussed and they are used to provide the common ground of both theories. Then a comparison of the solutions can be made.

In financial markets, an option holder observes how Option values vary from time to time in relationship to the price of the underlying stock. A fundamental problem in investment theory is how to determine the value of an option. Also the holder of the option is concerned to find out what it is the best time for him to exercise his option. At every decision time the choice of actions is to exercise the option *i.e., stop*, to get involved in the trading procedure, or not to exercise the option *i.e., continue* to observe the market. At the expiry time t the holder receives a payoff $f_t = f_t(S_0, S_1, \cdots, S_t)$ which is determined by the "history" of the underlying asset price S_t up to time t. The problem in making his decision is he does not know what state the underlying asset price process is in at the future times and therefore has no information of the future profit he may gain if he postpones the exercise action now. However, he knows the probability distribution of the "future". After the expiry date the option has no value.

In the bandit problem, the decision maker's objective is to choose a project selection sequence to maximize some function of the expected total reward over the planning horizon. He has the policy to engage the project which gives the highest current reward rate and he needs to know for how long the already engaged project will continue to give the highest reward rate. Alternatively, one can suppose that, every time the decision maker uses the project j, he pays a fixed charge γ_j which is dependent on the project and its state only. When this charge is set up to such a value that it is neither a profitable nor a loss-making action to use the project j for a specified period, we say that this is the fair value. If the charge is greater than its fair value the project will yield some loss for the use of the specified period (see also, Weber, 1992).

One can start by comparing the two frameworks the models are built on. We only consider options with payoff functions which are path type dependent that is, its payoff $f_t = f_t(S_0, S_1, \cdots, S_t)$ is determined by what have been observed so far. If the option is not of this type, it can not be related to the Multi-armed bandit problem for reasons which will be obvious later.

In the next section we discuss the classical optimal stopping problem and introduce the concept of the Snell envelope. The Multi-armed bandit problem is presented as a control problem as was reformulated for its first time by Mandelbaum (1986). The problem of pricing of an American option is analyzed and connected with Multi-armed bandit problem.

6.4.1 Optimal Stopping

Suppose that we can observe a sequence of random variables $x_1, x_2, \dots$, which are defined on a Probability space $(\Omega, \mathcal{F}, \mathcal{P})$. We assume an increasing sequence $(\mathcal{F}_n)$ of sub-σ-algebras of $\mathcal{F}$ and that the n^{th} element of the sequence x_n is measurable with respect to $\mathcal{F}_n$, $n = 1, 2, \dots,$. A pair of sequences $\{x_n, \mathcal{F}_n\}_1^\infty$ is called a stochastic sequence.

A stopping time is a random variable τ from Ω into $\{1, 2, \dots\}$, such that $P(\tau < \infty) = 1$ and for all n, the event $\{\tau = n\}$ is in $\mathcal{F}_n$.

If we stop the observation process at some point n, we receive a reward y_n, which is dependent only on the already observed values $x_1, \dots, x_n$. To define the actual reward received, think about the stochastic sequence $\{y_n, \mathcal{F}_n\}_1^\infty$ and t any stopping variable then the random variable y_t defined by

$$y_t = \sum_{n=1}^{\infty} y_n I_{\{t=n\}} = \begin{cases} y_n & \text{on } \{t = n\} \text{ for } n = 1, 2, \ldots \\ 0 & \text{on } \{t = \infty\} \end{cases}$$

is the reward, where $I_{\{t=n\}}$ is the indicator of the set $\{t = n\}$. The value V of the stochastic sequence $\{y_n, \mathcal{F}_n\}_1^\infty$ is defined to be $\sup_\tau Ey_t$, where the supremum is over all the stopping times t such that Ey_t exists. We assume that $E|y_n| < \infty$ $(n \geq 1)$.

Let D be any class of stopping times t such that Ey_t exists, we define $V = \sup_{t\in D} Ey_t$. The problem of the optimal stopping is to find an optimal stopping $t \in D$, such that

$$V = \sup_{t\in D} Ey_t.$$

V should be understood as the optimal average gain.

Optimal Stopping with a Finite Horizon

In the finite case we observe an integrable stochastic sequence $\{y_n, \mathcal{F}_n\}_1^\infty$ and we are looking for a stopping time $t \leq N$ where N is a fixed integer number. Denote by C_n^N the class of stopping times $n \leq t \leq N$, then we define

$$V_n^N = \sup_{t\in C_n^N} Ey_t,$$

that is, the supreme average gain Ey_t when one is allowed to stop observations at time t, such that $n \leq t \leq N$. One can solve this problem with the dynamic programming approach based on the principle of backward induction as follows: Suppose that $n = N$ then the only stopping rule in C_N^N is $t = n$ and we have $V_N^N = Ey_N$. If we go one step back in time, then $n = N - 1$ we need to compare y_{N-1} with $E(y_N|\mathcal{F}_{N-1})$ and determine the rule

$$t = \begin{cases} N-1 & \text{if} \quad y_{N-1} \geq E(y_N|\mathcal{F}_{N-1}), \\ N & \text{if} \quad y_{N-1} \leq E(y_N|\mathcal{F}_{N-1}). \end{cases}$$

The solution to this optimal stopping problem is given by the following theorem:

Theorem 6.1 (Chow, Robbins, and Siegmund (1971), p. 50)

Let N be a fixed positive integer. Define successively $\gamma_N^N, \gamma_{N-1}^N, \dots \gamma_1^N$ by setting

$$\gamma_N^N = y_N, \tag{6.21}$$

$$\gamma_n^N = \max[y_n, E(\gamma_{n+1}^N|\mathcal{F}_n)], \qquad n = N-1, \dots, 1. \tag{6.22}$$

For each $n = 1, 2, \dots N$, let

$$s_n^N = \text{first } i \geq n \text{ such that } y_i = \gamma_i^N.$$

Then $s_n^N \in C_n^N$ and

$$E(y_{s_n^N}|\mathcal{F}_n) = \gamma_n^N \geq E(y_t|\mathcal{F}_n), \; t \in C_n^N;$$

so that

$$Ey_{s_n^N} = E\gamma_n^N \geq Ey_t, \; t \in C_n^N, \text{ and } \nu_N^N = E\gamma_n^N.$$

This theorem implies that

$$\gamma_n^N = \operatorname*{ess\,sup}_{t \in C_n^N} E(y_t|\mathcal{F}_n), \qquad n = 1, 2, \dots, N.$$

The sequence $\gamma = (\gamma_n^N)_{0 \leq n \leq N}$ is called the Snell envelope of y_t which is defined as follows.

Definition 6.1 Let $y = (y_n)_{0 \leq n \leq N}$ be an adapted sequence of real valued integrable variable. The Snell envelope of y is the sequence $U = (U_n)_{0 \leq n \leq N}$ defined by

$$U_n = \operatorname*{ess\,sup}_{T_{n,N}} E(y_t|\mathcal{F}_n), \qquad 0 \leq n \leq N$$

where $T_{n,N}$ is the set of all stopping times with values in $\{n, n+1, \dots, N\}$.

Optimal Stopping with an Infinite Horizon

In the optimal stopping problem with an infinite horizon, an infinite sequence of random variable $y_1, y_2, \dots$ is observed which is an integrable stochastic sequence and the set of possible stopping times is extended to the set $t \geq 1$. It can be shown that the same result of theorem (6.1) can be extended in the infinite time horizon.

Theorem 6.2 (Chow et al. (1971), p. 66)
Define γ_n by

$$\gamma_n = \operatorname*{ess\,sup}_{C_n} E(y_t|\mathcal{F}_n), \quad (n = 1, 2, \dots)$$

where $\{y_n, \mathcal{F}_n\}_1^\infty$ is an integrable stochastic sequence and C_n is the set of stopping time t such that $t \geq n$. Then

$$\gamma_n = \max[y_n, E(\gamma_{n+1}|\mathcal{F}_n)], \qquad n = 1, 2, \dots, \tag{6.23}$$

$$\nu_n = E\gamma_n. \tag{6.24}$$

where $\nu_n = \sup_{C_n} Ey_t$.

From the above theorem one can conclude that the Snell envelope γ_n is the minimal supermartingale which dominates y_n. This conclusion can be derived if one notices that the inequalities $\gamma_n \geq E(\gamma_{n+1}|\mathcal{F}_n)$ and $\gamma_n \geq y_n$ imply (6.23). Also, if V is a supermartingale majorant of y then $V_n \geq y_n$. This property of the Snell envelope enable us to price options when their expiry time is not fixed. We consider the case of the American option.

6.4.2 Valuation of an American option

Recall that an option is a contract between two parties and the holder has the right to sell(buy) the underlying asset (e.g share or currency) under the contracting conditions. The options providing the right to buy are known as call options. If the option gives the right to sell, it is called a put option.

Options are also classified in terms of the expiration time. A European option is the one which has fixed expiration date. In contrast to the European option, the American type option can be exercised at any time up to expiration date.

In this section, valuation of an American option is assumed to be done in a viable complete market as was introduced in section (4.6). We consider a market at times $n = 0, 1, \dots, N$ with a stock process $S = (S_n)$ and bank account process $B = (B_n)$. We denote such a market by (B, S).

Given a (B, S) financial market, we examine an American type option with expiration time $N < \infty$ and a collection of non-negative payment functions $f = (f_n)_{0 \leq n \leq N}$. We assume that $f_n = f_n(S_0, S_1, \dots, S_n)$ are given and represents non-negative payments if the option is exercised at any time $n (n \leq N)$. f_n can be understood as the payment to the option buyer if the option is exercised at time n given the realization of stock prices $(S_0, S_1, \dots, S_n)$.

The holder of the American option has to make a decision when to exercise his option depending on the history of (B,S)-market, that is if τ denotes the exercise time, his decision to exercise its option at time $\tau = n$ or not $(\tau > n)$ is determined only by the available information up to time n. Therefore it is natural to assume that τ is a stopping time.

Consider the problem of pricing an American option. An investor faces the problem of replicating the value of an American option by constructing an admissible strategy, which gives the "fair" value of the American option.

Passing to the problem of valuing an American option, we have assumed that we operate in a viable complete market (see def. 4.6). This implies that, the absence of arbitrage possibilities is equivalent to the existence of a probability measure $\mathbf{P}^*$ which is absolutely continuous with respect to the basic probability measure $\mathbf{P}$ with strictly positive and bounded density and such that all the security prices are martingales with respect to $\mathbf{P}^*$.

Assume that there exists a self-financing strategy $\pi = (\pi_n)_{0 \leq n \leq N}$ with initial capital $X_0^\pi = x$ which will provide , at time n the capital $X_n^\pi(\omega)$ for $\omega \in \Omega$.

Definition 6.2 ((Shiryaev et al., 1994), p. 21) A fair or rational price of an American option with last expiration time N and a system of payment functions $f = (f_n)_{0 \leq n \leq N}$ is defined to be the minimal initial amount $X_0^\pi = x = \mathbb{C}_N^*$ required for any π self financing investment strategy having the property for all $0 \leq n \leq N$,

$$X_n^\pi(\omega) \geq f_n\left(S_0, S_1(\omega), \dots, S_n(\omega)\right), \qquad \omega \in \Omega.$$

In a pricing procedure one wonders what is the best strategy of the holder of the option with regard to his right to exercise it or not. Again, this is a situation of two possible actions as in Bandit problem. More importantly, this choice is reflected through the mathematics of the optimal stopping problem. It can be proved that the investor is perfectly hedged and the best strategy of the holder is to price the option as follow:

Let $Y = (Y_n)_{0 \leq n \leq N}$ be a positive sequence adapted to $(\mathcal{F}_n)$ which represents the value of the option at any arbitrary time n, where $N \in \mathbb{N}$, the maturity time. If we assume that $\tau = \tau(\omega)$ is an expiration time for a given American option, the contract requires the seller to be ready to make a payment $f_\tau = f_{\tau(\omega)}(S_0, S_1(\omega), \dots, S_{\tau(\omega)})$.

Definition 6.3 ((Shiryaev et al., 1994), p. 21) A stopping time $\tau^* = \tau^*(\omega)$ is said to be rational or reasonable expiration time (withdrawal, repayment) of American option, if for initial capital $\mathbb{C}_N^*$, for any self-financing strategy π having the property

$$X^{\pi}_{\tau^*(\omega)}(\omega) \geq f_{\tau^*(\omega)}(S_0, S_1(\omega), \dots, S_{\tau(\omega)}), \qquad \omega \in \Omega,$$

the equality

$$X^{\pi}_{\tau^*(\omega)}(\omega) = f_{\tau^*(\omega)}(S_0, S_1(\omega), \dots, S_{\tau(\omega)}), \qquad \omega \in \Omega,$$

is valid.

According to the contract

Before he exercises his option at time $\tau = n$ he compares the following two amounts:

1. The discounted value of profit, $\dfrac{f_n}{B_n}$, if the option is exercised at time $\tau = n$,
2. The expected discounted value of the contract at the next point in time $n+1$, $\mathbf{E}^*(Y_{n+1}|\mathcal{F}_n)$, if the option is not exercised at time at n and the option to proceed optimally remains open at all future times $\tau = n+1, \dots, N$.

Since an American option can be exercised at any time between 0 and N, we define the value process $Y_n = (Y_n)_{0 \leq n \leq N}$ of an American option described by a sequence (f_n) such that,

$$Y_N = \frac{f_N}{B_N}$$

and, for $n = 0 \leq n \leq N-1$,

$$Y_n = \max\left\{\frac{f_n}{B_n}, \mathbf{E}^*(Y_{n+1}|\mathcal{F}_n)\right\}.$$

Thus, the sequence $Y = (Y_n, \mathcal{F}_n, \mathbf{P}^*)$ is a supermartingale dominating the sequence (f_n/B_n).

Theorem 6.3 ((Shiryaev et al., 1994), p. 41)

1. Under the conditions of a (B, S)-market, the rational price C_N^* of American option with last expiration time N and a system of non-negative payment functions $f = (f_n)_{0 \leq n \leq N}$ is given by

$$C_N^* = \sup \mathbf{E}^* \alpha^\tau f_\tau,$$

where $\alpha = (1+r)^{-1}$ and sup is taken over all Markov times $\tau = \tau(\omega)$ such that $0 \leq \tau(\omega) \leq N$, $\omega \in \Omega$, and is achieved for some τ^*.

2. At time τ^* is rational if and only if

$$\mathbf{E}^* \alpha^{\tau^*} f_{\tau^*} = \sup_\tau \mathbf{E}^* \alpha^\tau f_\tau.$$

The above theorem establishes the basic principles in the theory of American options that the pricing problem is essentially an optimal stopping problem in the sense that if we manage to solve the optimal stopping problem "$\sup_\tau \mathbf{E}^* \alpha^\tau f_\tau$", the result is the option price (rational cost) $\mathbb{C}_N^*$ and at the same time the rational stopping time τ^* is known. Therefore one can characterize the arbitrage-free price process $Y = (Y_n)$ as:

$$Y(n) = \operatorname*{ess\ sup}_{\tau \in \mathcal{T}_{n,N}} \mathbf{E}^* \left[\alpha^{(\tau - n)} f_\tau \middle| \mathcal{F}_n \right]$$

where $\mathcal{T}_{n,N}$ denotes the set of all stopping times τ, $n \leq \tau \leq N$ with respect to the current information field $\mathcal{F}_n$ of the price process and the conditional expectation is with respect to the equivalent martingale measure $\boldsymbol{P}^*$.

6.4.3 Bandit problem as a control problem

Suppose that there is a collection of d independent projects in the sense that each project yields a sequence of random rewards, and these are considered to be d independent stochastic processes on a given probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The reward process for the project i is denoted by $Z_i = \{Z_i(t), t = 1, 2 \dots\}$ which is a bounded real value stochastic process on $(\Omega, \mathcal{F}) : \mathcal{F}^i = \{\mathcal{F}_t^i, t = 0, 1, \cdots, \infty\}$, that is, the information process for the project i which is a non-decreasing family of sub σ -fields of $\mathcal{F}$. More precisely $\mathcal{F}_t^i$ is the information accumulated during the first t uses of project i.

For the option pricing model, $\mathcal{F}^i = \{\mathcal{F}_t^i, t = 0, 1, \cdots, \infty\}$ should be thought of as the time t price of the risky asset i, this may be the price of one share of stock in a given market. The filtration $\mathcal{F}$ describes how the information is revealed to the investors. The investors know past and present information but nothing about the future.

In the same way, the decision maker in the bandit problem knows all the past and present rewards gained for each project worked out so far. The reward process $Z_i = \{Z_i(t), t = 1, 2 \dots\}$ is adapted to the information process $\mathcal{F}_t^i$ for the given project. It is assumed that:

1. $E \sum_{t=0}^{\infty} \alpha^t |Z_i(t)| < \infty$, where E is the conditional expectation given $\mathcal{F}_0^i$, for a given discount rate $0 < \alpha < 1$ and all $i = 1, 2, \dots d$,

2. Reward processes are statistically independent, *i.e.,* $\mathcal{F}_\infty^i$ are independent σ – fields, for $i = 1, \dots d$,

3. $Z_i(t) \in \mathcal{F}_{t-1}^i$ for $t = 1, 2 \dots \infty$.

One needs to have a definition of an allocation strategy.

Definition 6.4 (Mandelbaum, 1987)

An allocation strategy $T = \{T(t),\ t = 0, 1, \dots\}$ is a d-dimensional discrete stochastic process $T(t) = (T_1(t), T_2(t), \dots T_d(t))$ where $T_i(t) \in \{0, 1, \dots\}$ is the total time allocated to the i^{th} project by the calendar time t, with $T(0) = 0$ such that for all $t = 0, 1, \dots$,

1. $T(t+1)$ is a direct successor of $T(t)$,
2. $T(t)$ is a stopping time with respect to $F = \{F(s), s \in S\}$,

where the filtration F is given by

$$F(s) = F_1(s_1) \vee \cdots \vee F_d(s_d), \qquad s = (s_1, \dots, s_d) \in N^d,$$

and N denotes the set of non-negative integers.

One should note that $\sum_{i=1}^{d} T_i(t) = t$. Given an allocation strategy T, it yields future rewards with present value $R(T)$:

$$R(T) = E \sum_{t=0}^{\infty} \alpha^t \sum_{i=1}^{d} Z_i(T_i(t+1))[T_i(t+1) - T_i(t)] \quad \text{for } t = 0, 1, \dots \tag{6.25}$$

The reward obtained at time $t+1$ comes from the project i for which $T_i(t+1) - T_i(t) = 1$.

The decision maker's objective is to schedule how to choose projects sequentially in time, *i.e.*, to find an allocation strategy $T(t)$ so as to attain the maximal total expected discounted reward

$$\sup_{T(t)} \mathrm{E} \sum_{t=0}^{\infty} \alpha^t \sum_{i=1}^{d} Z_i(T_i(t+1))[T_i(t+1) - T_i(t)]. \tag{6.26}$$

The solution to the problem described by (6.26) is in terms of an Index process.

Index process

We are looking for the strategy $T = \hat{T}$ which achieves the optimal value $V = \sup_T R(T)$. The strategy $\hat{T}$ is described in terms of adapted index processes as follows:

The index process $\Gamma^i = \{\Gamma^i(t),\ t = 0, 1, \dots\}$ is a stochastic sequence associated with project i given by:

$$\Gamma^i(n) = \operatorname*{ess\,sup}_{\tau \geq n+1} \frac{E^{F_i(n)} \sum_{t=n}^{\tau-1} \alpha^t Z_i(t)}{E^{F_i(n)} \sum_{t=n}^{\tau-1} \alpha^t}, \qquad n \in \mathbb{N}, \tag{6.27}$$

where τ is a stopping time with respect to $\mathcal{F}_n^i$ and $E^{F_i(n)}$ is the conditional expectation given $\mathcal{F}_n^i$. We also denote by $\underline{\Gamma}^i(t)$, the lower envelope of $\Gamma^i(t)$, that is,

$$\underline{\Gamma}^i(t) = \inf_{0 \leq t \leq n} \Gamma^i(n).$$

The following theorem explains how these indices are used to identify the optimal policy and expresses the optimal value V in terms of the lower envelopes of the indices.

Theorem 6.4 (Kaspi & Mandelbaum, 1998)
The class of optimal strategies coincides with the class of index strategy $\hat{T}$, which chooses projects with the highest index. Formally, an index strategy is a strategy $\hat{T}$ for which

$$\hat{T}^i(t+1) = \hat{T}^i(t) + 1 \quad \text{only when } \Gamma^i\{\hat{T}^i(t)\} = \sup_{j=1,\dots,d} \Gamma^j\{\hat{T}^j(t)\},$$

for all $t \in N$ and $i \in \{1, 2, \dots, d\}$. Furthermore, the optimal value is given by

$$V = \sum_{t=0}^{\infty} \alpha^t [\sup_{j=1,\dots,d} \underline{\Gamma}^i \{\hat{T}^i(t)\}].$$

The interpretation of this theorem, is that the index process(or rule) selects the project i with the largest current index (6.27). For example, suppose that the project i is engaged at the stage n. Starting from state n, then the maximum attainable expected average reward rate between state n and any arbitrary future state $\tau(\tau > n)$, is achieved if the project is worked out up to the stopping time $\tau - 1$ than any other time t, and is equal to $\Gamma^i(n)$.

6.4.4 Discussion

Starting from the Option pricing problem, an option can be thought of as a game where the reward is the payoff of the option and the option holder pays a fee, that is, the option price for playing this game. For example, in the American option case, the option fee was the rational cost $\mathbb{C}_N^*$ which was the result of optimal stopping problem, that is,

$$C_N^* = \sup_{\tau} \mathbf{E}^* \alpha^\tau f_\tau$$

with discounting factor α and a system of payments $f = (f_n)_{0 \le n \le N}$. The value of the game at time n is given by

$$Y(n) = \operatorname*{ess\,sup}_{\tau \in \tau_{n,N}} \mathbf{E}^* \left[\alpha^{(\tau - n)} f_\tau \middle| \mathcal{F}_n \right]$$

where $\tau_{n,N}$ denotes the set of all stopping times τ, $n \le \tau \le N$ with respect to the current information field $\mathcal{F}_n$ of the price process and the conditional expectation is with respect to the equivalent martingale measure $\boldsymbol{P}^*$.

In the bandit problem, the gambler pays a fixed charge γ to play a bandit after time n until stopping where the total wealth gained up to stopping time $\tau - 1$ is given by

$$u_n(\gamma) = \operatorname*{ess\,sup}_{\tau \ge n+1} \mathrm{E} \left[\sum_{m=n}^{\tau-1} \alpha^m [Z(m) - \gamma] \middle| \mathcal{F}_n \right].$$

In both cases, each game has some rules. In the option pricing case, the investor wants to solve the "investment problem" to reproduce wealth of at least size f_τ at time τ, that is the (random) time the option buyer decides to exercise his option and the investor must be able to pay the amount f_τ as the game (option) requires him to do so.

For the bandit problem, the charge γ can also be understood as a rent per unit time the gambler is asked to pay to have the right to use the bandit j for a given period. If the decision maker decides to use project j for period $(n, \tau - 1)$, the use of this project, it comes to a point which is not profitable. A fair charge would be the charge which if the project is chosen optimally the decision maker will experience no profit or loss. The fair charge is given by:

$$\gamma(x_j) = \sup \left\{ \gamma \,:\, \sup_{\pi} \mathrm{E}_{\pi} \left[\sum_{m=n}^{\tau-1} \alpha^m [Z(m) - \gamma] \middle| \mathcal{F}_n \right] \geq 0 \right\}$$

where the policy π determines a stopping time $\tau \geq n + 1$.

In the case that there are more than one fair charge, the decision maker chooses the smallest one that is $\min_{0 \leq s \leq t} \{\gamma_j(x_j(s))\}$.

It is remarkable to note that both problems have the concept of the "fair value" and that in both cases its value is unique. The trading strategy followed to replicate the value of an option is a martingale transform since there is no arbitrage opportunities as stated in the theorem (4.2). The problem of valuing the American option becomes an Optimal Stopping problem since the trading strategy process should be stopped sometime, that is, the time the option holder decides to exercise his option or the expiry time.

Chapter 7

Conclusions and Further Research

In this thesis, we focused on the Pearson and the Gittins indices. The Pearson index can be viewed as the right decision rule when a set of projects needs to be divided into two groups; the projects which are going to be implemented and the projects which will never be considered. The Pearson index indicates how many projects one should prioritize in order to maximize profitability in an optimisation problem of the Neyman-Pearson Lemma type. The Gittins Index gives the solution to the problem of allocating one's effort over projects sequentially in time so as to maximize expected total discounted reward. Consequently, one can combine the methods by selecting a subset of projects using the Pearson index and then prioritize them sequentially by using the Gittins index.

It would be interesting to develop prioritization indices for parallel-series system and series-parallel system. A prioritisation rule for a sequential decision processes which its each stage is regarded as a successful one if at least k out of n of its sub-stages are successful may be of interest.

In Chapter 3, we studied the problem of maximizing the expected utility of the terminal profit of the chosen projects, in a one period model, subject to linear inequality budget constraint. We showed that when the random profit follows a Normal distribution and the utility function is negative exponential, the problem is equivalent to Quadratic Knapsack problem. We suggested a prioritisation index for selecting projects in a similar fashion to the Pearson index. We also pointed out that it is not always optimal to spend as much as we can, in an optimisation problem with budget constraint when we try to maximise the expected value of exponential utility function.

In Chapter 4, we presented forecasting systems. Probabilities and forecasts are essential to calculate parameters for ranking indices. We showed that the sequence of forecasts is a martingale process. The conditional variance is a supermartingale process. We also noted the similarity of a forecasting system with the pricing process of a European option in complete market due to the martingale property.

In Chapter 5, we introduced the real option approach to capital budgeting problems. We explained why the decision tree approach seemed to give different results than the option pricing approach. We concluded that the option value created from the operating flexibility and strategy flexibility has significant value. Indices studied in Chapter 2 solved problems without taken into account the value associated with flexibility which may arise from operating options. Adjustments should be made in these problems to derive indices which will capture options which may be found in any R&D portfolios problem.

In Chapter 6, we showed that the Multi-Armed bandit problem can be reconciled with the problem of pricing an American option. The common point is that both problems can be thought of as one game with some rules and a strategy which is dependent on the experience of the decision maker. Equivalently, the concepts of the fair value of a project with that of the fair price of the option represent a kind of "fair" charge for both games.

References

Arnold, L. (1971). *Stochastic Differential Equations: Theory and Application.* John Wiley & Sons.

Asawa, M., & Teneketzis, D. (1996). Multi-Armed Bandits with Switching Penalties. *IEEE Transactions on Automatic Control, 41*(3), 328-348.

Baxter, M., & Rennie, A. (1994). *Financial calculus: An introduction to derivative pricing.* Cambridge University Press.

Bellman, R. E. (1956). A problem in the sequential design of experiments. *Sankhya, 16*, 221-229.

Ben-Dov, Y. (1981). Optimal resting procedures for special structure of coherent systems. *Management Sci., 27*(12), 1410-1420.

Berger, J. . (1980). *Statistical Decision Theory.* Springer-Verlag.

Bergman, S. W., & Gittins, J. C. (1985). *Statistical Methods for Pharmaceutical Research Planning.* Marcel Dekker.

Björk, T. (1998). *Arbitrage theory in continuous time.* OUP.

Björk, T., & Johansson, B. (1992). Adaptive prediction and reverse martingales. *Stochastic Process and their Applications, 43*, 191-222.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy, 81*, 637-659.

Black, W. (1965). Discrete Sequential Search. *Inform. and Control, 8*, 156-162.

Bradt, R. N., Johnson, S. M., & Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist., 27*, 1060-1074.

Bretthauer, K. M., & Shetty, B. (1995). The nonlinear resource allocation problem. *Operational Research, 43*(4), 670-683.

Bretthauer, K. M., Shetty, B., & Syam, S. (1995). A branch and bound algorithm for integer quadratic knapsack problems. *ORSA Journal on Computing, 7*(1), 109-116.

Chow, Y., Robbins, H., & Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin Company.

Chun, Y. (1994). Sequential decisions under uncertainty in the R&D project selection problem. *IEEE Transactions on Engineering Management, 41*(4), 404-413.

Chun, Y., & Platt, M. (1992). On the Sequential Selection Problem. *Decision Sciences, 23*, 1251-1262.

Clark, P. K. (1973). A subordinate stochastic process model with finite variance for speculative prices. *Econometrica, 41*, 135-155.

Dean, V. D. (1966). *"chapter XII: Stochastic Networks in Research Planning", in research program effectiveness, m. c. yonitis et al., eds.* New York: Gordon and Breach.

DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill.

Dixit, A., & Pindyck, R. (1994). *Investment Under Uncertainty.* Princeton University Press.

Dixit, A., & Pindyck, R. (1995). The Options Approach to Capital Investment. *Harvard Business Review, 81*, 105-115.

Elton, E. J., & Gruber, M. J. (1975). *Finance as a Dynamic Process.* Prentice Hall.

Gallo, G., Hammer, P., & Simeone, B. (1980). Quadratic Knapsack Problems. *Mathematical Programming, 12*, 132-149.

Gittins, J. C. (1979). Bandit Processes and Dynamic Allocation Indices. *J. Roy. Stac. Soc.*, *B*(14), 148-177.

Gittins, J. C. (1989). *Multi-armed Bandit Allocation Indices.* John Wiley & Sons.

Gittins, J. C. (1996). Quantitative methods in the planning of pharmaceutical research. *Drug Information Journal, 30*, 479-487.

Gittins, J. C. (1997). Why crash pharmaceutical research? *R&D Management, 27*(1), 79-85.

Gittins, J. C., & Jones, D. M. (1974). A dynamic allocation index for sequential design of exeperiments. *In progress in Statistics(J. M. Gani, K. Sarkadi and I. Vincze, Eds.)*, 241-266.

Gittins, J. C., & Wang, Y. G. (1992). The learning component of dynamic allocation indices. *The Annals of Statistics, 20*(3), 1625-1636.

Gourieroux, C. (1997). *ARCH Models and Financial Applications.* Springer.

Hadley, G. (1964). *Nonlinear and dynamic programming.* Addison-Wesley Publishing Company.

Harrison, M., & Kreps, D. (1979). Martingales and Multiperiod Securities Markets. *Journal of Economic Theory, 20*, 381-408.

Harrison, M., & Pliska, S. (1981). Martingales and stochastic intergrals in the theory of continuous trading. *Stochastic Processes and their Applications, 11*(3), 215-260.

Hodges, S. D. (1998). Introduction to Financial Economics. *Statistics in Finance edited by David J. Hand and Saul D. Jacka*, 187-203.

Huang, C.-f., & Litzenberger, R. (1988). *Foundations for Financial Economics.* Prentice Hall.

Ibaraki, T., & Katoh, N. (1988). *Resource Allocation Problems.* MIT Press, Cambridge, Mass.

Ishikida, T., & Varaiya, P. (1994). Multi-Armed Bandit Problem Rivisited. *Journal of Optimization Theory and Applications*, *83*(1), 113-154.

Joyce, W. B. (1971). Organizations of Unsuccessful R&D Projects. *IEEE Trans. Engineering Management*, *EM-18*(3), 57-65.

Kabanov, Y. M., & Kramkov, D. O. (1994). No-arbitrage and equivalent martingale measures: An elementary proof of the Harrison-Pliska theorem. *Theory Probab. Appl.*, *39*, 523-527.

Kaspi, H., & Mandelbaum, A. (1998). Multi-Armed Bandits in Discrete and Continuous Time. *The Annals of Applied Probability*, *8*(4), 1270-1290.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objective: Preferences and Value Tradeoffs.* John Wiley & Sons.

Kester, W. C. (1984). Today's Options for Tomorrow's Growth. *Harvard Business Review*, *62*, 153-160.

Kwan, C. C. Y., & Yuan, Y. (1988). Optimal sequential selection in capital budgeting: A shortcut. *Financial Management*, *17*, 54-59.

Lamberton, D., & Lapeyre, B. (1996). *Introduction to stochastic calculus applied to finance.* Chapman & Hall.

Lo, A. W., & Wang, J. (1995). Implementing Option Pricing Models When Asset Returns Are Predictable. *Journal of Finance*(1), 87-129.

Mandelbaum, A. (1986). Discrete Multi-Armed Bandits and Multiparameter Processes. *Probability Theory Related Fields*, *71*, 129-147.

Mandelbaum, A. (1987). Continuous Multi-Armed Bandits and Multiparameter Processes. *The Annals of Probability*, *15*(4), 1527-1556.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*, 77-91.

Markowitz, H. (1990). *Portfolio selection.* Blackwell.

Martello, S., & Toth, P. (1997). *Knapsack Problems.* John Wiley & Sons.

McDonald, R., & Siegel, D. (1986). The value of waiting to invest. *Quarterly Journal of Economics*, *101*, 707-727.

Myers, S. C. (1984). Finance theory and financial strategy. *Interfaces*, *14*(1), 126-137.

Nielsen, S., & Zenios, S. (1992). Massively parallel algorithms for singly constrained convex programs. *ORSA Journal on Computing*, *4*, 166-181.

Øksendal, B. (1995). *Stochastic Differential Equations.* Springer.

Ostaszewski, A. (1990). *Advanced Mathematical Methods.* CUP Press.

Pearson, A. (1972). The use of ranking formulae in R&D projects. *R&D*, *2*, 69-73.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Addison-Wesley, Reading, MA.

Regan, P. J., & Senn, S. (1997). Project evaluation in drug development: a case study in collaboration. In *The Practice of Bayesian Analysis* (p. 259-277). Arnold.

Rothkopf, M. H. (1966). Scheduling independent tasks on parallel processors. *Management Science*, *12*, 437-447.

Rothkopf, M. H., & Smith, S. A. (1984). There are no undiscovered priority index sequencing rules for minimizing total delay. *Operations Res.*, *32*, 451-456.

Senn, S. J. (1996). Some statistical issues in project prioritization in the pharmaceutical industry. *Statistics in Medicine*, *15*(1), 2689-2702.

Senn, S. J. (1997). *Statistical Issues in Drug Development.* John Wiley & Sons.

Senn, S. J. (1998). Further statistical issues in project prioritization in the pharmaceutical industry. *Drug Information Journal*, *32*, 253-259.

Shiryaev, A. N., Kabanov, Y. M., Kramkov, D. O., & Melnikov, A. V. (1994). Towards the theory of pricing of options of both of European and American types. I. Discrete time. *Theory Probab. Appl.*, *39*, 14-60.

Smith, J. E., & Nau, R. F. (1995). Valuing Risky Projects: Option Pricing Theory and Decision Analysis. *Management Science*, *41*, 795-816.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of evidence of two samples. *Biometrica*, *25*, 285-294.

Thompson, W. R. (1935). On the theory of apportionment. *Amer. J. Math.*, *57*, 450-456.

Trigeorgis, L. (1996). *Real Options: Managerial Flexibility and Strategy in Resource Allocation.* MIT Press.

Walrand, J. C. (1988). *An Introduction to Queueing Networks.* Englewood Cliffs, NJ: Prentice-Hall.

Weber, R. (1992). On the Gittins Index for Multi-Armed Bandits. *The Annals of Applied Probability*, *2*(4), 1024-1033.

Whittle, P. (1980). Multi-Armed Bandits and the Gittins Index. *J. R. Statist. Soc. B*, *42*(2), 143-149.

Zipkin, P. H. (1980). Simple ranking methods for allocation of one resource. *Management Science*, *26*(1), 34-43.