# Financial Mathematics

## An Introduction



S. Chandra
S. Dharmaraja
Aparna Mehra
R. Khemchandani

Alpha Science

# Financial Mathematics

## An Introduction

# Financial Mathematics

## An Introduction

Suresh Chandra
S. Dharmaraja
Aparna Mehra
Reshma Khemchandani



Alpha Science International Ltd.
Oxford, U.K.

**Financial Mathematics**
An Introduction
516 pgs. | 48 figs. | 23 tbls.

**Suresh Chandra**
**S. Dharmaraja**
**Aparna Mehra**
Department of Mathematics
Indian Institute of Technology
Hauz Khas, New Delhi

**Reshma Khemchandani**
RBS Global Banking and Markets
DLF Building, 7, Cybercity
DLF Phase 3, Sector 25 A
Gurgaon, Haryana



ALPHA SCIENCE INTERNATIONAL LTD.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

**www.alphasci.com**



Printed from the camera-ready copy provided by the Authors.

ISBN 978-1-84265-654-9
E-ISBN 978-1-78332-002-8

Printed in India

*To all Financial Mathematics students & faculty*

# Preface

Last decade has seen a sudden upsurge in courses on Financial Mathematics at various Universities/Institutes. There are many Universities and Institutes running full time Masters Program in Financial Mathematics. Though there are large number of financial mathematics books published, a close look at the available titles indicate two extreme views on the presentation style. There are excellent books requiring very heavy dose of Measure Theory and Functional Analysis and at the same time, there are very good books requiring only college level Calculus and Probability Theory. Thus these texts can possibly be classified as either "too difficult" or "too easy" depending upon the level of Mathematics used. There is a general feeling that those texts that are "too difficult" lack in economic/physical interpretations and those which are "too easy" forgo precise mathematical presentation. Therefore there seems to be a strong need to develop a text on Financial Mathematics that takes a balanced approach and evolves an optimal trade-off between "precise mathematical presentation" and "economic/physical interpretations". This textbook is a modest attempt in this direction which has evolved from certain courses on Financial Mathematics that we have been teaching at I.I.T. Delhi for last many years.

This book attempts to provide an introductory text on Financial Mathematics to cater to the needs of students at various Universities/Institutes. Apart from presenting two Nobel Prizes winning theories of *Black, Scholes and Merton for option pricing* and *mean-variance approach of Markowitz for portfolio optimization*, the text also includes now standard topics of *interest rate* and *interest rate derivatives*. Certain interesting and useful topics like *optimal trading strategies, credit risk management* and *Monte Carlo simulation*, which are normally not covered in a text of this kind, are also included here. A significant portion of the book is devoted to study *stochastic of finance* that is very much needed to understand basic concepts related to pricing of derivatives.

This book is most suitable as a text for any introductory financial mathematics course in M.Sc (Financial Mathematics/Financial Engineering), M.Sc (Mathematics/Statistics/Operations Research), B.Tech/B.E., B.Sc(Hons.) and M.B.A. programs. The full text can be covered in two semesters. However if only 'one' one-semester course is to be designed then the instructor, depending upon the background of the audience, can choose appropriate topics. The book can also be used as a reference by researchers and practitioners. We sincerely hope that the present text will inspire students and faculty to learn and explore the challenges and fascination the subject "financial mathematics" offer. Believe us, it is an awesome subject.

The prerequisites to understand and appreciate this text are some standard undergraduate courses on Calculus, Linear Algebra, Probability Theory, and a course on Optimization Theory. The "difficult part", namely the material on *stochastic of finance* is not a prerequisite as it has been presented here in the book. The book is written in a simple classroom teaching style and a special effort is made so as to make the readers appreciate both - the *mathematics of finance* and *economic/physical interpretations*. There are numerous small illustrative examples throughout the book and also end chapter exercises for practice by students. Each chapter ends with "Summary and Additional Notes" section, which tries to provide some recent references for further study. An added feature of the book is MATLAB codes for some selected problems in Financial Mathematics.

Although every care has been taken to make the presentation error free, some errors may still remain and we hold ourselves responsible for that. The readers are requested to kindly communicate errors, if any, at chandras@maths.iitd.ac.in (e-mail address of S.Chandra).

In the long process of writing this book we have been encouraged and assisted by many individuals. In particular, an honorable mention goes to our long standing friends and well-wishers; Professors - C. R. Bector, M. M. Chawla, S. K. Gupta, S. K. Bhatt, S. S. Appadoo, V. Arunachalam, Karmeshu, B. R. Handa, S. S. Yadav, Jayadeva, Wagish Shukla, B. Chandra, R. K. Sharma, B. S. Panda, S. Chandrasekhra Rao, K. Sreenadh, Nomesh Bolia, Joydeep Dutta, Pankaj Gupta, C. S. Lalitha and Anulekha Dhara. Thank you for always being very supportive and encouraging and especially for your interest in this work.

We gratefully acknowledge the book grant provided by the I.I.T. Delhi and thank Professor Prem K. Kalra and Professor R. K. Sharma for help in this regard.

A very heartfelt thanks and appreciation to our students Ms. Arti Singh, Ms. Priyanka Kedia, Ms. Supriya, Ms. Amita Sharma and Mr. Krishna Prasad for sparing their valuable time during the preparation of the manuscript. Ms. Arti Singh read the entire manuscript from a student point of view and made useful suggestions. The MATLAB programs included here are due to Ms. Amita Sharma and Mr. Krishna Prasad. We like to exclusively mention our sincere thanks to Dr. Abha Aggarwal who undertook overall responsibility of preparing the manuscript. Without her help it would have been extremely difficult to complete this book in the present form.

Finally yet importantly, we are thankful to Mr. N. K. Mehra and his team at Narosa Publishing House for their co-operation and understanding in the presentation of the book.

Suresh Chandra
Dharmaraja Selvamuthu
Aparna Mehra
Reshma Khemchandani

# Contents

# 1
# Financial Mathematics: An Overview

## 1.1 Introduction

The word *financial mathematics* refers to the application of mathematics to study the problems arising in the area of finance. Here the word mathematics has been used in a wider sense so as to include subjects like *probability* and *statistics*, *stochastic processes*, *optimization, econometrics, numerical analysis* and *partial differential equations.* Also the scope of the word *finance* could be very broad but here it should be understood in terms of *investment* of money for the purpose of receiving more money (hopefully!!) at sometime in future. Thus *financial mathematics* could also be termed as *investment science* or *investment theory.*

Though investment as an art form has always been there, it developed as science/theory due to the pioneering work of Fisher Black, Myron Scholes, Robert Merton, Harry Markowitz, William Sharpe and John Lintner amongst many others. In fact, Merton and Scholes got Nobel Prize in Economics (in 1997) for their work on pricing of option carried out during late 60's and early 70's. Black could not share this prize as he died of illness before 1997. Earlier Markowitz along with Sharpe (Markowitz's Ph.D. student) were awarded Nobel Prize (in 1990) for their work on portfolio optimization carried out during late 50's and early 60's. They shared Nobel Prize with another famous economist Merton Miller.

The above extra ordinary theoretical developments had great impetus on research in finance. This together with globalization of investment activities and tremendous power of computing technology has made the financial mathematics area intellectually very fertile. There is a significant growth in theory as well as in applications in the areas of derivative pricing, portfolio management, interest rate modeling, credit risk methodologies and algorithmic trading.

The aim of this chapter is to introduce basic notions and assumptions of a simple financial market model so as to facilitate the presentation of some simple investment problems for illustration purposes. Our discussion here is motivated by some of the excellent texts. We shall particularly like to mention Capinski and Zastawniak [25], Hull [65] and Luenberger [85].

## 1.2 Some Basic Terminology

Any financial market model has to deal with the trading of financial assets. By a financial asset we mean a financial instrument that can be bought or sold. Here an asset will always mean a financial asset. It is common knowledge that some assets are *risk free* and some are *risky*. For example a bank deposit or a bond issued by a government are risk free where as stocks are risky. To introduce certain notations and terminology for a financial market we first assume a very simple scenario. We assume that only two assets are being traded. These are, one risk free asset (bond) and one risky asset (stock). The other parameter in modeling a financial market is the *time horizon*. The simplest of these is the *single period model* where we restrict the time scale to be two instants only, namely $t = 0$ and $t = 1$ and any transaction takes place at these two time instants only. Realistically it has to be a multi period or continuous time scenario. But here we restrict our discussion to the case of single period only, the multi period and continuous time models will be discussed later.

Let $t = 0$ and $t = 1$ be two time instants. Let $B(0)$ denote the price of one bond at $t = 0$ and $B(1)$ be the corresponding price at $t = 1$. Similarly let $S(0)$ denote the price of one share of stock at $t = 0$ and $S(1)$ be the corresponding price at $t = 1$. As bond is a risk free asset, both $B(0)$ and $B(1)$ are known at $t = 0$. But for the case of stock, which is a risky asset, though $S(0)$ is known, $S(1)$ remains uncertain. The stock price at $t = 1$, namely $S(1)$, may go up or down with certain probability, i.e. $S(1)$ is a random variable. Thus for a risk free asset both $B(0)$ and $B(1)$ are deterministic, but for a risky asset $S(0)$ is deterministic while $S(1)$ is a random variable. We next define the *rate of return of an asset*. For the case of bond it is defined as

$$r_B = \frac{B(1) - B(0)}{B(0)} ,$$

and for the case of stock it is defined as

$$r_S = \frac{S(1) - S(0)}{S(0)} .$$

Here we note that $r_B$ is known exactly but $r_S$ is a random variable because $S(1)$ is a random variable. In the finance literature, rate of return is called *return* only.

Thus when we refer to the return of an asset we really mean the rate of return of that asset.

Let an investor hold $x$ shares of stock and $y$ units of bond at $t = 0$. The pair $(x, y)$ is called a *portfolio* and it is denoted by $P : (x, y)$. The value of the portfolio $P : (x, y)$ at time $t = 0$ is given by

$$V_P(0) = x\ S(0) + y\ B(0).$$

Similarly the value of the portfolio at time instant $t = 1$ is given by

$$V_P(1) = x\ S(1) + y\ B(1).$$

The quantity

$$r_P = \frac{V_p(1) - V_p(0)}{V_p(0)}\ ,$$

is called the *rate of return of the portfolio* $P : (x, y)$. We note that $r_P$ is a random variable because $V_P(1)$ is so.

The units of $B(0)$, $B(1)$, $S(0)$ and $S(1)$ are units of money and here they are always taken in Rupees. Thus $B(0) = 100$ etc. will always mean $B(0) = Rs\ 100$ etc. We shall follow this convention throughout the book.

Since the portfolio return $r_P$ is a random variable we can talk of its expected value and variance. The quantity $E(r_P)$ is called the *expected return of the portfolio* $P : (x, y)$ and the *standard deviation* $\sigma_P$ *of* $r_P$ is called its *risk*. Obviously if we are given a choice between two portfolios with the same expected return, we should choose the one for which the risk is least. In a similar manner, if the risk levels of two portfolios are same we should choose the one for which return is maximum. This is a typical problem of portfolio optimization. We shall have opportunity of discussing a general $n$-asset problem of portfolio optimization in context of mean-variance theory of Markowitz in later chapters of the book. The concept of risk itself has been a topic of research in the recent past. Is standard deviation or variance the *correct* measure of risk? Are there some other *better* definitions of risk? All these questions need to be answered and we plan to discuss them in greater detail in a later chapter.

**Example 1.2.1** *Let* $B(0) = 100$, $B(1) = 110$, $S(0) = 80$ *and*

$$S(1) = \begin{cases} 100, & \text{with probability } 0 \cdot 8 \\ 60, & \text{with probability } 0 \cdot 2. \end{cases}$$

*What is the expected return on the stock? Consider the portfolio* $P : (x = 50, y = 60)$. *Determine the expected return and the risk for the given portfolio.*

**Solution** The return on bond is obviously 10%. But for the stock

$$r_S = \begin{cases} \dfrac{100-80}{80}, & \text{with probability } 0\cdot 8 \\[2ex] \dfrac{60-80}{80}, & \text{with probability } 0\cdot 2, \end{cases}$$

i.e.

$$r_S = \begin{cases} 0\cdot 25, & \text{with probability } 0\cdot 8 \\ -0\cdot 25, & \text{with probability } 0\cdot 2. \end{cases}$$

Therefore $E(r_S) = (0\cdot 25)(0\cdot 8) + (-0\cdot 25)(0\cdot 2) = 0\cdot 15$.

Further as $x = 50$ and $y = 60$, we have

$$V_P(0) = (50 \times 80) + (60 \times 100) = 10,000.$$

Also

$$\begin{aligned} V_P(1) &= \begin{cases} (50 \times 100) + (60 \times 110), & \text{with probability } 0\cdot 8 \\ (50 \times 60) + (60 \times 110), & \text{with probability } 0\cdot 2 \end{cases} \\ &= \begin{cases} 11,600, & \text{with probability } 0\cdot 8 \\ 9,600, & \text{with probability } 0\cdot 2. \end{cases} \end{aligned}$$



Therefore

$$\begin{aligned} r_P &= \frac{V_P(1) - V_P(0)}{V_P(0)} \\ &= \begin{cases} 0\cdot 16, & \text{with probability } 0\cdot 8 \\ -0\cdot 04, & \text{with probability } 0\cdot 2. \end{cases} \end{aligned}$$

This gives the expected return on the portfolio as

$$E(r_P) = (0\cdot 16)(0\cdot 8) + (-0\cdot 04)(0\cdot 2) = 0\cdot 12.$$

We have defined the standard deviation of the portfolio as its risk. Therefore the risk of given portfolio is

$$\sigma_P = \sqrt{E\left(r_p - E(r_p)\right)^2} = \sqrt{(0\cdot 16 - 0\cdot 12)^2 \times (0\cdot 8) + (-0\cdot 04 - 0\cdot 12)^2 \times (0\cdot 2)}$$
$$= 0\cdot 08.$$

□

## 1.3 Certain Assumptions

We now state certain assumptions which help in building a simple mathematical model of a typical financial market. The real world financial market is extremely complex and any mathematical model has to trade off between reality and simplifications which make the model amenable to analytical or numerical solution. Apart from the time horizon, it is the dynamics of the assets which plays a very crucial role in the modeling aspect of a financial market. Thus, for a general case, we need to describe the dynamics of the stock price $S(t)$ as $t$ varies over a continuum. Since dynamics generally refers to a differential equation, we expect that the stock price movement will be represented by a differential equation. But things are not that simple as for each $t$, $S(t)$ is a random variable i.e. it is a random process. So there has to be an appropriate stochastic process component in this differential equation leading to something called a *stochastic differential equation* (SDE). So for a proper understanding of this dynamics we need to study stochastic calculus and stochastic differential equations more deeply. Since more than one stock may be in the portfolio, one has to deal with more than one SDE and also to take into account their interaction as stocks may be correlated. But as agreed in Section 1.2, in this chapter we restrict our discussion to the case when only two assets (one bond and one stock) are being traded and the time horizon is only one period (from $t = 0$ to $t = 1$).

For the present we do not take the general case and state certain assumptions for our single period case only. We should take all these assumptions only as a first step towards modeling and try to modify them later so as to make the model more realistic.

**Assumption 1 (Randomness)**
The stock price $S(1)$ at $t = 1$ is a random variable with at least two different values.

**Assumption 2 (Positivity of Prices)**
All bond and stock prices are strictly positive, i.e. $B(0) > 0$, $B(1) > 0$, $S(0) > 0$ and $S(1) > 0$.

**Assumption 3 (Divisibility, Liquidity and Short Selling)**
Let the investor hold $x$ shares of stock and $y$ units of bond. Then $x \in \mathbf{R}$ and $y \in \mathbf{R}$.

Thus the investor is permitted to hold any units of bond and any number of shares of stock. These numbers may be zero, positive or negative reals. This property is referred to as *divisibility* property. In a real market this property is expected to hold when the volume of transactions is large compared to unit prices.

As there are no bounds on $x$ and $y$ they can be as large or as small as we wish. In other words, an investor can buy or sell arbitrary quantities at the market price on demand. This property is called *liquidity*. In reality there is always a restriction on the volume of trading so this property may not hold in general.

The next thing to understand is the interpretation of sign restriction on $x$ and $y$. If $x > 0$ then it obviously means that at $t = 0$, the investor is buying $x$ shares of stock and this we state as "The investor has taken a *long position* on stock". Similarly we interpret $y > 0$. However, if $x < 0$ (respectively $y < 0$), we say that "the investor has taken a *short position* on stock (respectively bond)". Thus if $x > 0$ and $y < 0$, we say that investor has taken a long position on stock and a short position on bond. If the investor has taken a short position on an asset we say that the asset is *shorted*.

But what is the physical interpretation of a short position ? A short position in a risk free asset (bond) means borrowing cash, where the interest rate is determined by the bond price. Thus if $y = -10,\ B(0) = 100$ and $B(1) = 110,$ then it means that the investor is borrowing $Rs$ 1000 from the bank where the interest rate is 10%. We say that the short position in the bond is closed when the borrowed cash along with interest is repaid to the bank or lender.

A short position in the stock is slightly more complicated. This is best understood in terms of *short selling*. Short selling is a way to make profit if price goes down. Suppose that the investor feels that the price of a stock is overvalued, so it makes sense in selling the stock and purchasing it when the price goes down. In this situation the investor can resort to short selling. Here the investor borrows the stock from someone (say broker), sells it and uses the amount received to make some other investment. But the original owner of the stock has all the rights vested to him/her. Thus he/she receives all dividends due and has the full right to sell the stock any time. In this situation, the investor has to return the stock to the original owner. If he/she does not have stock with him/her, he/she may have to buy from the market at the prevailing market price. Therefore it becomes imperative for the investor to keep sufficient resources (cash) with him/her to fulfil this obligation. We say that the short position on the stock is closed when the stock is returned to the original owner by the investor. The investor may return the stock of his/her own or when the owner asks for the same. In practice the investor is also asked to give a security deposit with the broker from which the stock is borrowed. But in theory we do not consider this.

The above discussion suggests that to meet his/her obligations as per Assumption 3, the wealth of the investor should never be negative. This is precisely the Assumption 4 given below.

**Assumption 4 (Solvency)**
The wealth of the investor should never be negative, i.e. $V_P(t) \geq 0$ at all times $t$. For the single period case it simply means that $V_P(0) \geq 0$ and $V_P(1) \geq 0$.

A portfolio $(x, y)$ satisfying the *solvency* property is termed as an *admissible portfolio*.

Short selling is generally considered to be extremely risky because in short selling there are possibilities of unlimited losses. Let the borrowed asset be sold for an amount $X_0$ at $t = 0$ and later at $t = 1$ it be purchased for an amount $X_1$ so as to close the short position. If $X_1$ is less than $X_0$ then a profit of $(X_0 - X_1)$ is made. In other words short selling is profitable if the asset price goes down. But if the asset value increases then the loss is $(X_1 - X_0)$, which can be arbitrary large since $X_1$ can increase arbitrary. Because of this not all financial institutions allow short selling. In some institutions it may be allowed only on some assets and that too with certain restrictions.

Let us consider an example in this regard. Let $S(0) = Rs\ 10$ and let an investor decide to short 100 shares of this stock at $t = 0$. For this he/she borrows 100 shares from a broker and sells these in the stock market to receive an amount of $Rs\ 1000$. Let at $t = 1$, the stock price $S(1)$ become $Rs\ 9$, so the investor buys back 100 shares from the market at $Rs\ 900$ and gives these shares to the broker to close the short position. This has been a favorable position to the investor as stock price has gone down. In fact the investor has made a profit of $Rs\ 100$ due to short selling. If the price had gone up, the investor would obviously had made losses.

Now imagine another investor who has not gone for short selling but rather has purchased 100 shares in the beginning at $t = 0$. Imagine that this investor has to sell these shares at $t = 1$ (the problem being a single period only). This will entail a loss of $Rs\ 100$ to this investor. The rate of return on the asset for this investor will be

$$r = \frac{900 - 1000}{1000} = -0 \cdot 10$$

i.e. $-10\%$.

This example clearly illustrates that short selling allows an investor to convert a negative rate of return into a profit. This is because the original investment of the investor is also negative. It gives a profit of (original investment)×(rate of return), i.e. $(-1000)(-0 \cdot 10) = 100$.

**Assumption 5 (Discrete Unit Prices)**
By Assumption 1, the stock price $S(1)$ at time $t = 1$, is a random variable taking at least two distinct values. What we assume here is that $S(1)$ is a random variable taking only finitely many values.

This assumption is again not realistic as $S(t)$ at any time $t$ can take values in a continuum. But again it is taken only for the sake of simplicity and will be generalized later.

We next present *no arbitrage principle* and *the law of one price*. In fact the law of one price is a consequence of no arbitrage principle. We say that there is an arbitrage opportunity in the market when someone, with no invested capital, has a positive probability of achieving a positive return with no risk of loss. So in our common usage of English language, arbitrage opportunity really means *free lunch*. It is certainly not desirable in a fair market, and therefore we make the assumption given below.

**Assumption 6 (No Arbitrage Principle)**
There is no portfolio with initial value $V(0) = 0$ such that $V(1) \geq 0$ with probability 1 and $V(1) > 0$ with non-zero probability. Thus if the initial value of a portfolio is zero, i.e. $V(0) = 0$ (no initial investment), and $V(1) \geq 0$ (no risk of loss), then $V(1) = 0$ (no profit) with probability 1.

As a consequence of *no arbitrage principle*, no investor can lock in a profit without risk of loss and no initial endowment. In other words, there are no zero cost investment strategies that allow an investor to make a profit without taking some risk of a loss. In this context, we may note that investing in a Government bond (or fixed deposit) is not an arbitrage though it guarantees a positive net return because capital is invested initially. Some readers may like to analyze if short selling is an arbitrage opportunity? The answer is obviously in negative.
Assumption 6 can also be stated mathematically as follows
There is no portfolio such that for some $t$

(i) $V(0) = 0$,
(ii) $V(t) \geq 0$ with probability 1,
(iii) $P[V(t) > 0] > 0$ ,

where $V(t)$ is the wealth of the portfolio at time $t$. For the single period case it means that there is no portfolio such that

(i) $V(0) = 0$,
(ii) $V(1) \geq 0$ with probability 1,
(iii) $P[V(1) > 0] > 0$.

The following is rather an artificial example of arbitrage opportunity. Consider two banks (say Bank 1 and Bank 2) that offer to loan money or accept deposits at the same rate of interest but these are different in the two banks. Let the rate at Bank 1 for loan and deposits be $10\%$ and at Bank 2 be $12\%$. So an investor can go to Bank 1 and borrow say $Rs\ 10,000$ at $10\%$ and then deposit that $Rs\ 10,000$ at $12\%$ interest at Bank 2. At $t = 1$, the investor will get $Rs\ 11,200$ and then he/she can return $Rs\ 11,100$ (borrowed money and $10\%$ interest on it) to Bank 1. In this process he/she has earned $Rs\ 100$ (guaranteed) without making any investment of his/her own. Obviously such opportunities are rare and instantaneously both banks will adjust the same rate of interest and there will be no arbitrage.

The above discussion suggests that in a financial market, arbitrage opportunity cannot prevail for extended periods of time and therefore it justifies Assumption 6. This assumption is extremely important for any meaningful study in financial mathematics and we shall have opportunity to see its application throughout the book.

As a consequence of no arbitrage principle we have the *law of one price* or also called the *comparison principle*. It essentially tells that if two financial instruments have exactly the same pay-offs, then they have the same price.

To have a better idea of the law of one price, let us imagine two portfolios, say Portfolio $P$ and Portfolio $Q$. Let us consider the case of single period only and refer to the time instants as $t = 0$ and $t = 1$. The two portfolios $P$ and $Q$ have known (determinate) value at $t = 0$ but their values at $t = 1$ (also called pay-offs) are random variables. Thus $V_P(0)$ and $V_Q(0)$ are deterministic, but $V_P(1)$ and $V_Q(1)$ are random variables, whose values will depend upon the state of economy $\omega \in \Omega$. Therefore,

$$V_P(1) \in \{V_{P,1}(\omega) : \omega \in \Omega\} \ ,$$

where $V_{P,1}(\omega)$ refers to the value taken by the random variable $V_P(1)$ when the state of the economy is $\omega$. In particular, if there are $m$ possible states of economy, i.e. $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ then

$$V_P(1) \in \{V_{P,1}(\omega_1), \ldots, V_{P,1}(\omega_m)\} \ ,$$

where $V_{P,1}(\omega_i)$ refers to the value of the random variable $V_P(1)$ when the state of economy is $\omega_i (i = 1, 2, \ldots, m)$. In a similar manner

$$V_Q(1) \in \{V_{Q,1}(\omega) : \omega \in \Omega\}.$$

The law of one principle states that if $V_{P,1}(\omega) = V_{Q,1}(\omega)$, for all $\omega \in \Omega$, then $V_P(0) = V_Q(0)$.

To see that it should be true, suppose that $V_P(0) > V_Q(0)$. Then at $t = 0$ an investor can sell the costlier portfolio (Portfolio $P$) and purchase the cheaper portfolio (Portfolio $Q$), pocketing the difference. At $t = 1$, no matter what the state of economy is in, the investor receives the common final value of the portfolios. Thus the investor looses nothing at the end and made profit. But this violates the principle of no arbitrage. Hence $V_P(0)$ cannot be more than $V_Q(0)$. In a similar manner we can show that $V_Q(0)$ cannot be more than $V_P(0)$. Hence $V_P(0) = V_Q(0)$.

The law of one price has proved to be a very effective tool in pricing *derivative securities*, which we shall discuss in detail in later chapters.

Apart from the above main assumptions it also assumed that *there are no commissions or transaction costs and the lending rate is equal to the borrowing rate.*

## 1.4 Derivative Securities

Some financial securities have the property that their values depend upon the value of another security. In this case, the former security is called a *derivative security* of the later security which is called the *underlying security* for the given derivative security. We often call a derivative security simply as *derivative* and the corresponding underlying security simply as *underlying*. Derivative securities are also called *contingent claims*.



Our aim here is to introduce some simple derivatives and illustrate them through examples. In particular we shall describe a *forward contract, a call option* and a *put option*. These derivative securities in their pure form have stock as underlying security. But sometimes it is possible to have a derivative whose underlying is another derivative. So we can have options on *futures contracts*. Thus a given financial entity can be derivative under some circumstances and an underlying in other circumstances. In fact sometimes underlying may even be interest rate and currency exchange rate which do not look like typical financial securities. Our discussion here remains valid even when some other asset, other than stock, is the underlying security. We shall also illustrate the application of no arbitrage principle (Assumption 6) for their *pricing*. All these topics will be further continued in the later chapters of the book.

**Forward Contracts**

A *forward contract* is a derivative security whose underlying is a stock and it has the below given basic characteristics.

(i) It is an agreement to buy or sell the stock at a specified future time, called *delivery date*, for a fixed price $F$, called the *forward price* which is agreed at $t = 0$. Obviously for the single period model, delivery date has to be $t = 1$.

(ii) An investor who agrees to buy the stock is said to enter into a *long forward contract* or take a *long forward position*.

(iii) An investor who agrees to sell the stock is said to enter into a *short forward contract* or take a *short forward position*.

(iv) No money is paid at $t = 0$ when a forward contract is exchanged.

(v) A forward contract guarantees that the stock will be bought (for long position) or sold (for short position) for forward price $F$ at the delivery date.

In general, the party holding the long forward contract will benefit if $S(1) > F$ and suffer a loss if $S(1) < F$. Therefore the pay-off of a long forward contract is $S(1) - F$, which could be zero, positive or negative. For a short forward position, the pay-off is obviously $F - S(1)$.

The question which we have to answer is as follows: What should be the forward price $F$? Though a formal theoretical formula will be derived in Chapter 2, here we explain the logic through an example.

**Example 1.4.1** *Let* $B(0) = 100$, $B(1) = 110$, $S(0) = 50$. *Determine the forward price* $F$ *of a forward contract on the given stock.*

**Solution** Here delivery is one period (from $t = 0$ to $t = 1$) which let us take as one year. As $B(0) = 100$ and $B(1) = 110$, the nominal rate of interest is 10%. Since $S(0) = 50$, the forward price $F$ should at least be the amount which the holder of the forward contract would have got by depositing $Rs$ 50 in the bank for one year with rate of interest 10%. Thus $F$ should be at least $Rs$ 55.

We shall now argue that $F$ should be exactly $Rs$ 55. If possible let $F > 55$. Then at $t = 0$, we construct the portfolio $P : (x = 1, y = -\frac{1}{2}, z = -1)$, where $z$ denotes the units of forward contracts in the portfolio. The interpretation of the portfolio $P$ so constructed is that at time $t = 0$ we borrow $Rs$ 50 as $(y = -\frac{1}{2})$, buy the stock for $S(0) = 50$ as $(x = 1)$ and enter into a short forward contract as $(z = -1)$ with forward price $F$ and delivery date $t = 1$. This is the interpretation of $x = 1, y = -\frac{1}{2}$ and $z = -1$.This gives

$$V_P(0) = x\ S(0) + y\ B(0) = (1 \times 50) + (-\frac{1}{2} \times 100) = 0.$$

Here it may be noted that in $V_P(0)$ there is no term corresponding to the short forward contract $(z = -1)$ because in forward contract there is no exchange of money at the time when the contract is initialized, i.e. $t = 0$.

Now at $t = 1$, we close the short forward position by selling the asset for $Rs\ F$ and close the risk free position by paying $\frac{1}{2}B(1) = \frac{1}{2} \times 110 = Rs\ 55$. Therefore at $t = 1$, $V_P(1) = F - 55$, which is strictly positive as we have assumed that $F > 55$. This clearly violates the no arbitrage principle. Hence $F = 55$.

□

**Call and Put Options**

Similar to forward contracts, the call and put options in their basic form are also derivative securities whose underlying is stock. An option has the below given characteristics.

(i) The owner of a call option gets the right to buy stock at a future date (called *maturity* or *expiration date*) at a predetermined price (called *strike* or *exercise price*). In a similar manner, the owner of a put option gets the right to sell stock at maturity at a predetermined strike price.

(ii) The owner of the option (whether call or put) has the right without any obligation. Thus at maturity the owner may not exercise the option if he/she so wishes.

(iii) The person who has bought the option is called its *holder* and the person who has sold the option is called its *writer*.

(iv) As explained in (ii) above, the holder of the option has the right but no obligation whereas the writer of the option does have a potential obligation. Thus in case of a call option, the writer is duty bound to sell the stock should the holder choose to exercise the call option. Similarly in case of a put option, the writer is duty bound to purchase the stock should the option is exercised by the holder.

What we have described above are basic characteristics of a *European call/European put option*. Here the option can be exercised at maturity only and not any time before that. Thus for a single period model, maturity has to be $t = 1$. In contrast, if the option can be exercised at any time before or at maturity, then it is called an *American option*. Thus for a single period model, an American option can be exercised at $t = 1$ or at $t = 0$. Unless we mention it specifically, for us an option shall always mean a European option.

Since an option confers on its holder a right without any obligation, the holder needs to pay some amount at $t = 0$ to have this right. This amount which is paid at $t = 0$ when option is bought is called the *premium* or *price* of the option. Thus we need to determine the price of a European call, American call, European put and American put etc.

**Remark 1.4.1** *At first, a call option resembles very much with a long forward contract because both involve buying stock at a future date for a fixed price in advance. But the holder of long forward contract is committed for buying the asset for the fixed price whereas the holder of a call option has the right but no obligation to do so. For this reason there is no premium in case of a long forward contract but there is a premium in the case of an option.*

**Remark 1.4.2** *In the case of derivatives e.g. forward contracts, options etc., there are two parties which get together and set a rule by which one of two parties will receive a payment from the other party depending upon the value of some financial variables. Also profit of one party is the loss of other party. This is a typical situation of a two person zero sum game. Therefore one can visualize many of the derivative pricing and other finance related problems as game theoretic problems and use the rich theory available there.*

**Example 1.4.2** *Let* $B(0) = 100$, $B(1) = 110$, $S(0) = 100$ *and*

$$S(1) = \begin{cases} 120, & \textit{with probability } 0 \cdot 8 \\ 80, & \textit{with probability } 0 \cdot 2. \end{cases}$$

*Consider a European call option* $C$ *on the given stock for which the strike price* $K = 100$ *and expiration date is 1 year. Determine* $C(0)$, *i.e. the call price of this option.*



**Solution** Here $S(1)$, the stock price at $t = 1$, takes two values. Denoting these by $u\ S(0)$ and $d\ S(0)$, we obtain $u = 1.2$ and $d = 0.8$. Such a stock dynamics is called a *single period binomial model.* Here

$$S(1) = \begin{cases} u\ S(0), & \text{with probability } p \\ d\ S(0), & \text{with probability } (1-p)\ , \end{cases}$$

with $0 < p < 1$.

Also as $B(1) = 110$, $B(0) = 100$ we have nominal rate of interest $r = 10\% = 0 \cdot 1$. It is important to note here that $d < (1 + r) < u$ which in the context of binomial model is a requirement for no arbitrage principle to hold. This result is proved analytically in Chapter 3.

Though this dynamics looks extremely simple and probably unrealistic, it forms the basis of other multi period and continuous pricing models.

To determine the call price $C(0)$ we shall make use of the law of one price. For this we shall first determine the pay-off of the given call, i.e its value $C(1)$ at $t = 1$;

and then construct a portfolio $P:(x,y)$ of stock and bond such that $V_P(1)$ equals $C(1)$. Then the law of one price tells that the required price of the call has to be the value of the portfolio at $t=0$ ; i.e. $C(0)=V_P(0)$.

To compute the pay-off $C(1)$ we argue that the holder will exercise the call option only when the stock price $S(1)$ is more than $K$ and receive the amount $S(1)-K$. If $S(1)\le K$, then it is not beneficial for the holder and therefore he/she will use his/her right not to exercise the option. Thus

$$C(1)=\begin{cases} S(1)-K\ , & \text{if } \ S(1)>K \\ 0\ , & \text{if } \ S(1)\le K\ , \end{cases}$$

i.e.

$$C(1)=\max\left((S(1)-K),\ 0\right)=(S(1)-K)^+\ .$$

Here for $x\in R, x^+$ denotes $\max(x,0)$.
For our problem

$$C(1)=\begin{cases} 20, & \text{with probability } \ 0\cdot 8 \\ 0, & \text{with probability } \ 0\cdot 2\ . \end{cases}$$

Next we need to determine x and y such that for the portfolio $P:(x,y)$, $V_P(1)=C(1)$. This gives

$$x\,S(1)+y\,B(1)=C(1)\ ,$$

i.e.

$$\begin{aligned} 120x+110y&=20\\ 80x+110y&=0. \end{aligned}$$

Solving above two equations we get $x=\dfrac{1}{2}$ and $y=-\dfrac{4}{11}$. Hence by the law of one price we obtain

$$\begin{aligned} C(0)=V_P(0)&=x\ S(0)+y\ B(0)\\ &=\frac{1}{2}(100)-\frac{4}{11}(100)\\ &=13\cdot 6364. \end{aligned}$$

Thus the buyer (holder) of the call option has to pay a premium of $Rs$ 13.64 to the seller (writer) of the option. □

Here it must be noted that the above arguments are not valid if the inequality $d < (1 + r) < u$ is not true. This is because if this inequality is not true, then no arbitrage principle does not hold, and hence the law of one price is not valid.

**Remark 1.4.3** *At the time of writing the option $S(1)$ is unknown. Therefore the first problem of option pricing is to find how much the holder should pay the writer for an asset (stock) worth $(S(1)-K)^+$ at time $t = 1$? From the writers point of view it is imperative to use the premium $C(0)$ in constructing a portfolio $(x, y)$ such that it generates an amount $(S(1) - K)^+$ at $t = 1$. This is the problem of hedging the option and the portfolio $(x, y)$ is called the replicating portfolio because in terms of pay-off at $t = 1$ it replicates the given option. Thus while pricing a derivative, two problems are being solved simultaneously namely the pricing problem for the holder and hedging problem for the writer. Here again a game theoretic interpretation could be given so as to interpret these two problems as primal-dual pair.*

**Remark 1.4.4** *On first reading, option may look superfluous in a market because we are replicating in terms of two basic securities of stock and bond. This is certainly true for a single period model. But for multi period and continuous time models, this is not that simple because it involves re-balancing of the replicating portfolio at every time instant. Therefore options and other derivatives are treated as separate and independent financial instruments and not something which can be expressed in terms of stock and bond.*

**Example 1.4.3** *Let $B(0)$, $S(0)$, $B(1)$, $S(1)$, $K$ and $T$ be same as in Example 1.4.2. Let the option $P$ be a put option. Determine $P(0)$.*

**Solution** It is simple to argue that for the case of put option the pay-off at time $t = 1$ is Max $((K - S(1)),\ 0) = (K - S(1))^+$.
Thus

$$P(1) = \begin{cases} 0, & \text{with probability } 0 \cdot 80 \\ 20, & \text{with probability } 0 \cdot 20. \end{cases}$$

This gives

$$\begin{aligned} 120x + 110y &= 0 \\ 80x + 110y &= 20\ , \end{aligned}$$

i.e. $x = -\dfrac{1}{2}$ and $y = \dfrac{6}{11}$. Therefore,

$$P(0) = x\ S(0) + y\ B(0) = -\frac{1}{2}(100) + \frac{6}{11}(100) = 4.54. \qquad \square$$

**Remark 1.4.5** *In above examples we have seen that while determining* $C(0)$ *and* $P(0)$ *we have never used the actual probabilities* $p = 0 \cdot 8$ *and* $(1 - p) = 0 \cdot 2$*. It is probably not possible to explain its reason at this stage because it has to be explained in terms of risk neutral probability measure (RNPM) which will be discussed in Chapter 3 and also in other chapters on option pricing.*

## 1.5 Some More Terminology

We have already introduced terms like bonds, stocks, portfolios, securities and derivatives. In this section we first introduce some more terminology with respect to bond and stock and then proceed to discuss *time value of money* and *money market account.*

Security is a general term used for bonds, stocks and derivatives. Stocks are also referred as *equities*. We can treat bonds as *fixed income securities*, because at maturity they pay fixed amounts of money to their owners. Regular savings accounts, certificates of deposits are other examples of fixed income securities.

Thus a bond is a security that gives its owner the right to a fixed predetermined payment at a future predetermined date, called *maturity*. The amount of money that a bond will pay in future is called *nominal value*, *face value*, *par value* or *principal*. Thus bonds represent the general paradigm of risk free securities. In fact we can treat money (cash) also as bond with zero interest rate and zero (immediate) maturity.



There are bonds that involve only an initial payment (the initial price) and a final payment (the final value or principal). Such bonds are called *pure discount bonds* or *zero coupon bonds*. In contrast there are *coupon bonds* where the holder of the bond gets periodic payments from the issuing institution during the life of the bond. These periodic payments are specified percentage of the principal and are called *coupons.* If the price at which the bond is sold is same as its nominal value we say that the bond sells *at par*. The terms *above par* and *below par* are defined in an obvious manner.

Sometimes the institution issuing the bond does not meet the promise and it defaults in payment. This type of risk is called *credit risk* or *default risk*. The bonds issued by the Government are considered to be free of default risk. Even when there is no default risk, there is always *inflation risk* because future prices of goods are uncertain. If the rate of interest does not take into consideration the inflation risk then it is called *nominal interest rate*. If a bond guarantees a payment that takes care of inflation level, it is called *inflation indexed bond.*

In context of stock, we call the owner of the stock as *stockholder*. The profit that the company distributes to the stockholders is called *dividend*. Dividend in general is not known in advance because it depends upon the company's profit and its policy. The difference between the selling price and the initial price is called *capital gain* or *loss*.

**Time Value of Money**

The way in which money changes its value in time is of utmost importance in finance. Though a general scenario may be very complex, the below given questions are simple to answer.

(i) What is the future value of an amount invested or borrowed today?
(ii) What is the present value of an amount to be paid or received at a certain time in future?

We all have studied simple and compound interest in our schools. Let an initial amount (called *principal*) $P$ be deposited into a bank account where it earns interest. The future value of this investment consists of the principal $P$ together with all the interest earned for the period the money remained deposited. Let the interest rate be $r\%$ per year and $t$ be the time (expressed in years) for which the money was deposited. Let $V(t)$ denote the value of this investment at time $t$. Then under the assumption that the principal $P$ attracts *simple interest rate* only, we have



$$V(t) = P(1 + rt).$$

We next consider the case of *periodic compounding*. Here again an amount $P$ is deposited in a bank account attracting a constant interest rate $r\%$ per year. But in contrast to the simple interest case, here the interest earned is added to the principal periodically. This is called *periodic compounding* which could be annually, semi-annually, quarterly, monthly, weekly or daily basis. Therefore interest will be attracted not only on the principal $P$ but also by all the interest earned so far. In this case we have

$$V(t) = P\left(1 + \frac{r}{m}\right)^{tm},$$

where $m$ refers to the period used in the compounding. Thus if it is compounded semi-annually then $m = 2$ because there are two periods in an year. For quarterly compounding $m = 4$ and for monthly compounding $m = 12$, etc. Here the number $\left(1 + \frac{r}{m}\right)^{tm}$ is called the *growth factor*.

**Example 1.5.1** *How long will it take for a sum of* Rs 800 *attracting simple interest to became* Rs 830 *if the rate of interest is* 9% *per year? Also compute the return on this investment.*

**Solution** From the formula for the simple interest we have

$$830 = 800(1 + 0 \cdot 09\ t)\ ,$$

which gives $t = 0 \cdot 417$ year which is approximately 152 days. Further the return on this investment will be

$$\frac{V(t) - V(0)}{V(0)} = \frac{830 - 800}{800} = 0 \cdot 0375\ ,$$

i.e. $3 \cdot 75\%$.

□

**Example 1.5.2** *Which will deliver a higher future value after one year, a deposit of* Rs 1,000 *attracting interest at* 15% *compounded daily, or at* 15·5% *compounded semi-annually?*

**Solution** At 15% compounded daily the deposit will grow to

$$V(t) = 1000\left(1 + \frac{0 \cdot 15}{365}\right)^{1\times 365} \cong Rs\ 1161 \cdot 80\ ,$$

after one year. If the interest is compounded semi-annually at $15 \cdot 5\%$, the value after one year will be

$$V(t) = 1000\left(1 + \frac{0 \cdot 155}{2}\right)^{1\times 2} \cong Rs\ 1161 \cdot 01\ .$$

So interest at 15% compounded daily will give higher future value.

□

We next introduce the concept of *continuous compounding*. Let us recall the formula

$$V(t) = P\left(1 + \frac{r}{m}\right)^{tm}\ ,$$

which can be written as

$$V(t) = P\left[\left(1 + \frac{r}{m}\right)^{\frac{m}{r}}\right]^{rt}\ .$$

Therefore if we take the limit as $m \to \infty$ we get $V(t) = Pe^{rt}$. This is known as continuous compounding. Here the growth factor is $e^{rt}$.

**Example 1.5.3** *It is given that the future value of Rs* 950 *subject to continuous compounding will be Rs* 1,000 *after half year. Find the interest rate.*

**Solution** The rate $r$ satisfies

$$950\ e^{r\times 0.5} = 1000.$$

This gives $r = 0.1026$, i.e. 10.26%.

□

Let us recall the simple interest formula $V(t) = P(1 + rt)$. Denoting the initial deposit $P$ as $V(0)$ (value of this investment at $t = 0$) we have $V(t) = V(0)(1 + rt)$, which gives $V(0) = V(t)(1+rt)^{-1}$. This last formula gives the *present* or *discounted value* of $V(t)$ and $(1 + rt)^{-1}$ is the *discount factor*.

For the case of periodic compounding the *present* or *discounted value* $V(0)$ of $V(t)$ is given by $V(t)\left(1 + \frac{r}{m}\right)^{-tm}$ and $\left(1 + \frac{r}{m}\right)^{-tm}$ is the discount factor. In case of continuous compounding we have the present or discounted value as $V(0) = V(t)e^{-rt}$ and the discount factor as $e^{-rt}$. If the principal $P$ is invested at time $s$ rather than at time 0, then the modifications are obvious.

**Example 1.5.4** *Find the present value of Rs* 1,00,000 *to be received after* 100 *years if the interest rate is assumed to be* 5% *throughout the whole period and* (*a*) *daily or* (*b*) *annual compounding applies.*

**Solution** Under daily compounding the present value is given by

$$V(0) = 100000\left(1 + \frac{0.05}{365}\right)^{-100\times 365} \cong Rs\ 674\cdot 03.$$

Under annual compounding the present value is given by

$$V(0) = 100000(1 + 0.05)^{-100} \cong Rs\ 760\cdot 45.$$

□

It is obvious that frequent compounding will produce a higher future value than less frequent compounding if the initial deposit $P$ and the interest rate $r$ are same. But for the general case, to compare two compounding methods we need to compare the corresponding growth factor over the same base, say one year. This motivates the below given definition.

**Definition 1.5.1 (Comparison of compounding methods)** *Two compounding methods are said to be equivalent if their corresponding growth factors over a period of one year are same. If one of the growth factor exceeds the other, then the compounding method having the higher growth factor is preferred over the other.*

**Example 1.5.5** *Is semi-annual compounding at* 10% *is preferable to monthly compounding at* 9%*?.*

**Solution** For semi-annual compounding at 10%, the growth factor over a period of one year is

$$\left(1 + \frac{0 \cdot 1}{2}\right)^2 = 1 \cdot 1025.$$

But for monthly compounding at 9%, the growth factor is

$$\left(1 + \frac{0 \cdot 09}{12}\right)^{12} \cong 1 \cdot 0938.$$

Therefore 10% semi annual compounding is preferable to 9% monthly compounding.

□

There is another equivalent convention to compare the two compounding methods. Here rather than comparing the corresponding growth factors, we compare the corresponding *effective rate* $r_e$ as defined below.

**Definition 1.5.2 (Effective Rate)** *For a given compounding method with interest rate* $r$*, the* effective rate $r_e$ *is the one which gives the same growth factor over a one year period under annual compounding.*



In view of Definition 1.5.2, for the case of periodic compounding, the effective rate $r_e$ satisfies the equation

$$\left(1 + \frac{r}{m}\right)^m = (1 + r_e) \ .$$

If it is the case of continuous compounding then $r_e$ satisfies the equation

$$e^r = 1 + r_e \ .$$

In terms of effective rates, two compounding methods are equivalent if the corresponding effective rates $r_e$ and $r_e^{'}$ are equal. if $r_e > r_e^{'}$, then the compounding method with effective rate $r_e$ is preferred over the compounding method with effective rate $r_e^{'}$. This follows directly from Definitions 1.5.1 and 1.5.2.

**Example 1.5.6** *Find the effective rate* $r_e$ *if it is semi-annual compounding at* 10%*.*

**Solution** We have

$$\left(1 + \frac{0.01}{2}\right)^2 = (1 + r_e) ,$$

which gives $r_e = 10 \cdot 25\%$.

□

**Example 1.5.7** *Is daily compounding at* 15% *preferable to semi-annual compounding at* $15 \cdot 5\%$*?*

**Solution** Let the corresponding effective rates be $r_e$ and $r_e'$. Then

$$1 + r_e = \left(1 + \frac{0.15}{365}\right)^{365} \cong 1 \cdot 1618 ,$$

and

$$1 + r_e' = \left(1 + \frac{0.155}{2}\right)^2 \cong 1 \cdot 1610.$$

Thus $r_e \cong 16 \cdot 18\%$ and $r_e' \cong 16 \cdot 10\%$. Therefore daily compounding at 15% is preferable.

□

We can also express future value in terms of effective rate $r_e$ because of its relationship with the growth rate. In fact we can show that



$$V(t) = P\,(1 + r_e)^t$$

for all $t \geq 0$. This formula holds both for the periodic compounding and continuous compounding, but it cannot hold for the case of simple interest. This is because for the simple interest case, $V(t)$ is a linear function of $t$, whereas for the periodic/continuous compounding it is an exponential function of $t$.

**Streams of Payments**

An *annuity* is a sequence of finitely many payments of a fixed amount due at equal time intervals. Let payment of an amount $C$ are to be made once a year for $n$ years, the first one is due a year hence. If we assume that annual compounding applies, then the present value of such a stream of payments is given by

$$\frac{C}{(1+r)} + \frac{C}{(1+r)^2} + \frac{C}{(1+r)^3} + \ldots + \frac{C}{(1+r)^n}.$$

If we denote

$$PA(r,n) = \frac{1}{(1+r)} + \frac{1}{(1+r)^2} + \ldots + \frac{1}{(1+r)^n},$$

then the present value of the given stream of payments is $(C \times PA(r,n))$. Here the number $PA(r,n)$ is called the *present value factor for an annuity*. Obviously

$$PA(r,n) = \frac{1-(1+r)^{-n}}{r}.$$

We define *perpetuity* as an infinite stream of payments of a fixed amount $C$ occurring at the end of each year. Therefore the *perpetuity* can be computed by $\lim_{n\to\infty} PA(r,n) \times C$ which equals $\frac{C}{r}$.

**Example 1.5.8** *Consider a loan of* $Rs\ 1,000$ *to be paid back in* $5$ *equal instalments due at yearly intervals. The instalments include both the interest payable each year calculated at* $15\%$ *of the current outstanding balance and the repayment of the fraction of the loan. Determine the amount of each instalment.*

**Solution** We can look the loan as equivalent to annuity from the point of view of the lender. Therefore if $C$ is the amount of each instalment then

$$1000 = C \times PA(0\cdot 15, 5).$$



But

$$PA(0\cdot 15, 5) = \frac{1-(1+0\cdot 15)^{-5}}{0\cdot 15}.$$

This gives $C \cong Rs\ 298\cdot 32$.

□

**Example 1.5.9** *Let the interest rate be* $18\%$ *and a person can afford to pay* $Rs\ 10,000$ *at the end of each year for next* $10$ *years. Determine the amount which the person can borrow?*

**Solution** The person can borrow the amount $PA(0\cdot 18, 10) \times 10000 = Rs\ 44,941$.

□

### Money Market

By the money market we mean default free securities. Thus money market consists of risk free bonds like treasury bills and notes and other financial securities which promise the holder a sequence of guaranteed future payments. Risk free here means default free i.e. these payments will be delivered with certainty. However in practice the risk, even for risk free securities, cannot be completely avoided since the market prices of such securities may fluctuate unpredictably and there may be default in payments. But in our discussion here we shall assume that it is a default free, and hence risk free situation.

As mentioned earlier, the simplest case of a bond is *zero coupon bond*. Here the issuing institution (a government, a bank, or a company) promises to exchange the bond for a certain amount of money $F$, called the *face value*, on a given day $T$, called the *maturity date*. A typical zero coupon bond has the face value of $Rs$ 100 (or some other round figure) and maturity date as one year (or multiple of a year). If an investor buys a bond, he/she becomes its holder and its effect is that he/she is lending money to the bond writer (a government, a bank, or a company). Here it must be noted that investor need not always be an individual, it could also be a financial institution.

Given the interest rate, the present value of such a bond can be computed easily. Thus if the face value of a bond is $F = Rs$ 100 and maturity date $T = 1$ year, then for 12% compounding annually we have $V(0) = F(1 + r)^{-1} \cong Rs\ 89 \cdot 29$.

For simplicity, we shall assume that the face value $F$ of the given bond is $Rs$ 1. Such bonds are called *unit bonds*. Typically a bond can be sold any time prior to maturity at the prevailing market price. Let $B(t, T)$ denote the price of the given unit bond at time $t$ when the maturity is $T$. Thus $B(T, T) = 1$ and $B(t, T) = e^{-r(T-t)}$. Here we have assumed that $r$ remains constant throughout the period up to maturity.

But in reality the bonds are freely traded in the market and their prices are determined by market forces. Therefore the *market interest rate* (depending upon the market price of the given bond) may be different from the *implied interest rate* $r = \dfrac{F}{V(0)} - 1$ given by the formula $V(0) = F(1 + r)^{-1}$. In fact the implied interest rate should not be constant as it should depend on the trading time $t$ as well as on the maturity time $T$. Not only this, but also the bond prices $B(t, T)$ are determined by (random) market forces. Therefore the implied interest rate should be taken as a random process $r(t, T)$. We shall have occasion to discuss this aspect in a later chapter.

As our bond is a unit bond we can treat $B(0, T)$ as the discount factor and $B(0, T)^{-1}$ as the growth factor. This observation allows to compute time value of money without restoring to the corresponding interest rates.

### Coupon Bonds

In zero coupon bonds, there is just a single payment at maturity. But in *coupon* *bonds*, there is a sequence of payments. These payments consist of the face value due at maturity, and *coupons* paid regularly, say annually, semi-annually or quarterly. The last coupon becomes due at maturity only. If the interest rate $r$ is assumed to be constant throughout, then it is simple to compute the price of a coupon bond by discounting all the future payments. We explain this process with the help of the next example.

**Example 1.5.10** *Find the price of a bond with face value* $Rs\ 100$ *and* $Rs\ 5$ *annual coupons that matures in* 4 *years, if the continuous compounding rate is* 8%.

**Solution** The price of the given bond is obtained as

$$5e^{-0.08} + 5e^{-2\times 0.08} + 5e^{-3\times 0.08} + 105e^{-4\times 0.08} \cong Rs\ 89 \cdot 06.$$

Here we should note that as maturity is 4 years, the payments have to be $Rs\ 5, 5, 5$ and105.

□

**Money Market Account**

An investment in the money market can be realized by means of an investment bank which buys and sells bonds on behalf of its customers. The risk free position of an investor is given by the level of his/her account with the bank. We can think of this account as a tradable asset, since the bond themselves are tradable. A long position in the money market involves investing the money and a short position means borrowing money.

Thus an investment of amount $A(0)$ in the money market amounts to purchasing $\dfrac{A(0)}{B(0,T)}$ bonds at $t = 0$. At time $t$, the value of each bond will be

$$B(t,T) = e^{-r(T-t)} = e^{rt}e^{-rT} = e^{rt}B(0,T).$$

As a result, at time $t,\ \ t \leq T$, the investment $A(0)$ will become

$$A(t) = \frac{A(0)}{B(0,T)}B(t,T) = A(0)e^{rt}.$$

The investment in bond has a finite time horizon, say $T$. To extend the position in the money market beyond $T$, the investor can invest the amount received at maturity into a newly issued bond at time $T$, but maturing at $T'(T' > T)$. We explain this by the below given example.

**Example 1.5.11** *Suppose Rs* 1 *is invested in zero coupon bonds maturing after one year. At the end of each year the proceeds are re-invested in new bonds of the same kind. How many bonds will be purchased at the end of year* $n$ $(n \geq 1)$*? It may be assumed that the interest rate is constant at* $r$ % *and continuous compounding is followed.*

**Solution** At $t = 0$, the investor buys $\dfrac{1}{B(0,1)} = e^r$ bonds. At time $t = 1$, increases his/her holdings to $\dfrac{e^r}{B(1,2)} = e^{2r}$ bonds. Therefore at the end of year $n$, the investor will purchase $e^{(n+1)r}$, one year bonds.

□

## 1.6 Summary and Additional Notes

- The date of March 29, 1900 should be considered as the origin of Financial Mathematics. On that day, a French postgraduate student, Louis Bachelier successfully defended his thesis *Théorie de la spéculation*. Bachelier presented a model of Brownian motion (see Chapter7) while deriving the evolution of Paris asset prices. This work of Bachelier was neglected by the economic community for many years but probabilists took great interest in this development. Kolmogorov's analytical theory of Markov process has its origin in the theory proposed by Bachelier. The economic community also later recognized Bachelier's contribution and named *International Finance Society* after him.
  Bachelier's thesis topic at that time was out of ordinary. The appropriate topic of research in that era were on the theory of functions. Therefore this topic was not initially accepted as a research topic. For these and other historical details on Bachelier's work we may refer to Vassiliou [144].
- In this chapter we have presented an overview of the topic of *financial mathematics*. Here our main aim has been to introduce the basic terminology of finance and motivate the readers to the art of mathematical modeling to deal with the problems arising in this area. In the process, we have also introduced the readers to the three core topics of the book, namely *option pricing, portfolio optimization* and *interest rate theory* through elementary illustrative examples.
- Though the illustrative examples presented here require only an elementary knowledge of algebra, calculus and probability theory, it has been emphasized that real life problems in finance are extremely complex. Therefore we need to get ourselves equipped with certain advanced topics in mathematics, e.g.

stochastic process and stochastic calculus, numerical optimization, and partial differential equations etc.

- The importance of *no arbitrage principle* and the *law of one price* is illustrated repeatedly in various illustrative examples.
- It has been remarked that the standard deviation is not always a good measure of risk for studying problems in the finance area and therefore the definition of *risk* becomes significant.
- The relevant texts for this chapter are Capinski and Zastawniak [25], Hull [65] and Luenberger [85].

## 1.7 Exercises

**Exercise 1.1** *Let* $B(0) = Rs\ 100,\ B(1) = Rs\ 110$ *and* $S(0) = Rs\ 80.$ *Also let*

$$S(1) = \begin{cases} 100, & \text{with probability } p = 0.80 \\ 60, & \text{with probability } p = 0.20. \end{cases}$$

*Design a portfolio with initial wealth of* $Rs\ 100,000,$ *split fifty-fifty between stock and bond. Compute the expected return and the risk of the portfolio so constructed.*

**Exercise 1.2** *Let* $B(0),\ B(1)\ S(0)$ *and* $S(1)$ *be as specified in Exercise 1.1. Also let* $C$ *and* $P$ *respectively be a European call and European put with* $K = Rs\ 100$ *and* $T = 1$ *year.*

*(i) Determine* $C(0)$ *and* $P(0)$.

*(ii) Find the final wealth of an investment with initial capital of* $Rs\ 900$ *being invested equally in the given stock, the given call and the given put.*

**Exercise 1.3** *In the data of Exercise 1.2, let there be a minor change. Instead of* $B(1) = Rs\ 110$, *let* $B(1) = Rs\ 120$. *Find the replicating portfolio* $(x, y)$ *for the call* $C$. *Are we justified in taking* $C(0) = x\ S(0) + y\ B(0)$? *Give appropriate mathematical reason for your answer. Will your answer change if instead of call* $C$ *we take the put* $P$ *?*

**Exercise 1.4** *Spot an arbitrage opportunity (if it exists) in the following situation. Suppose that a dealer A offers to buy British pounds in an year from now at a rate of* $Rs\ 79$ *a pound, while dealer B would sell British pounds immediately at a rate of* $Rs\ 80$ *a pound. Assume that a rupee can be borrowed at an annual rate of* $4\%$ *and a British pound can be invested in a bank at* $6\%$ *annual interest.*

**Exercise 1.5** *Let* $B(0) = Rs\ 100$, $B(1) = Rs\ 112$, $S(0) = Rs\ 34$ *and* $T = 1$. *Find the forward price* $F$. *Also find an arbitrage opportunity if* $F$ *is taken to be* Rs 38.60.

**Exercise 1.6** *Assume that a person can afford to pay Rs* 10,000 *at the end of each year. How much the person can borrow if the interest rate is* 12% *and he/she wishes to clear the loan in 5 years?*

**Exercise 1.7** *What will be the difference between the value after 1 year of* Rs 10,000 *deposited at* 10% *compounded monthly and compounded continuously. How frequently should the periodic compounding be done for the difference to be less than* Rs 1*?*

**Exercise 1.8** *Consider a loan of* Rs 10,000 *to be paid back in 10 equal instalments at yearly intervals. The instalments include both the interest payable each year calculated at* 15% *of the current outstanding balance and repayment of a fraction of the loan. Determine the amount of each instalment. What is the amount of interest included in each instalment and what is the outstanding balance of the loan after each instalment is paid?*

**Exercise 1.9** *Find the return on a 75-day investment in zero-coupon bonds if* $B(0,1) = Rs\ 0.89$.

**Exercise 1.10** *After how many days will a bond purchased for* $B(0,1) = Rs\ 0.92$ *produces a* 5% *return?*

# 2
# Forward and Futures Contracts

## 2.1 Introduction

A *forward contract* is probably the simplest of all derivative securities. It has a simple pricing mechanism and has wide applications, particularly in commodity and foreign exchange markets. We had already initiated some discussion on forward contracts in Chapter 1 where we had also explained the pricing methodology through a simple example.

Though *futures contracts* are very much in the spirit of forward contracts, they have specially been designed to standardize these contracts so as to eliminate the risk of default by the party suffering the loss. For this, a process called *marking to market* is conceptualized which requires an individual to open a *margin account* which is managed by an organized clearing house/exchange.

This chapter continues our earlier discussion on forward contracts and explains various concepts related with the working of futures contracts. We also describe very briefly another derivative security called *swap* which is again very popular in commodity and foreign exchange markets.

## 2.2 Forward Contract

Let us recall the definition of a *forward contract* from Chapter 1 and note that it is an agreement between two investors (also called parties) to buy or sell a risky asset at a specified future time (called the *delivery date*) for a price F (called the *forward price*) fixed at the present time (say $t = 0$). The party which agrees to buy the asset is said to enter into a *long forward contract* or to take a *long forward position.* The party which agrees to sell the asset is said to enter into a *short forward contract* or to take a *short forward position.* There is no payment of money by either party when the agreement for the forward contract is made. Some typical examples of forward contract could be a farmer wishing to fix the sale price of his/her crops in advance, an importer arranging to buy foreign currency at a fixed rate in future, a fund manager wishing to sell stock for a fixed price in future or a country wishing to import commodities (like wheat, sugar, oil etc.) from another country at a fixed rate at some specified future date. In these situations, forward contracts become very handy because they provide an opportunity to hedge against the unknown future price of the underlying risky asset.

Let $t = 0$ denote the time when the two parties enter into the specific forward contract agreement and $t = T$ be the delivery date. Let the agreed forward price be $F(0, T)$. Here the two arguments in $F$ denote the time $t = 0$ and $t = T$ respectively. In case the context is clear and there is no ambiguity, we write $F(0, T)$ simply as $F$.

Let us now analyze the two scenarios, namely $F(0, T) < S(T)$ and $F(0, T) > S(T)$. Let us first consider the case $F(0, T) < S(T)$. In this case the party having the long forward contract will benefit because it can buy the asset for $F(0, T)$ and sell the same for the market price $S(T)$, making a profit of $S(T) - F(0, T)$. But the other party which has taken a short forward position will suffer loss of $S(T) - F(0, T)$ because it will have to sell below the market price. Fig. 2.1 clearly exhibits these two positions.



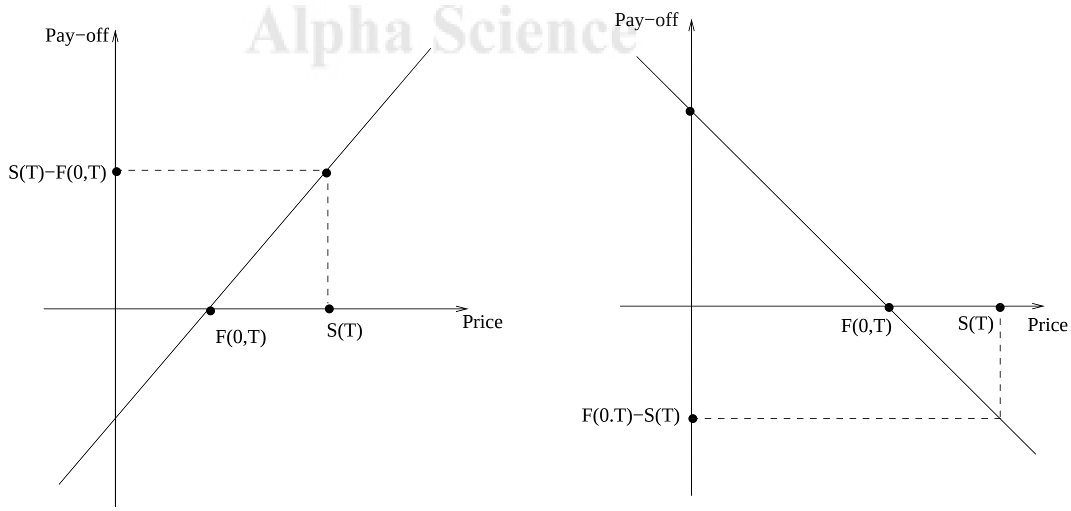



**Fig. 2.1.** Pay-off of a forward contract

If $F(0, T) > S(T)$ then these arguments will simply get reversed. In this case the party taking a short forward position will benefit and make a profit of $F(0,T)-S(T)$ whereas the other party (which has taken a long forward position) will make a loss of $F(0,T)-S(T)$ and analogue of Fig. 2.1 can be drawn for this case as well.

Therefore in either case the pay-off of a long forward contract at delivery is $S(T)-F(0,T)$ and that of a short forward contract is $F(0,T)-S(T)$. Sometimes the contract may be initiated at time $t<T$ rather than $t=0$. In this scenario the pay-offs of a long forward position and a short forward position are $S(T)-F(t,T)$ and $F(t,T)-S(T)$ respectively where $F(t,T)$ denotes the forward price.

## 2.3 Forward Price Formula

We now look into the problem of determining the price of the forward contract, i.e. the forward price $F(0,T)$. We shall make use of the *no arbitrage principle* and follow the methodology indicated in Chapter 1 to get the next theorem.

**Theorem 2.3.1 (Forward Price Formula)** *Let the price of an asset at $t=0$ be $S(0)$. Then the forward price $F(0,T)$ is given by*

$$F(0,T)=\frac{S(0)}{d(0,T)}\ , \tag{2.1}$$

*where $d(0,T)$ is the discount factor between $t=0$ and $t=T$.*



**Proof.** If possible let $F(0,T)>(S(0)/d(0,T))$. Then at $t=0$ we construct the portfolio P as follows

(i) borrow $S(0)$ amount of cash until $t=T$,
(ii) buy one unit of the underlying asset on the spot market at price $S(0)$ (if the underlying is the stock then it will mean buying one share of stock for amount $S(0)$),
(iii) take a short forward position with delivery at $t=T$ and forward price as $F(0,T)$.

This gives the value of the portfolio P at $t=0$ as

$$V_P(0)=-S(0)+S(0)=0.$$

Now at $t=T$, we sell the asset for $F(0,T)$ and pay the amount $(S(0)/d(0,T))$ to clear the loan with interest and thereby close the short position. This gives

$$V_P(1)=F(0,T)-(S(0)/d(0,T))\ ,$$

which is strictly positive under the stated assumption. Thus there is a strictly positive amount of risk free profit with zero net investment. This is contrary to the no arbitrage principle.

Next suppose that $F(0,T) < (S(0)/d(0,T))$. In this case, at $t = 0$, we construct the portfolio Q as follows

(i) short sell one unit of the underlying asset (thus if the underlying asset is stock, then we borrow one share of stock and sell the same in the market with all conditions and procedure as specified in short selling),
(ii) invest the proceeds (i.e. the amount $S(0)$ received by selling one unit of the asset) at the risk free rate up to $t = T$,
(iii) enter into a long forward contract with delivery as $t = T$ and forward price $F(0,T)$.

This gives

$$V_Q(0) = S(0) - S(0) = 0.$$

Now at $t = 1$, we

(i) cash the risk free investment with interest, i.e. collect cash $(S(0)/d(0,T))$,
(ii) buy the underlying asset for $F(0,T)$ using the long forward position,
(iii) close the short position on the underlying asset by returning the borrowed unit of asset to the owner.

This gives

$$V_Q(T) = (S(0)/d(0,T)) - F(0,T),$$

which is strictly positive. This again allows a strictly positive risk free profit with zero net investment contradicting the no arbitrage principle. Thus $F(0,T) = S(0)/d(0,T)$.

□

**Remark 2.3.1** *For the case of constant interest rate $r$ being compounded continuously, we have $d(0,T) = e^{-rT}$ and hence*

$$F(0,T) = S(0)\ e^{rT}. \tag{2.2}$$

**Remark 2.3.2** *If the contract is initiated at some intermediate time $t$, $0 < t < T$, then $d(t,T) = e^{-r(T-t)}$ and*

$$F(t,T) = S(t)\ e^{r(T-t)}, \tag{2.3}$$

*where $r$ is same as in Remark 2.3.1.*

**Remark 2.3.3** *In terms of zero coupon bond prices, the forward price formula 2.1 becomes*

$$F(0,T) = \frac{S(0)}{B(0,T)}\ , \tag{2.4}$$

*which is more convenient to use as it does not require the assumption that the interest rate r is constant.*

It is a matter of common knowledge that holding of physical assets like gold, sugar, oil, wheat etc. entail inventory carrying costs, such as rental for storage and insurance fees etc. These costs affect the theoretical forward price formula as given at (2.1). The standard forward price formula (2.1) can be generalized in several ways to take into consideration the costs of carry and also to include dividends. We have the below given result in this regard.

**Theorem 2.3.2 (Forward Price Formula with Carrying Costs)** *Let an asset carry a holding cost of c(i) per unit in period* $i$ $(i = 0, 1, 2, \ldots (n-1))$. *Also let at* $t = 0$, *the price of this asset be* $S(0)$ *and short selling be allowed. Then*

$$F(0,T) = \frac{S(0)}{d(0,T)} + \sum_{i=0}^{n-1} \frac{c(i)}{d(i,n)}\ , \tag{2.5}$$

*where delivery date is* $t = T$ *and between* $t = 0$ *and* $t = T$ *there are n periods which have appropriately been identified as per the given context.*

Next we consider the case when the underlying asset (e.g. stock) is dividend paying.

**Theorem 2.3.3 (Forward Price Formula with Dividend)** *Let an asset be stored at zero cost and also sold short. Let the price of this asset at* $t = 0$ *be* $S(0)$ *and a dividend of* $Rs\ div$ *be paid at time* $\tau$, $0 < \tau < T$. *Then*

$$F(0,T) = \frac{[S(0) - (div)\,(d(0,\tau))]}{d(0,T)}\ . \tag{2.6}$$

*For the case of constant interest rate r being compounded continuously, the above formula becomes*

$$F(0,T) = \left[S(0) - (div)\ e^{-r\tau}\right] e^{rT}\ . \tag{2.7}$$

Sometimes the asset (stock) may pay dividends continuously at a rate of $r_{div} > 0$. This is called *continuous dividend yield*. If the dividends are reinvested in the stock, then an investment in one share held at $t = 0$ will become $e^{(r_{div})T}$ shares at $t = T$. This is very similar to continuous compounding and therefore $e^{-(r_{div})T}$ shares at $t = 0$ will give one share at $t = T$. This observation leads to Theorem 2.3.4.

**Theorem 2.3.4 (Forward Price Formula with Continuous Dividend )** *Let an asset be stored at zero cost and also sold short. Let the price of the asset at* $t = 0$ *be* $S(0)$ *and the asset pay dividends continuously at a rate of* $r_{div}$*. Then*

$$F(0, T) = S(0)\ e^{(r - r_{div})T}. \tag{2.8}$$

Theorems 2.3.2, 2.3.3 and 2.3.4 are not proved here as their proofs can be constructed analogous the proof of Theorem 2.3.1. Interested readers may refer to Luenberger [85] and Capinski and Zastawniak [25] in this regard.

Forward contracts in foreign currency market are very common and we wish to derive a formula for determining the price of the same. To be specific, let us consider the two currencies as British pound and US dollars with the latter as the underlying. The readers may imagine a British importer of US goods requiring US dollars after $t = T$. So this British importer may think of taking a forward contract on US dollars with delivery as $t = T$ to hedge against the fluctuating exchange rate of British pound versus US dollars.

Let at $t = \tau$, the buying and selling exchange rate be: 1 British pound = $P(t)$ US dollars. Also let the risk free interest rates for investments in British pounds and US dollars be $r_{GBP}$ and $r_{USD}$ respectively. Then the forward price is given by

$$F(0, T) = P(0)\ e^{(r_{USD} - r_{GBP})T}. \tag{2.9}$$

The above formula gives the agreed exchange rate at $t = T$, i.e. at $t = T$, the British importer will be able to buy $F(0, T)$ US dollars for one British pound.

To prove formula (2.9), let us consider two strategies as described below.

**Strategy A**

Invest $P(0)$ US dollars at the rate of $r_{USD}$ until $t = T$.

**Strategy B**

Buy one British pound for $P(0)$ US dollars, invest it until $t = T$ at the rate of $r_{GBP}$, and take a short position in $\exp(r_{GBP} \cdot T)$ British pound forward contract with delivery as $t = T$ and forward price as $F(0, T)$. This gives

$$P(0)\ e^{(r_{USD})T} = F(0, T)\ e^{(r_{GBP})T},$$

and hence the result.

**Example 2.3.1** *The current price of gold is* Rs 18,000 *per* 10$gm$*. Assuming a constant interest rate of* 8% *per year compounded continuously, find the theoretical price of gold for delivery after 9 months.*

**Solution** Here $S(0) = 18,000$, $r = 8\%$ per year and $T = 9$ months $= 3/4$ year. Hence by the forward price formula

$$\begin{aligned} F &= S(0)\ e^{rT} \\ &= (18000)\ e^{(0.08\times 3/4)} \\ &= (18000)\ e^{0.06} \\ &= Rs\ 19113.06. \end{aligned}$$

□

**Example 2.3.2** *Find the forward price of a non-dividend paying stock traded today at* $Rs$ 100*, with the continuously compounded interest rate of* 8% *per year, for a contract expiring seven months from today.*

**Solution** We have $S(0) = 100$, $r = 8\%$ per year and $T = 7/12$ year. This gives

$$F = 100\ e^{(0.08)(7/12)} = Rs\ 104.78.$$

□

**Example 2.3.3** *The current price of sugar is* $Rs$ 60 *per Kg and its carrying cost is* 10 *paisa per Kg per month to be paid at the beginning of each month. Let the constant interest rate* $r$ *be* 9% *per annum. Find the forward price of the sugar (Rupees per Kg) to be delivered in* 5 *months.*

**Solution** The interest rate is $(.09)/12 = .0075$ per month. Therefore the reciprocal of one month discount rate for any month is (1.0075). Therefore we have

$$\begin{aligned} F(0,5) &= (1.0075)^5(60) + \Big(\sum_{i=1}^{5}(1.0075)^i\Big)(.1) \\ &= Rs\ 62.79. \end{aligned}$$

□

**Example 2.3.4** *Let the price of a stock on* 1*st April* 2010 *be* 10% *lower than it was on* 1*st January* 2010*. Let the risk free rate be constant at* $r = 6\%$*. Find the percentage drop of the forward price on* 1*st April* 2010 *as compared to the one on* 1*st January* 2010 *for a forward contract with delivery on* 1*st October* 2010.

**Solution** It is convenient to take 1*st* January 2010 as $t = 0$. Then 1*st* October 2010 is 9 months, i.e. 3/4 year. Thus $t = 3/4$. Also $t = \tau$ is 1*st* April 2010, i.e. $3/12 = 1/4$. Therefore using Theorem 2.3.1 we get

$$F(0, 3/4) = S(0)\ e^{(0.06\times 3/4)}$$
$$F(1/4, 3/4) = S(0)(0.9)\ e^{(0.06\times 2/4)}.$$

Here it may be noted that on $1st$ April 2010 the price of the stock is 10% lower than that on $1st$ January 2010, i.e. $(0.9)S(0)$. Now

$$\frac{F(0,3/4) - F(1/4,3/4)}{F(0,3/4)} = \frac{e^{(0.06\times 3/4)} - (0.9)\ e^{(0.06\times 1/2)}}{e^{(0.06\times 3/4)}}$$
$$= 0.1134.$$

Therefore the percentage drop in the forward price is 11.34%.

□

**Example 2.3.5** *An Indian importer wants to arrange a forward contract to buy US dollars in half a year. The interest rates for investment in Indian rupees and US dollars are $r_{INR} = 6\%$ and $r_{USD} = 4\%$ respectively. Also the current rate of exchange is Rs 46 to a dollar, i.e. $\frac{1}{46}$ dollars to a rupee. Find the forward price of dollar-rupee exchange rate.*



**Solution** Here US dollars play the role of underlying. Therefore from (2.9) we get

$$F(0, 1/2) = (1/46)\ e^{(0.5\times(0.06-0.04))}$$
$$= (1/46)\ e^{(0.5\times 0.02)}$$
$$= 0.022,$$

dollars to a rupee, i.e. $\frac{1}{(0 \cdot 022)} = 45 \cdot 54$ rupees to a dollar.

□

## 2.4 The Value of a Forward Contract

Let a forward contract be initiated at $t = 0$ with delivery $t = T$. Also let $F(0, T)$ be the forward price of this contract. Consider an intermediate time $0 < \tau < T$ and let $F(\tau, T)$ be the forward price of the contract initiated at $t = \tau$ with delivery $t = T$.

Thus we have two forward contracts; one initiated at $t = 0$ and other initiated at $t = \tau$, both having the same delivery date, namely $t = T$. Then $F(0, T)$ and $F(\tau, T)$ are the respectively the forward prices of these two contracts. Now as time progress, *the value of the forward contract* initiated at $t = 0$ will be changing. Let at $t = \tau$, its value be $f(\tau)$. Is there a relationship connecting $f(\tau)$, $F(0, T)$ and $F(\tau, T)$? The below given theorem is precisely the answer to this question.

**Theorem 2.4.1 (Value of a Forward Contract)** *Let $f(\tau)$, $F(0, T)$ and $F(\tau, T)$ be as explained above. Then*

$$f(\tau) = (F(\tau, T) - F(0, T))\ d(\tau, T), \tag{2.10}$$

*where $d(\tau, T)$ is the risk free discount factor over the period $t = \tau$ to $t = T$.*

**Proof.** If possible let

$$f(\tau) < (F(\tau, T) - F(0, T))\ d(\tau, T).$$

Now at $t = \tau$ we construct a portfolio P as follows

(i) borrow the amount $f(\tau)$ to enter into a long forward contract with forward price $F(0, T)$ and delivery $t = T$,
(ii) enter into a short forward position with forward price $F(\tau, T)$ for which there is no cost as per the definition of forward contract.

Then the value of the portfolio P at $t = \tau$ is given by

$$V_P(\tau) = 0.$$



Next at time $t = T$, we

(i) close the forward contracts by collecting (or paying, depending upon the sign of pay-offs) the amounts $S(T) - F(0, T)$ for the long forward position and $-S(T) + F(\tau, T)$ for the short forward position,
(ii) pay back the loan amount with interest, i.e. amount $(f(\tau)/d(\tau, T))$.

Therefore the value of the portfolio at $t = T$ is

$$V_P(T) = F(\tau, T) - F(0, T) - (f(\tau)/d(\tau, T)),$$

which is strictly positive and risk free. This violates no arbitrage principle.

Let us now consider the second case, namely $f(\tau) > [F(\tau, T) - F(0, T)]\, d(\tau, T)$. In this case our strategy is to construct a portfolio Q at $t = \tau$ as follows

(i) sell the forward contract which was initiated at $t = 0$ for the amount $f(\tau)$,
(ii) invest this amount $f(\tau)$ risk free from $t = \tau$ to $t = T$,

(iii) enter into a long forward contract with delivery time $t = T$ and forward price as $F(\tau, T)$.

Then $V_Q(\tau) = 0$ and $V_Q(T)$ is given by

$$\begin{aligned} V_Q(T) &= \frac{f(\tau)}{d(\tau, T)} + (S(T) - F(\tau, T)) + (F(0, T) - S(T)) \\ &= \frac{f(\tau)}{d(\tau, T)} + F(0, T) - F(\tau, T)\ , \end{aligned}$$

which is strictly positive and risk free. This is not possible due to no arbitrage principle.

□

**Example 2.4.1** *Let at the beginning of the year, a stock be sold for* Rs 45 *and risk free interest rate be* 6%. *Consider a forward contract on this stock with delivery date as one year. Find its forward price. Also find its value after 9 months if it is given that the stock price at that time turns out to be* Rs 49.

**Solution** From (2.2), the initial forward price $F(0, 1)$ is given by

$$F(0, 1) = S(0)\ e^{rT} = 45\ e^{0.06} = Rs\ 47.78.$$

Also it is given that $S(9/12) = 49$ and hence

$$F(9/12, 1) = 49\ \ \exp\left((\cdot 06)(3/12)\right) = Rs\ 49 \cdot 74.$$

Therefore by Theorem 2.4.1, the value of the forward contract after 9 months is

$$\begin{aligned} f(9/12) &= [F(9/12, 1) - F(0, 1)]\ \ \exp((-0 \cdot 06)(1 - (9/12))) \\ &= 1 \cdot 93. \end{aligned}$$

□

**Example 2.4.2** *Consider the data of Example 2.4.1 and assume that a dividend of* Rs 2 *is being paid after 6 months. Find the forward price of the contract and also its value after 9 months.*

**Solution** By (2.7), the initial forward price $F(0, 1)$ is given by

$$\begin{aligned} F(0, 1) &= [S(0) - (div)e^{-rt}]\ e^{rT} \\ &= \left[45 - 2e^{-0.06(1/2)}\right]\ e^{\cdot 06} \\ &= Rs\ 45 \cdot 72. \end{aligned}$$

Also by (2.3)

$$F(9/12, 1) = S(9/12)\ e^{(0\cdot 06)(1-9/12)} = Rs\ 49 \cdot 74.$$

Hence by Theorem 2.4.1,

$$f(9/12) = Rs\ 3 \cdot 96.$$

□

## 2.5 Futures Contract

It is obvious that when two parties enter into a forward contract, one of them is certainly going to lose money. Therefore a forward contract is always exposed to a risk of default by the party suffering a loss. To eliminate this risk of default, futures contracts have been introduced which are managed by an organized exchange. Thus even though futures contracts are very much in the spirit of forward contracts, their working is entirely different. Here the role of an organized exchange becomes very important because individual contracts are made with the exchange, which itself becomes counter party to both long and short trades. Therefore individuals themselves do not need to search for an appropriate counterparty because this job is taken care off by the exchange itself. The exchange also takes appropriate measures so as to eliminate the risk of counterparty default. This is done by a process called *marking to market*. Before we explain the working of the process, we list the basic features of a futures contract.



(i) Similar to a forward contact, a futures contract also has an underlying asset (e.g. stock) and a delivery date, say time $t = T$.

(ii) In addition to the (underlying) asset prices, the market also dictates the futures price $f(t, T)$ at discrete time steps, say $f(n, T)$ for $n = 0, 1, 2, \ldots$ with $nt \leq T$. In practice these discrete time steps may refer to day1, day2 etc.

(iii) $f(0, T)$ is known but $f(n, T)$ for $n = 1, 2 \ldots$ with $nt \leq T$ are unknown and are treated as random variables. Also the (underlying) asset prices $S(n)$ for $n = 1, 2 \ldots$ with $nt \leq T$ are random variables.

(iv) As in the case of a forward contract, there is no cost involved in initiating a futures contract, but there is a major difference in the cash flow. A long forward contract involves just a single payment $S(T) - F(0, T)$ at the delivery time $t = T$, whereas a futures contract involves a random cash flow at each time step $n = 0, 1, 2 \ldots$ with $nt \leq T$. Thus the holder of a long futures position will receive the amount $f(n, T) - f(n-1, T)$ if positive, or will have to pay if it is negative. For a short futures position, exactly opposite payments apply. These payments are managed by a clearing house for the futures market.

(v) The futures price satisfy the condition that $f(t,T) = S(T)$. This condition has to hold because the futures cost of immediate delivery of goods has to be the *market price* or *spot price*.

(vi) It costs nothing to close, open or alter a futures position at any time step between $t = 0$ and $t = T$. This condition can be met if at each time step $n = 0,1,2\ldots$ with $nt \le T$, the value of futures position is zero. For $n \ge 1$, this value is computed after marking to market.

But what is the physical meaning of futures price $f(n,T)$? Suppose a forward contract initiated at time $t = 0$ has forward price $F(0,T)$, where time $t = T$ is the delivery date. Let the forward price for a new contract initiated at time $t = 1$ with delivery date time $t = T$ be $F(1,T)$. Now the clearing house comes into picture. At the second day it revises all earlier contracts to the new delivery price $F(1,T)$ and accordingly an investor which holds a long forward contract initiated at $t = 0$ receives or pays the difference of two prices depending upon if the change in price reflects a loss or gain. So if $F(1,T) > F(0,T)$, then an investor holding a long forward contract receives $F(1,T) - F(0,T)$ from the clearing house because at the delivery he/she has to pay $F(1,T)$ rather than $F(0,T)$. Continuing in this manner and assuming that the investor stays until maturity, the profit/loss pay-off of a long futures position will be $[F(1,T) - F(0,T)] + [F(2,T) - F(1,T)] + \ldots + [F(T,T) - F(T-1,T)]$, which equals $S(T) - F(0,T)$ because $F(T,T) = S(T)$.

The above discussion demonstrates that the pay-off of a futures contract is the pay-off of the corresponding forward contract, except that the pay-off is paid throughout the life of the contract rather than at maturity. However there is no requirement that the investor of a futures contract has to stay till maturity. He/she can come out of the contract any time by taking the opposite position in the contract with the same maturity. Thus $f(n,t)$, $n = 0,1,2,\ldots$ with $nt \le T$ are essentially the forward prices as perceived by the market.

Here it may be noted that $f(n,T)$ are not known and are taken as random variables dictated by the market. This is because interest rate r is rarely constant and, in general, it is stochastic in nature.

This process of adjusting the contract is called *marking to market*. Here an individual is required to open a *margin account* with the clearing house. This account must have a specified amount of cash for each futures contract. In practice it is about 5 to 10% of the value of futures contract. The margin account is compulsory for all contract holders whether they have long or short position. If the price of futures contract increased that day, then the parties having long position receive an amount which equals (change in price)$\times$ (the contract quantity) which is deposited in their account. The short parties loose the same amount and therefore this amount is deducted from their account. This process is termed as *marking of accounts to the market* which is carried out at the end of each trading day. This guarantees that both parties of the contract cover their obligations.

Thus each margin account value fluctuates from day to day according to change in futures price. At the delivery date, delivery is made at the futures contract price at that time which may be different from the futures price when the contract was first initiated.

Margin accounts serve a dual role. They serve as accounts to collect or pay out daily profits and also guarantee that contract holders do not default on their obligations. If the value of a margin account drops below a pre-defined margin level (in practice about 75% of the initial margin requirement), a *margin call* is issued to the contract holder demanding additional margin. Otherwise futures position will be closed by taking an equal and opposite position. Also any excess amount above the initial margin can be withdrawn by the investor. This margin account is totally managed by the futures clearing house. There are other rules/procedures/practices for managing the market account but these details are not presented here.

**Example 2.5.1** *Suppose that the initial margin is set at* 10% *and the maintenance margin at* 5% *of the futures price. Suppose for* $n = 0, 1, 2, 3, 4$, *the futures prices are* 140, 138, 130, 140 *and* 150 *respectively. Show the working of marking to market and margin account in a tabular form.*

**Solution** In the below given table, the two columns termed as Margin 1 and Margin 2, respectively refer to the deposits at the beginning and end of the day.

| $n$ | f($n$,T) | Cash Flow | Margin 1 | Payment | Margin 2 |
|---|---|---|---|---|---|
| 0 | 140 | opening | 0 | -14 | 14 |
| 1 | 138 | -2 | 12 | 0 | 12 |
| 2 | 130 | -8 | 4 | -9 | 13 |
| 3 | 140 | 10 | 23 | 9 | 14 |
| 4 | 150 | 10 | 24 | 9 | 15 |
| | | closing | 15 | 15 | 0 |
| | | | Total | 10 | |

Here on day 0, a futures position is opened and 10% deposit (i.e. *Rs* 14) paid to open the market account. On day 1, the futures price drops by *Rs* 2 which is subtracted from the deposit. On day 2, there is further drop of *Rs* 8 in the futures price which makes the deposit below 5%. Therefore there is a margin call and the investor has to pay *Rs* 9 to restore the deposit to 10% level. On day 3, the futures price increases and $Rs$ 9 is withdrawn leaving a 10% margin. On day 4, the futures price again goes up, which allows the investor to further withdraw $Rs$ 9. At the end of day, the investor decides to close the position collecting the balance of the deposit. The total of all payments is $Rs$ 10 which is the same as the increase in future price from day 0 to day 4.

□

## 2.6 Futures Pricing

We now wish to establish that if the interest rate $r$ is constant and it is being compounded continuously then $f(0,T) = F(0,T)$. Obviously there is nothing very special here about $t = 0$. We can take any t, $0 < t < T$ and obtain

$$f(t,T) = F(t,T) = S(t)\, e^{r(T-t)}\ , \tag{2.11}$$

for an asset which is non dividend paying. For a dividend paying stock, the above formula (2.11) could be modified in an obvious manner.

Earlier we have noted that the futures prices are random and dictated by the market perception. What is important to note here is that if the market futures prices depart significantly from the values given by formula (2.11), it indicates that the market does not believe that the interest rate is constant. The difference between the two values signifies the market's perception of future interest rate changes.

**Theorem 2.6.1** *Let the interest rate r be constant and compounded continuously. Then*

$$f(0,T) = F(0,T).$$

**Proof.** For the sake of simplicity, let us assume that in the entire life of futures contract (i.e. from $t = 0$ to $t = T$), the marking to market is carried out at two time steps only. Let these be denoted by $t = t_1$ and $t = t_2$ with $0 < t_1 < t_2 < T$. The general case of $n$ time steps can be dealt with by making suitable modifications in the arguments.
We now consider two strategies, namely strategy A and strategy B, as follows

**Strategy A**
At $t = 0$: we take a long forward position with delivery as time T and forward price as $F(0,T)$. Also we invest the amount $e^{-rT}\ F(0,T)$ risk free until time T.

At $t = T$: we close the risk free investment and collect the amount $F(0, T)$. We then use this amount to purchase one share of the asset, and sell the same at market price $S(T)$.
Therefore at time T our final wealth will be $F(0, T) + S(T) - F(0, T)$ i.e. $S(T)$.

**Strategy B**
At $t = 0$: we initiate $e^{-r(T-t_1)}$ units of long futures position. This does not involve any cost. Also we invest the amount $e^{-rT} f(0, T)$ risk free until time $T$.
At $t = t_1$: we receive (or pay) the amount $e^{-r(T-t_1)}$ $[f(t_1, T) - f(0, T)]$ as a result of marking to market. We invest (or borrow, depending upon to sign) $e^{-r(T-t_1)}$ $[f(t_1, T) - f(0, T)]$ risk free. We further increase our long futures position to $e^{-r(T-t_2)}$ units; which does not involve any cost.
At $t = t_2$: we cash (or pay) $e^{-r(T-t_2)}$ $[f(t_2, T) - f(t_1, T)]$ as a result of marking to market. We invest (or borrow, depending upon the sign) the amount $e^{-r(T-t_2)}$ $[f(t_2, T) - f(t_1, T)]$ risk free until time $T$. We increase the long futures position to 1, which does not involve any cost.
At $t = T$: we collect the risk free investments which are $f(0, T)+[f(t_1, T)-f(0, T)]+[f(t_2, T)-f(t_1, T)]$, i.e. $f(t_2, T)$. We close the futures position, receiving (or paying) the amount $S(T) - f(t_2, T)$.
Therefore at $t = T$, the final wealth of strategy B is $f(t_2, T)+S(T)-f(t_2, T) = S(T)$.
Since at $t = T$, the final wealth of both strategies are same, under the principle of no arbitrage, the initial wealth needed to initiate Strategy A and Strategy B should be same. Thus

$$e^{-rT} f(0, T) = e^{-rT} F(0, T),$$

i.e.

$$f(0, T) = F(0, T).$$

□

**Example 2.6.1** *Let the interest rate r be* $6\%$ *which is compounded continuously. Let* $S(0) = Rs\ 80$*, and* $S(1)$ *be the asset price after one day. If the marking to market of a futures contract initiated on day 0 (i.e. first day of the year) with delivery in 3 months is zero, then find* $S(1)$*.*

**Solution** We are given that $t = T = 3$ months$= \frac{3}{4}$ year, $t = 0 =$ day $1 = \dfrac{1}{365}$ year, $r = 6\%$ and $S(0) = Rs\ 80$. Hence by formula (2.11)

$$f\left(\frac{1}{365}, \frac{3}{4}\right) - f\left(0, \frac{3}{4}\right) = S\left(\frac{1}{365}\right)\ e^{0.6((3/4)-(1/365))} - S(0)\ e^{(0.6\times 3/4)}.$$

From the given condition, on day 1 the marking to market is zero, i.e.

$$f\left(\frac{1}{365}, \frac{3}{4}\right) - f\left(0, \frac{3}{4}\right) = 0.$$

This gives

$$S\left(\frac{1}{365}\right) = 80\ e^{0.6/365},$$

which is the same as investing *Rs* 80 for one day with risk free rate of 6%.

□

## 2.7 Swaps

There are many investment situations where we desire to transform one cash flow stream into another by appropriate market activity. For example, a company which has sold fixed-coupon bonds and is paying fixed interest may wish to switch into paying the floating rate instead. This can be realized by writing a floating coupon bond and paying a fixed coupon bond with the same present value. This may be achieved by a financial instrument (contract/derivate) called *swap*. Here one party *swaps* a series of fixed-level payments for a series of variable level payments. We can visualize a swap as a series of forward contracts, and hence price the same by using the concept of forward pricing.

We shall have occasion to discuss *interest rate swaps* in a later chapter and therefore, here, we shall introduce a *commodity swap* only. For this consider the below given example.

Let us consider an electric power company that has to purchase oil every month for its power generation facility. If it purchases the oil from the spot market, the company will experience randomly fluctuating cash flows caused by fluctuating spot prices. Therefore the company may desire to swap this payment scheme for the one that is constant. But for this, the company needs to find a counter party willing to swap. Here the swap counter party agrees to pay to the power company the amount ((spot price of oil) × (a fixed number of barrels)), and in return the power company pays to the counter party a fixed price per barrel for the same number of barrels over the life of the swap. In this way, the variable cash flow stream is transformed to a fixed cash flow stream which is depicted in Fig 2.2.

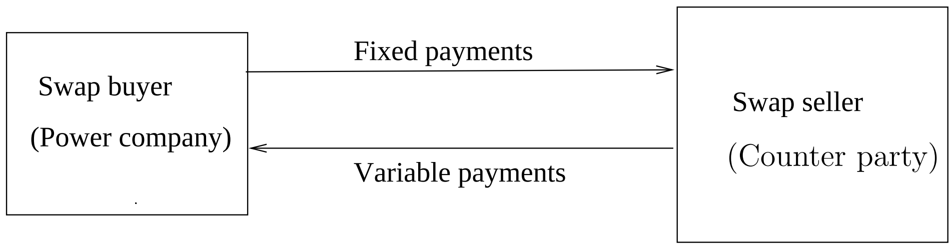



**Fig. 2.2.** Swap

## 2.8 Value of Commodity Swap

Let us consider an agreement where party A receives spot price for N units of a commodity each period while paying a fixed amount X per unit for N units. If the agreement is made for M periods (i.e. the life of the swap in M periods), the net cash flow stream received by A is $(S_1 - X, S_2 - X, \ldots, S_M - X)$ multiplied by the number of units N, where $S_i$ denotes the spot price of the commodity at time $i$.

We can value this stream using the concepts of forward markets. At time zero forward price of the unit of commodity to be received at time $i$ is $F(i)$. By discounting back at time zero, we note that the current value of receiving $S_i$ at time $i$ is $d(0,i)F(i)$, where $d(0,i)$ is the discount factor at time zero for cash received at time $i$. By applying this argument each period, we find that the total value of the stream under consideration is

$$V = \sum_{i=1}^{M} d(0,i)\,(F(i) - X)\,N.$$

Hence the value of the swap can be determined by the series of forward prices. We choose $X$ to make the value $V$ as zero, so that the swap represents an equal exchange.

## 2.9 Summary and Additional Notes

- In this chapter we have discussed two simple, but very popular derivative securities. These are forward contracts and futures contracts. A pricing formula for forward contract has been derived for both, a non-dividend paying stock as well as a dividend paying stock.
- The process of *marking to market* is explained in the context of futures contracts, and a small numerical example is presented for illustration.
- A very brief discussion is included for *swap* and that too in the context of commodity swaps only.
- We have not discussed these (forward contracts, futures and swaps) derivative securities in the context of interest rates or foreign exchange as we plan to discuss them in a later chapter.
- Forward and futures contracts are used to hedge risk in commercial transactions. The simplest type of hedge is the perfect hedge, where the risk associated with a future commitment to deliver or receive an asset is completely eliminated by taking an equal and opposite position in the futures market.
- The perfect hedge may not always be available. This is because there may be no hedging instrument available that matches the commodity of the obligation exactly. In that situation a minimum-variance hedge can be construed using instruments that are correlated with the obligation. We shall refer to Luenberger [85] for details on minimum variance hedging.



## 2.10 Exercises

**Exercise 2.1** *A share of INFOSYS stock can be purchased at* Rs 2,500 *today or at* Rs 2,850 *six months from now. Which of these prices is the spot price, and which is the forward price?*

**Exercise 2.2** *An investor enters a futures contract on SBI stock at* Rs 534 *today. Tomorrow the futures price is* Rs 535. *How much goes into or out of investor's margin account? What will be the answer if tomorrow's price in equal to* Rs 532*?*

**Exercise 2.3** *Let* $B(0) = 100,\ B(1) = 112,\ S(0) = 34$ *and* $T = 1$. *Can the forward price* $F$ *of the stock can be* Rs 38.60*? Justify your answer mathematically.*

**Exercise 2.4** *Explain the difference between entering into a long forward contract with the forward price of* Rs 50 *and buying a call option with strike price of* Rs 50. *You may assume that the other parameters, namely* $S(0)$, $S(1)$, $r$ *and* $T$ *remain same.*

**Exercise 2.5** *At the beginning of April 1 year, the silver forward price (in Rs/Kg) were as follows*

| | |
|---|---|
| *April* | 60,650 |
| *July* | 61,664 |
| *Sept* | 62,348 |
| *Dec* | 63,384 |

*Assume that contracts settle at the end of the given month. The carrying cost of silver is* Rs 2,000 *per kg per year, paid at the beginning of each month. Estimate the interest rate at that time.*

**Exercise 2.6** *Suppose that interest rate* $r$ *is constant. Given* $S(0)$*, find the price* $S(1)$ *of the stock after one day such that the marking to market of futures with delivery in 3 months is zero on that day.*

**Exercise 2.7** *The current price of gold is* Rs 25,000 *per 10 gm. The storage cost is* Rs 200 *per gm per year, payable quarterly in advance. Assuming a constant interest rate of* 9% *compounded quarterly, find the theoretical forward price of gold for delivery in 9 months.*

**Exercise 2.8** *The April 14, 2012 edition of the* Economic Times *gives the following listing for the price of a USD*

- *today: Rs 46.50*
- *90 days forward: Rs 47.60*

*In other words, one can purchase 1 USD today at the price of* Rs 46.50. *In addition, one can sign a contract to purchase 1 USD in 90 days at a price to be paid on delivery of* Rs 47.60. *Let* $r_{IR}$ *and* $r_{USD}$ *be the nominal yearly interest rates being compounded continuously. Obtain the value of* $(r_{INR} - r_{USD})$.

**Exercise 2.9** *Suppose that the value of a stock exchange index (say SENSEX) is 16,500, the futures price for delivery in 9 months is 17,100 index points. If the interest rate* $r$ *is* 8%*, find the dividend yield* $r_{div}$.

**Exercise 2.10** *The difference between the spot and future prices is called the* basis $b(t, T)$. *Show that as* $t \to T,\ b(t, T) \to 0$. *Derive an explicit formula for* $b(t, T)$ *in a market with constant interest rate* $r$.

# 3
# Basic Theory of Option Pricing-I

## 3.1 Introduction

We have already seen some examples of European options in Chapter 1. Though these examples were specific to single step discrete time scenario, they were general enough to guide the basic principle of pricing methodology. There the strategy had been to replicate the given option in terms of stock and bond so that at the end of expiry, the pay-off of the option matches with the value of the replicating portfolio. Then, by no arbitrage principle, the initial value of the portfolio became the price of the given option.

The aim of this chapter is to formalize the above pricing methodology and introduce single and multi-period binomial lattice models for pricing of European and American call/put options. This discussion is continued in Chapter 4 as well, where first the Cox-Ross-Rubinstein (C.R.R.) model is presented, and then the celebrated Black-Scholes formula of option pricing is introduced. The Black-Scholes formula will be re-visited in Chapter 10 after readers have acquired the necessary background in stochastic calculus.

## 3.2 Basic Definitions and Preliminaries

In this section we define European and American call/put options and discuss some of their simple properties.

**Definition 3.2.1 (European Call Option)** *A European call option is a contract giving the holder the right to buy an asset, called the underlying (e.g. stock), for a price* $K$ *fixed in advance, called the strike price or exercise price, at a specified future time* $T$*, called the exercise time or the expiry time.*

Here the term 'underlying' has a general meaning. It could be stock, commodity, foreign currency, stock index or even interest rate. However, unless it is otherwise stated, we shall always mean stock option, i.e. the underlying asset is being taken as stock. Thus a European (stock) call option is a *derivative security* whose underlying security is stock. It gives the holder the right, to buy stock under specified terms as prescribed in Definition 3.2.1. There is absolutely no obligation for the holder to buy stock, but he/she has got the right to buy if he/she so wishes.

A *European put option* is exactly similar to a European call option, except that the word 'buy' changes to 'sell'. Therefore a European put option is a contract giving the holder the right to sell the underlying asset for the strike price $K$ at the exercise time $T$.

Along with the European call/put options another term, namely *American call/put options*, is also very commonly used. The basic distinction between a European option and an American option is that an American option allows exercise at any time before and including the expiry; whereas a European option can be exercised only at the expiry. Here we may note that the word *European* and *American* refer to two different conventions of exercise rather than having any geographical significance. Thus the words *European* and *American* have become two standards in option market, referring to two different structures, no matter where they are issued.

The above discussion suggests that an option can be described by describing its four basic features. These are as listed below.



(i) The description of the underlying asset.
(ii) The nature of the option - whether a call or a put; a European or an American.
(iii) The strike or exercise price.
(iv) The exercise or expiry date.

Next we try to understand the meaning of the term *option pricing*. By definition, the holder of an option gets the right to buy or sell the asset (depending upon the nature of the option) but has no obligation, so some amount has to be paid at the time of contract to get this right. This amount is termed as the *premium* or *price* of the option, and the problem of option pricing revolves around the methodologies to find this price in a *fair* way.

An option has two sides or parties. The party that grants the option is said to *write* the option and the party that obtains the option is said to *purchase* it. The party that purchases the option becomes its holder, and it has no risk of loss other than the original premium paid because it has the right to exercise the option and has no obligation attached to it. Whereas the party that writes the option has major risk. This is because if the option, say a call, is exercised then the writer has to arrange for the asset. If the writer does not already own the asset then it might have to be acquired from the market at a price higher than the agreed strike price. Similarly in the case of a put option, the writer may have to accept the asset at a much lower price (namely the strike price) than what is prevailing in the market. The problem of option pricing is to be *fair* to both the parties and determine the *fair* price of the option under consideration. There is another terminology used for the writer and the purchaser of the option. The buyer (purchaser) of the option is said to have taken the *long position* and the seller (writer) is said to have taken the *short position*. Thus we have terms like *long call*, *short call*, *long put* and *short put* etc.

The next important concept to understand is the *pay-off* or the *value of the option at expiry*. To fix our ideas, let us consider the case of a European call option with strike price $K$ and expiry as $T$. Let $S(T)$ denote the price of the underlying asset (stock) at the expiry. Then, the holder will not exercise the option at the expiry if $S(T) \leq K$ but will certainly exercise the option if $S(T) > K$. Therefore the pay-off to the holder is zero for $S(T) \leq K$ and $S(T) - K$ for $S(T) > K$. If we now introduce the notation $x^+$ as

$$x^+ = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise,} \end{cases}$$

then the pay-off of the given call option can be written as $(S(T) - K)^+$.



**Definition 3.2.2 (Pay-off of a European Call Option)** *Let $C$ be a European call option with specifications as prescribed in Definition 3.2.1. Then*

$$C(T, K, S) = (S(T) - K)^+ = Max\,(S(T) - K,\ 0)\,,$$

*is called the value or the pay-off of the call option $C$. Here $S(T)$ denotes the price of the underlying at the exercise time $T$.*

It is simple to define the pay-off or the value of a European put option $P$. Obviously the holder of a European put option will exercise the option only when $K > S(T)$ and get the profit as $K - S(T)$. Therefore the pay-off or the value of the given European put option is defined as

$$P(T, K, S) = (K - S(T))^+ = \text{Max}(K - S(T),\ 0)\,.$$

The left hand side diagram of Fig. 3.1 depicts the pay-off of a European call option whereas the pay-off of a European put option is depicted in the right hand side diagram of Fig. 3.1.

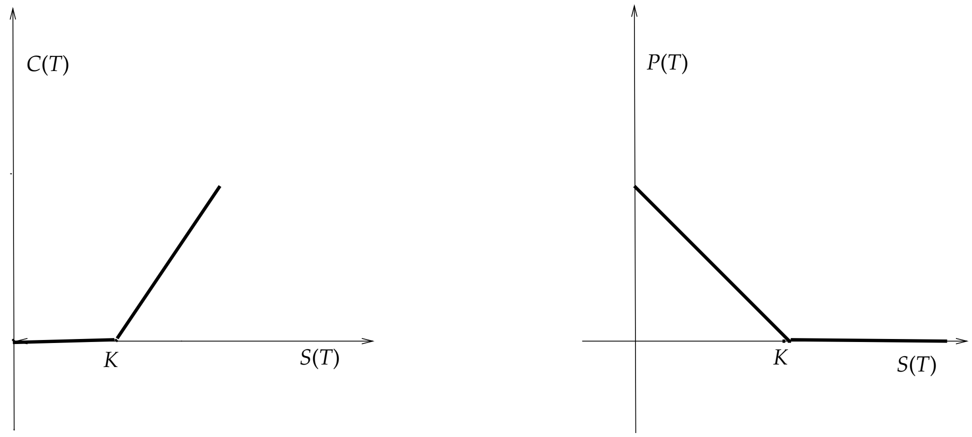



**Fig. 3.1.** Pay-off curves of call and put options

We say that a European call option is *in the money*, *at the money* or *out of money* depending upon $S(T) > K$, $S(T) = K$ or $S(T) < K$. A European put option is *in the money*, *at the money* or *out of money* depending upon $S(T) < K$, $S(T) = K$ or $S(T) > K$.

We also define the *gain* of an option. The gain of an option buyer is the pay-off modified by the premium $C(0)$ or $P(0)$ paid for the option. Thus for the call option $C$, the gain is $(S(T) - K)^+ - C(0)e^{rT}$. Similarly the gain of the buyer of the put option $P$ is $(K - S(T))^+ - P(0)e^{rT}$.

We shall discuss the pay-off corresponding to an American option at a later place. Also sometimes to make the context specific, we shall write $C^E$ or $C^A$ (respectively $P^E$ or $P^A$) to identify whether the option is European or American. However, if only $C$ or $P$ is used, it will always mean a European option.

Though there are many results connecting $C^E$, $P^E$ and $C^A$, $P^A$, the following lemma, called *put-call parity* is interesting and useful.

**Lemma 3.2.1.** ***(Put-Call Parity)*** *Let $C^E$ and $P^E$ be the prices of a European call and a European put defined over the same stock with price $S$. Let the given call and the given put have the same strike price $K$ and the same expiry date $T$. Further, let the underlying stock pay no dividend. Then*

$$C^E(0) - P^E(0) \;=\; S(0) - Ke^{-rT}$$

*where $r$ is the constant risk-free interest rate under continuous compounding.*

Though a formal proof could be given to the above lemma, we present here only an intuitive argument. Suppose we construct a portfolio by writing and selling one put and buying one call option, both with the same strike price $K$ and expiry date $T$. Now if $S(T) \geq K$, then the call will pay $S(T) - K$ and the put will be worthless. If $S(T) < K$ then the call will be worthless and the writer of the put will need to pay $K - S(T)$. In either case, the value of the portfolio will be $S(T) - K$ at expiry. But $S(T) - K$ is also the pay-off of a long forward contract with forward price $K$ and delivery time $T$. Therefore by no arbitrage principle, the current value of the constructed portfolio of options should be that of the forward contract, i.e. $C^E(0) - P^E(0)$ should be $S(0) - Ke^{-rT}$. Thus $C^E(0) - P^E(0) = S(0) - Ke^{-rT}$. Fig. 3.2. depicts this argument.

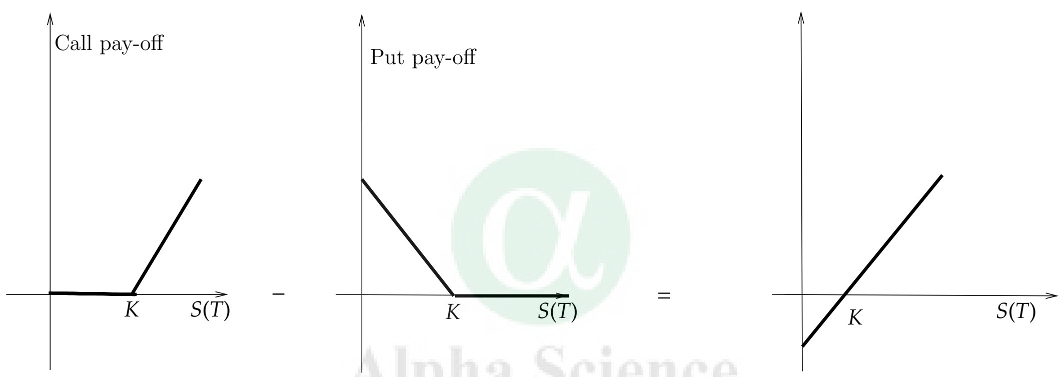



**Fig. 3.2.** Put-call parity



It is also quite obvious that $C^E(0) \leq C^A(0)$ and $P^E(0) \leq P^A(0)$. This is because an American option gives us more freedom to exercise the right and therefore we have to pay more premium in comparison to a European option. Here of course, it is being assumed that both American and European options have the same parameters, $T$, $K$ and $S$.

The below given lemma states that for a non-dividend paying stock, the American call option will never be exercised prior to expiry. This being so, it should be equivalent to the European call option, i.e. $C^A(0) = C^E(0)$.

**Lemma 3.2.2.** *Let $C^E$ and $C^A$ be respectively the prices of a European call and an American call defined over the same stock with price $S$. Let both these calls have the same strike price and the same expiry date. Further, let the stock be non-dividend paying. Then $C^A(0) = C^E(0)$.*

Again rather than giving a formal proof, let us give an intuitive argument for the above lemma. We have already seen that $C^A(0) \geq C^E(0)$ because in comparison to a European call option, an American option gives more right with regard to the time of exercise of the given option. Also, by using the put-call parity, $C^E(0) \geq S(0) - Ke^{-rT}$, as $P^E(0) \geq 0$. Therefore $C^A(0) \geq C^E(0) \geq S(0) - Ke^{-rT}$. But $r > 0$ and hence this last inequality gives $C^A(0) > S(0) - K$.

But $C^A(0) > (S(0) - K)$ implies that the price of the given American call option is greater than the pay-off. Therefore the option should be sold sooner than exercised at $t = 0$. Here the choice of $t = 0$ is for the purpose of reference time and is totally arbitrary. Hence taking any $t < T$ instead of $t = 0$, the above arguments show that the American call option will not be exercised at time t, i.e. the American call option will not be exercised prior to expiry, and then it is equivalent to the European call option. It is important to note here that the assumption of stock being non-dividend paying is crucial for the above assertion. The situation would be different for dividend paying stock. Moreover, this result does not hold for American put, i.e. $P^A(0)$ is different from $P^E(0)$ even when the stock is non-dividend paying (see Example 3.8.1).

Though the above lemma convinces that for a non dividend paying stock, $C^A(0) = C^E(0)$, we attempt to give an independent formal proof of this equality. This proof does not make use of the result that a non-dividend paying stock, an American call option will always be exercised at the expiry.

**Lemma 3.2.3.** *Let the stock be non-dividend paying. Then for the same* $K$, $T$ *and* $r$ *we have* $C^A(0) = C^E(0)$.

**Proof.** We know that $C^A(0) \geq C^E(0)$. Therefore we assume that $C^A(0) > C^E(0)$ and then arrive at a contradiction.

At time $t = 0$, let the investor write and sell an American call so as to get the amount $C^A(0)$. Further let a European call be bought for the amount $C^E(0)$ and the balance $C^A(0) - C^E(0)$ be invested risk free at the interest rate $r$.

If the American call is exercised at time $t < T$, then the investor borrows a share of stock and sells the same for $K$ to the buyer so as to settle his/her obligation as writer of the American call option. Further the investor invests this amount $K$ risk free up to time $T$. Now at time $t = T$, the investor exercises his/her European call option to buy a share of stock for $K$ and closes the short position on stock. This will result in an arbitrage profit of $\left(C^A(0) - C^E(0)\right)e^{rT} + Ke^{r(T-t)} - K > 0$. If the American option is not exercised at all, then the investor will end up with the European call and even an arbitrage profit $\left(C^A(0) - C^E(0)\right)e^{rT} > 0$. Therefore $C^A(0) = C^E(0)$. □

An obvious question at this stage is as follows : Do we have put call parity for American option as well? Unfortunately there is no exact put call parity for American option. But we have certain estimates for $C^A(0) - P^A(0)$ as detailed in the below given lemma.

**Lemma 3.2.4.** ***(Put-Call parity Estimate for American Option)*** *Let the stock be non dividend paying. Then for the same $K$, $T$ and $S$, we have*

$$S(0) - Ke^{-rT} \geq C^A(0) - P^A(0) \geq S(0) - K.$$

**Proof.** We shall first prove that $S(0) - Ke^{-rT} \geq C^A(0) - P^A(0)$. For this, we assume that $C^A(0) - P^A(0) - S(0) + Ke^{-rT} > 0$, and arrive at some contradiction. Now at $t = 0$, we write and sell an American call, buy an American put, buy a share of stock, and finance the transactions in the money market.

If the holder of the American call chooses to exercise it at time $t \leq T$, then we shall receive $K$ for the share of stock and settle the money market position, ending up with the put and a positive amount

$$\begin{aligned} K + \left(C^A(0) - P^A(0) - S(0)\right)e^{rt} &= \left(Ke^{-rt} + C^A(0) - P^A(0) - S(0)\right)e^{rt} \\ &\geq \left(Ke^{-rT} + C^A(0) - P^A(0) - S(0)\right)e^{rt} \\ &> 0. \end{aligned}$$

This argument presumes that $C^A(0) - P^A(0) - S(0) > 0$. In case $C^A(0) - P^A(0) - S(0) < 0$, then this much money is borrowed from the bank at $t = 0$. Then at $t$, we get $K$ because of the sold American call. But under our assumption $K > \left(C^A(0) - P^A(0) - S(0)\right)e^{rt}$ and therefore we close the position in the money market and still having a positive amount. This violates no arbitrage principle.

Next suppose that $C^A(0) - P^A(0) - S(0) + K < 0$. In this case at $t = 0$, we write and sell a put, buy a call, sell short one share of stock and finance the transactions in the money market. If the American put is exercised at $t \leq T$, then we can withdraw $K$ from the money market to buy a share of stock and close the short sale. We shall be left with the call option and a positive amount $\left(-C^A(0) + P^A(0) + S(0)\right)e^{rt} - K > Ke^{rt} - K \geq 0$. If the put is not exercised at all, then we can buy a share of stock for $K$ by exercising the call at time $T$ and close the short position on stock. On closing the money market position, we shall also end up with a positive amount. This again contradicts the no arbitrage principle.

□

## 3.3 Behavior of Option Prices With Respect to Variables

The price of an option depends on a number of variables. These are

(i) the strike price $K$
(ii) the expiry time $T$
(iii) the current stock price $S(0)$
(iv) the dividend rate $r_{div}$, and
(v) the risk free rate r.

Both from theoretical as well as from applications point of view, it should be useful to analyze the behavior of option price as function of one of the variables, keeping the remaining variables constant. We now present a sample of some results and refer to the text by Capinski and Zastawniak [25] for further details. The notation $C^E(K)$ denotes the price of European call options over the same underlying and with the same exercise time, but with different values of the strike price $K$. Similar interpretation is given for other notations, namely $C^E(S)$ and $C^E(T)$ etc.

**Lemma 3.3.1.** *Let $K_1 < K_2$. Then*

*(i)* $C^E(K_1) \geq C^E(K_2)$,
*(ii)* $P^E(K_1) \leq P^E(K_2)$.

Thus $C^E(K)$ is a non-increasing and $P^E(K)$ is a non-decreasing function of $K$. These assertions are obvious because to have the right to buy at a lower price, we have to pay more premium than to have the right to buy at a higher price.

**Lemma 3.3.2.** *Let $K_1 < K_2$. Then*

*(i)* $C^E(K_1) - C^E(K_2) \leq e^{-rT}(K_2 - K_1)$,
*(ii)* $P^E(K_2) - P^E(K_1) \leq e^{-rT}(K_2 - K_1)$.

**Proof.** From put-call parity, we have

$$C^E(K_1) - P^E(K_1) = S(0) - K_1 e^{-rt},$$

and

$$C^E(K_2) - P^E(K_2) = S(0) - K_2 e^{-rt}.$$

If we now subtract the above two equations, we get

$$C^E(K_1) - C^E(K_2) + P^E(K_2) - P^E(K_1) = (K_2 - K_1)e^{-rt}.$$

But then an application of Lemma 3.3.1 gives the result. □

**Lemma 3.3.3.** *Let* $K_1 < K_2$ *and* $0 \le \alpha \le 1$. *Then* $C^E(K)$ *and* $P^E(K)$ *satisfy*

*(i)* $C^E(\alpha K_1 + (1-\alpha)K_2) \le \alpha C^E(K_1) + (1-\alpha)C^E(K_2)$
*(ii)* $P^E(\alpha K_1 + (1-\alpha)K_2) \le \alpha P^E(K_1) + (1-\alpha)P^E(K_2)$.

*This means that* $C^E(K)$*and* $P^E(K)$ *are convex functions of* $K$.

**Proof.** Let us denote $\widehat{K} = \alpha\ K_1 + (1-\alpha)\ K_2$. Suppose that $C^E(\widehat{K}) > \alpha\ C^E(K_1) + (1-\alpha)\ C^E(K_2)$.
Now at $t = 0$, let the investor write and sell the option with strike $\widehat{K}$ and get the amount (premium) $C^E(\widehat{K})$. Then let the investor purchase $\alpha$ option with strike price $K_1$ and $(1-\alpha)$ options with strike price $K_2$. This will involve a payment (premium) of $\alpha\ C^E(K_1) + (1-\alpha)\ C^E(K_2)$. But then under the assumed inequality the investor is still left with the amount $C^E\ (\widehat{K}) - (\alpha\ C^E(K_1) + (1-\alpha)\ C^E(K_2)) > 0$ which he/she invests risk free.

At the expiry, if the option with the strike price $\widehat{K}$ is exercised, then the investor shall have to pay the amount $\left(S(T) - \widehat{K}\right)^+ = \text{Max}\left(S(T) - \widehat{K}, 0\right)$. But he/she can raise the amount $\alpha\ (S(T) - K_1)^+ + (1-\alpha)\ (S(T) - K_2)^+$ by exercising $\alpha$ calls with strike price $K_1$ and $(1-\alpha)$ calls with strike price $K_2$. But it is known that

$$\left(S(T) - \widehat{K}\right)^+ \le \alpha\ (S(T) - K_1)^+ + (1-\alpha)\ (S(T) - K_2)^+ ,$$

because the function $(S(T) - K)^+$ is a convex function of $K$. Therefore the investor will realize an arbitrage profit of $\left[C^E(\widehat{K}) - (\alpha\ C^E(K_1) + (1-\alpha)\ C^E(K_2))\right]\ e^{rT}$ at expiry. Therefore our supposition is wrong and we get the result.

□

The below given Lemmas can also be proved on the similar lines.

**Lemma 3.3.4.** *Let* $T_1 < T_2$. *Then*

*(i)* $C^E(T_1) \le C^E(T_2)$
*(ii)* $P^E(T_1) \le P^E(T_2)$.

**Proof.** If possible let $C^E(T_1) > C^E(T_2)$. We write and sell one option expiring at time $T_1$ and buy one with the same strike price but expiry as $T_2$, investing the balance without risk. If the written option is exercised at $T_1$, we can exercised the option immediately to cover our liability. The balance $(C^E(T_1) - C^E(T_2)) > 0$ invested without risk will be our arbitrage profit.

The inequality *(ii)* can be proved on similar lines. □

**Lemma 3.3.5.** *Let* $0 < x_1 < x_2$. *Further let at time* $t = 0$, $\overline{S} = x_1 S(0)$ *and* $\hat{S} = x_2 S(0)$. *Then*

*(i)* $C^E(\overline{S}) \leq C^E(\hat{S})$
*(ii)* $P^E(\overline{S}) \geq P^E(\hat{S})$.

**Proof.** We shall prove the first inequality. If possible let $C^E(\overline{S}) > C^E(\hat{S})$. We can write and sell a call on a portfolio with $x_1$ shares and buy a call on a portfolio with $x_2$ shares having the same strike price $K$ and exercise time $T$. Also we invest the amount $C^E(\overline{S}) - C^E(\hat{S})$ risk-free. As $x_1 < x_2$, we have $(x_1 S(T) - K)^+ \leq (x_2 S(T) - K)^+$. If the sold option is exercised at time $T$, we can exercised other option to cover our liability. The balance $C^E(\overline{S}) - C^E(\hat{S}) > 0$ invested risk-free will be our arbitrage profit.

□

**Lemma 3.3.6.** *Let* $\overline{S}$ *and* $\hat{S}$ *be as defined in Lemma 3.3.5. Then*

*(i)* $C^E(\hat{S}) - C^E(\overline{S}) \leq (\hat{S} - \overline{S})$
*(ii)* $P^E(\overline{S}) - P^E(\hat{S}) \leq (\hat{S} - \overline{S})$.

**Proof.** Using put-call parity we have

$$C^E(\hat{S}) - P^E(\hat{S}) = \hat{S} - Ke^{-rT},$$

and

$$C^E(\overline{S}) - P^E(\overline{S}) = \overline{S} - Ke^{-rT}.$$

On subtraction, we get

$$\left(C^E(\hat{S}) - C^E(\overline{S})\right) + \left(P^E(\overline{S}) - P^E(\hat{S})\right) = (\hat{S} - \overline{S}),$$

which gives inequalities (i) and (ii) because both terms on the right hand side are non-negative.

□

Results similar to above Lemmas also hold for American options. For these details we may refer to Capinski and Zastawniak [25].

## 3.4 Pay-off Curves of Options Combinations

The graph between the pay-off and the stock price at the expiry is called the *pay-off curve of the option*. Often we need to invest in combinations of options to take care of specific hedging or speculative strategies. Knowing the pay-off curves of individual options we can obtain the pay-off curve of the combinations of options. This curve will be a combination of connected straight line segments depending upon the number and nature of securities in the specific combination. This pay-off curve of the combination is called its *spread*. We now give examples of some of the most common spreads.

**Bull Spread**

Consider a scenario in which the investor expects the stock price to rise and wants to speculate on that. Obviously the investor should buy a call option say $C_{K_1}$, with strike price $K_1$ which is close to the current stock price. Here we have used the symbol $C_{K_1}$ rather than $C(K_1)$ for the sake of simplicity. The premium may be reduced by selling a call option, say $C_{K_2}$, with strike price $K_2 > K_1$. This strategy should bring good returns provided the stock price increases are moderate. The spread of the combination $C_{K_1} - C_{K_2}$ is called the *bull spread*, which is depicted in Fig 3.3.

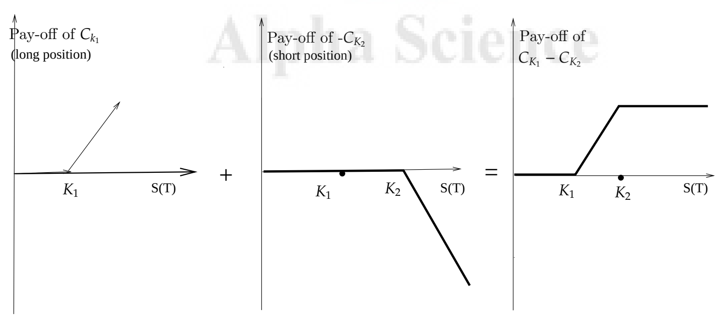





**Fig. 3.3.** Bull spread

**Bear Spread**

The pay-off curve of the combination $C_{K_1} - C_{K_2}$ with $(K_1 > K_2)$ gives rise to a *bear spread*. This is usually employed by an investor who is expecting a moderate decline in the stock price. The bear spread is depicted in Fig. 3.4.

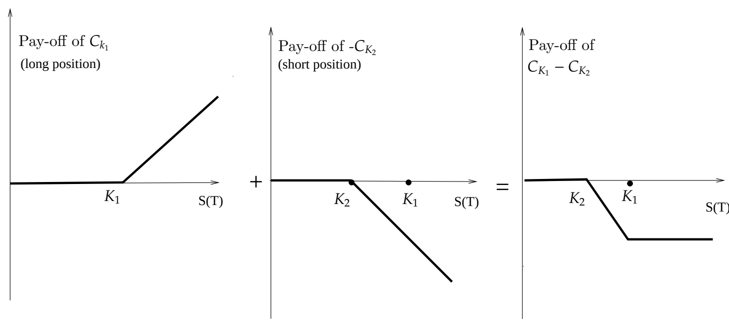



**Fig. 3.4.** Bear spread

## Butterfly Spread

Let $K_1 < K_2 < K_3$. Let $C_{K_1}$, $C_{K_2}$ and $C_{K_3}$ be call options with strike prices $K_1$,

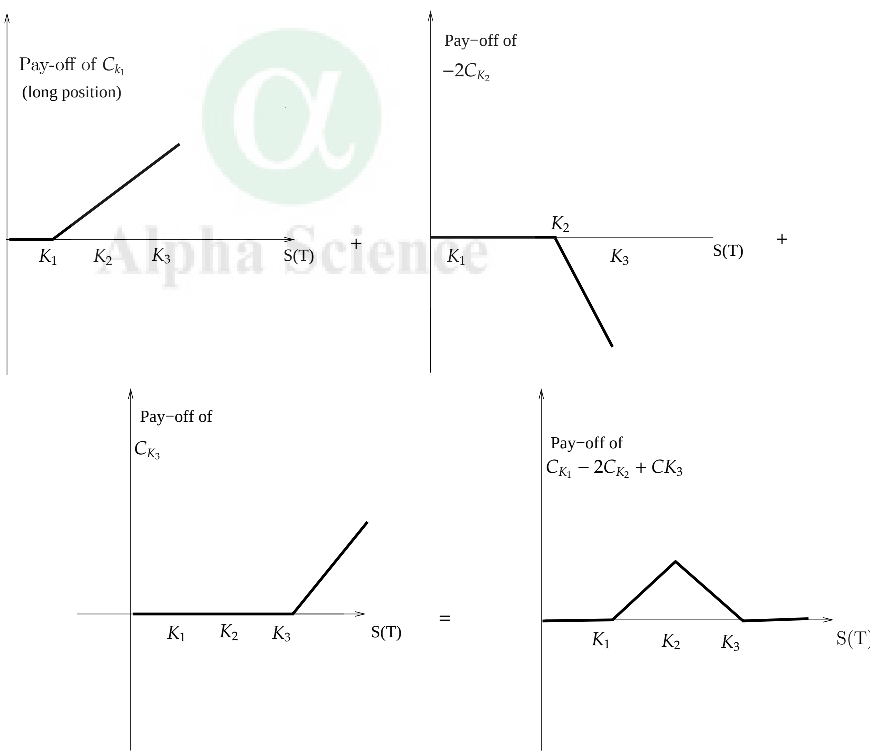





**Fig. 3.5.** Butterfly spread

$K_2$ and $K_3$ respectively. Let these calls be defined over the same stock and have common expiry $T$. Then the pay-off curve of the combination $C_{K_1} - 2C_{K_2} + C_{K_3}$ gives rise to a *butterfly spread.* This option combination is used by an investor who feels that the stock price will generally remain unaltered, i.e. it will not change significantly. The butterfly spread is depicted in Fig 3.5.

The above discussion suggests that by forming the combinations of options, any pay-off function can be approximated by a sequence of straight line segments. In other words, any continuous pay-off function can be made close to pay-off of an appropriate option combination.

Now similar to pay-off curve, we can also sketch *gain curve* of a given option combination. The readers can sketch the gain curves for the above examples in an obvious manner by utilizing the definition of gain of an option.

**Example 3.4.1** *Let the sale and purchase of options over the same stock and having the same expiry be given by the expression*

$$-P_{100} + P_{120} + 2C_{150} - C_{180}.$$

*Sketch the spread of the given option combination.*



**Solution** To draw the spread, we draw the individual pay-off curves and then combine them. The details in Fig. 3.6 are self explanatory.

□

## 3.5 Single Period Binomial Lattice Model for Option Pricing

The simple example of pricing a European option, presented in Chapter 1, is in fact an example of a single period binomial lattice model. Here we formally develop the theory for the single period case and extend the same for multi-period case in the next section. For this model we assume the following

(i) The initial value of the stock is $S(0)$, i.e. the stock price at $t = 0$ is $S(0)$.
(ii) At the end of the period, the price is either $u\,S(0)$ with probability $p$, or $d\,S(0)$ with probability $(1 - p)$, $0 < p < 1$.
(iii) $u > d > 0$.
(iv) At every period, it is possible to borrow or lend at a common risk free interest rate $r$. Let $R = (1 + r)$, and $u > R > d$. This assumption is needed to avoid arbitrage.

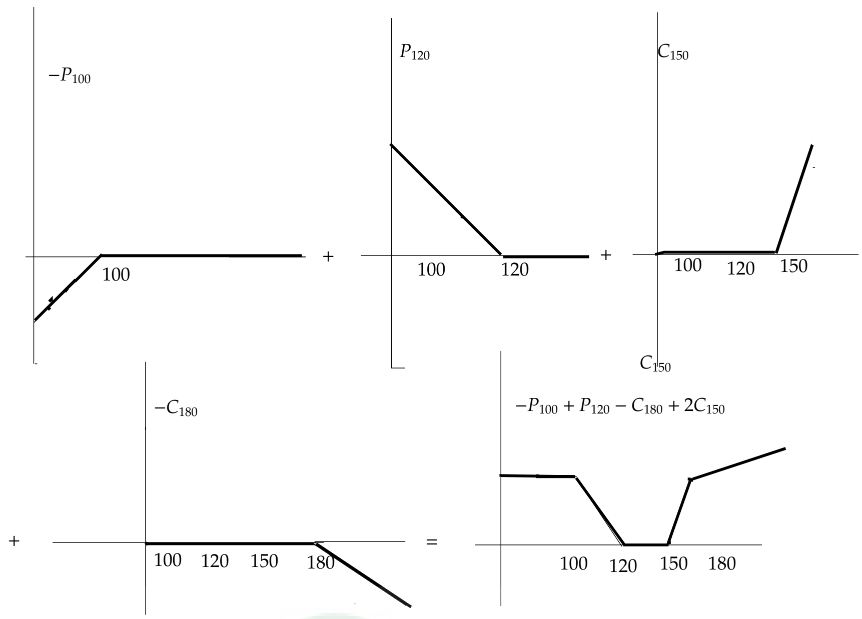



**Fig. 3.6.** Spread of Example 3.4.1



Let $C$ be the given European call option on the stock with strike price $K$ and expiry at the end of period (i.e. $t = 1$). Our aim is to find $C(0)$. Since at $t = 1$, the stock price could be either $u\ S(0)$ (with probability $p$) or $d\ S(0)$ (with probability $(1 - p)$). We have

$$C(1) = \begin{cases} C_u = \text{Max}(u\ S(0) - K,\ 0), & \text{with probability } \ p \\ C_d = \text{Max}(u\ S(1) - K,\ 0), & \text{with probability } \ (1-p). \end{cases}$$

Though the value of the risk free asset (say bond) is deterministic, we can take it as a (degenerate) derivative of the stock-degenerate in the sense that the same value $R$ is assigned at the end of each arc. Thus we have the following figures


Before we actually derive the pricing formula for the given scenario, let us try to justify the assumption $u > R > d$. For this we have below given lemma.

**Lemma 3.5.1.** *If $u > R > d$ does not hold then "no arbitrage principle" is violated.*

**Proof.** We first consider the case $R \geq u > d$. Now construct the portfolio $P$ : $\left(x = -\frac{1}{S(0)},\ y = \frac{1}{B(0)}\right)$, where $B(0)$ denotes the price of the bond at $t = 0$. Let $V_P(0)$ and $V_P(1)$ respectively denote the value of the portfolio P at $t = 0$ and $t = 1$. Then

$$V_{\mathrm{P}}(0) = x\ S(0) + y\ B(0) = \left(-\frac{1}{S(0)}\right)S(0) + \left(\frac{1}{B(0)}\right)B(0) = 0,$$

and

$$\begin{aligned} V_{\mathrm{P}}(1) &= x\ S(1) + y\ B(1) \\ &= \begin{cases} -\frac{1}{S(0)}u\ S(0) + \frac{1}{B(0)}R\ B(0), & \text{with probability } p \\ -\frac{1}{S(0)}d\ S(0) + \frac{1}{B(0)}R\ B(0), & \text{with probability } (1-p) \end{cases} \\ &= \begin{cases} R - u, & \text{with probability } p \\ R - d, & \text{with probability } (1-p). \end{cases} \end{aligned}$$

But under the assumption $R \geq u > d$, $(R - u) \geq 0$ and $(R - d) > 0$. As $0 < p < 1$, $V_P(1) \geq 0$ with probability 1 and it can take positive value with positive probability even though $V_P(0) = 0$. This clearly violates the "no arbitrage principle".

We next consider the case $u > d \geq R$, and construct the portfolio $Q$ : $\left(x = \frac{1}{S(0)}, y = -\frac{1}{B(0)}\right)$ and note that $V_Q(0) = 0$. Further $V_Q(1) \geq 0$ with probability 1 and it can take positive value $(u - R)$ with positive probability, which again violates the "no arbitrage principle". Therefore to avoid the arbitrage opportunity, we need to assume that $u > R > d$.

□

Now to find the price $C(0)$ of the call, we need to replicate it in terms of stock and bond so that at $t = 1$, the value of this replicating portfolio equals the pay-off of the call at $t = 1$. Let the replicating portfolio be $RP : (x = a, y = b)$, where $a$ is the number of shares of stock and $b$ is the units of bond. Then we need to find $a$ and $b$ such that $V_{RP}(1) = C(1)$. But

$$\begin{aligned} V_{\mathrm{RP}}(1) &= a\ S(1) + b\ B(1) \\ &= \begin{cases} a\ u\ S(0) + b\ R\ B(0), & \text{with probability } p \\ a\ d\ S(0) + b\ R\ B(0), & \text{with probability } (1-p), \end{cases} \end{aligned}$$

and

$$C(1) = \begin{cases} C_u, & \text{with probability } p \\ C_d, & \text{with probability } (1-p). \end{cases}$$

Hence $V_{\mathrm{RP}}(1) = C(1)$ gives

$$a\ u\ S(0) + b\ R\ B(0) = C_u \tag{3.1}$$

$$a\ d\ S(0) + b\ R\ B(0) = C_d. \tag{3.2}$$

Solving the above system we obtain

$$a = \frac{C_u - C_d}{S(0)(u-d)}, \tag{3.3}$$

and

$$b = \frac{u\ C_d - d\ C_u}{R\ (u-d)B(0)}. \tag{3.4}$$

Also

$$V_{\mathrm{RP}}(0) = a\ S(0) + b\ B(0). \tag{3.5}$$

Therefore using (3.3), (3.4) and (3.5) we have

$$\begin{aligned} V_{\mathrm{RP}}(0) &= a\ S(0) + b\ B(0) \\ &= \left(\frac{C_u - C_d}{S(0)(u-d)}\right) S(0) + \left(\frac{u\ C_d - d\ C_u}{R\ (u-d)A(0)}\right) B(0) \\ &= \left(\frac{C_u - C_d}{(u-d)}\right) + \left(\frac{u\ C_d - d\ C_u}{R\ (u-d)}\right) \\ &= \frac{1}{R}\left(\left(\frac{R-d}{u-d}\right) C_u + \left(\frac{u-R}{u-d}\right) C_d\right) \\ &= \frac{1}{R}\left(\hat{p}\ C_u + (1-\hat{p})\ C_d\right), \end{aligned} \tag{3.6}$$

where $\hat{p} = \left(\frac{R-d}{u-d}\right)$ and $(1-\hat{p}) = \left(\frac{u-R}{u-d}\right)$.

Since at $t = 1$, the value $V_{RP}(1)$ of the replicating portfolio and the pay-off $C(1)$ of the call are equal, by no arbitrage principle, we should have $C(0) = V_{\mathrm{RP}}(0)$. Hence from (3.6)

$$C(0) = \frac{1}{R}\left(\hat{p}\ C_u + (1-\hat{p})\ C_d\right). \tag{3.7}$$

In (3.7) we should note that $0 < \hat{p} < 1$ because $u > R > d$. Hence $\hat{p}$ can be considered a probability. This probability $\hat{p}$ is called the *risk neutral probability*. Before we discuss its interpretation and nomenclature, we observe from (3.7) that $C(0) = \frac{1}{R} E_{\hat{p}}(C(1))$, where $E_{\hat{p}}$ denotes the expectation under risk neutral probability $\hat{p}$. The formula

$$C(0) = \frac{1}{R} E_{\hat{p}}(C(1)) \tag{3.8}$$

is very general and is valid for any derivative security with "appropriate" modifications.

In the above derivation, there is no role of $B(0)$. In fact we may take $B(0) = 1$, giving $B(1) = R$. Therefore we can think of bond as cash, as Rs 1 gives $Rs\ (1+r) = Rs\ R$ after one time period. In this light, formula (3.8) tells that to find the price $C(0)$ of the call, we should first take the expectation of the pay-off at expiry with respect to the risk neutral probability, and then discount the same according to the risk free rate.

Here we should note that risk neutral probability $\hat{p}$ is different from $p$. The actual probability $p$ describes the (stochastic) price movement but the risk neutral probability $\hat{p}$ has a totally different interpretation. In fact the actual probability $p$ enters nowhere in the derivation of $C(0)$. However it gives a motivation to introduce risk neutral probability. This is because one would invest in stock only if the expected growth rate of stock is higher than that of the rate at money market (bank/bond/cash), i.e.

$$p\,(u\ S(0)) + (1-p)\,(d\ S(0)) > R\ S(0),$$

i.e.

$$S(0) < \frac{1}{R}\left[p\,(u\ S(0)) + (1-p)\,(d\ S(0))\right]. \tag{3.9}$$

But

$$\hat{p}\,(u\ S(0)) + (1-\hat{p})\,(d\ S(0)) = \left(\frac{R-d}{u-d}\right)(u\ S(0)) + \left(\frac{u-R}{u-d}\right)(d\ S(0)),$$

which on simplification gives

$$\frac{1}{R}\left[\hat{p}\,(u\ S(0)) + (1-\hat{p})\,(d\ S(0))\right] = S(0). \tag{3.10}$$

Equation (3.10) tells that under risk neutral probability $\hat{p}$, the mean rate of growth of stock equals the rate of growth in the money market account which is risk free. Now if this is the case then the investor must be neutral about the risk and hence the name *risk neutral probability*. The formula (3.7) is therefore referred as the *risk neutral pricing formula* for European call option.

The role of risk neutral probability is very important because it provides *fair* opportunity to the writer of the option to *hedge* his/her risk and thereby provides a *fair* price of the call. The hedging problem in the present context is to find an equivalent replicating portfolio, i.e. a portfolio of stock and bond at $t = 0$ such that the value of this portfolio at $t = 1$ matches with the pay-off of the call. This process of replicating the option is called the *hedging problem*, and the replicating portfolio is called the *hedge of the given option*.

**Remark 3.5.1** *Though the actual probabilities* $p$ *and* $(1 - p)$ *have not entered in the pricing formula (3.7), they have played an important role indirectly. Thus the investor will like to buy a call option provided he/she feels that there is high probability of the stock price going up at* $t = 1$*. Also the agreed strike price* $K$ *depends to some extent on these probabilities. This is because if there is a feeling that at* $t = 1$*, the stock price will go up with high value of* $p$ *then he/she may agree for a higher value of* $K$*.*

The above procedure of finding the call price is valid provided we guarantee that the risk neutral probability measure (RNPM) always exists and is unique. The below given result tells that under no arbitrage principle RNPM always exists. Further if in addition, the market is *complete*, then the RNPM is unique.



**Lemma 3.5.2.** *A risk neutral probability measure exists if and only if no arbitrage principle holds. Further if the market is complete then, the RNPM is unique.*

We shall partly prove this Lemma in Section 3.7 by making use of the concept of duality in linear programming. Specifically we shall prove that RNPM exists if and only if, no arbitrage principle holds. The discussion about the uniqueness of RNPM will be postponed till the concept of completeness of the market is introduced.

**Example 3.5.1** *Find the price of a European call option with the given data as* $B(0) = 100$*,* $B(1) = 110$*,* $S(0) = 100$*,* $K = 100$*,* $T = 1$ *and*

$$S(1) = \begin{cases} 120, & \text{with probability } 0 \cdot 6 \\ 80, & \text{with probability } 0 \cdot 4. \end{cases}$$

*Will the price change if the probabilities* $p$ *and* $(1 - p)$ *are taken as 0.3 and 0.7 respectively?*

**Solution** As $B(1) = 110$, $B(0) = 100$, we get $r = 10\%$, i.e. $R = 1 + r = 1.1$. Further $S(0) = 100$, $u\ S(0) = 120$ and $d\ S(0) = 80$ yield $u = 1.2$ and $d = 0.8$. Here the risk neutral probability $\hat{p}$ is $\dfrac{3}{4}$ as shown below.

$$\hat{p} = \frac{R-d}{u-d} = \frac{1.1-0.8}{1.2-0.8} = \frac{3}{4}$$

i.e. $1 - \hat{p} = 1 - \dfrac{3}{4} = \dfrac{1}{4}$.
Also

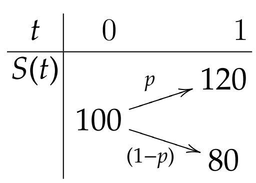
$$C_u = \text{Max}(u\ S(0) - K,\ 0) = \text{Max}(120 - 100, 0) = 20$$
$$C_d = \text{Max}(d\ S(0) - K,\ 0) = \text{Max}(80 - 100, 0) = 0.$$

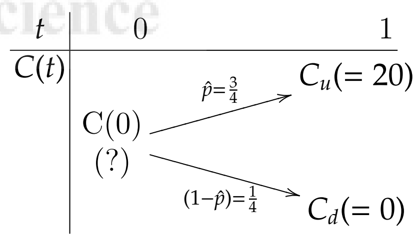
Therefore

$$\begin{aligned} C(0) &= \frac{1}{R}\,[\hat{p}\ C_u + (1-\hat{p})\ C_d] \\ &= \frac{1}{1.1}\left[\frac{3}{4}(20) + \frac{1}{4}(0)\right] \\ &= 13.63. \end{aligned}$$

We exhibit the above details in the form of following tables






This gives $C(0) = Rs\ 13.64$. As call price does not depend on $p$, it will not change if 0.6 and 0.4 are changed to 0.7 and 0.3 respectively.

□

To find the price of a European put option we can follow exactly the same derivation as for the European call. We need to use pay-off of the put at expiry instead of the pay-off of the call to get the following pricing formula

$$P(0) = \frac{1}{R}\,[\hat{p}\ P_u + (1-\hat{p})\ P_d]\ , \tag{3.11}$$

where $P_u = \text{Max}\,(K - u\ S(0),\ 0)$ and $P_d = \text{Max}\,(K - d\ S(0),\ 0)$.

**Example 3.5.2** *For the data given in Example 3.5.1, find the price of the corresponding European put option.*

**Solution** We have already obtained $\hat{p} = \frac{3}{4}$, $(1 - \hat{p}) = \frac{1}{4}$ and $R = 1.1$. Further $P_u = \text{Max}\,(K - u\ S(0),\ 0) = 0$ and $P_d = \text{Max}\,(K - d\ S(0),\ 0) = 20$. Therefore

$$
\begin{aligned}
P(0) &= \frac{1}{R}\,[\hat{p}\ P_u + (1-\hat{p})\ P_d] \\
&= \frac{1}{1.1}\left[\left(\frac{3}{4} \times 0\right) + \left(\frac{1}{4} \times 20\right)\right] \\
&= 4.54,
\end{aligned}
$$

and the corresponding table is

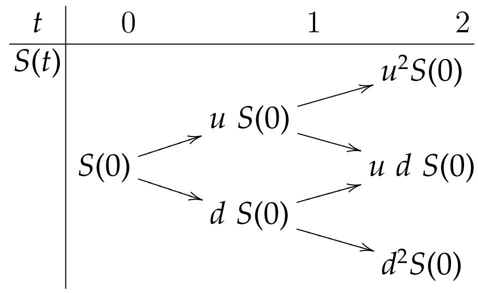
| $t$ | 0 | | 1 |
|---|---|---|---|
| $P(t)$ | | $\hat{p}$ | $P_u (= 0)$ |
| | $P(0)$ | | |
| | | $(1-\hat{p})$ | $P_d (= 20)$ |

At this stage we can also verify the put-call parity. For the given data we have already obtained $C(0)$ =Rs 13.63 and $P(0)$ =Rs 4.54. Hence the value of the expression $C(0) - P(0) + d(0,1)K$ comes out to be $13.64 - 4.54 + (1.1)^{-1} \times 100 = 100$, which is same as the value $S(0)$. Here $d(0,1) = \frac{1}{R}$ is the discount factor between $t = 0$ and $t = 1$.

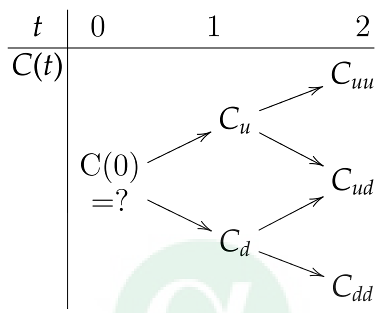
□

## 3.6 Multi Period Binomial Lattice Model for Option Pricing

We shall now extend the pricing formula (3.7) to multi period binomial lattice model. For this let us first take the two period case. Our notations are same as those in Section 3.5. Thus $S(0)$ is the price of the stock at $t = 0$ (initial time), $u$ and $d$ are up and down factors of the stock price movement, $K$ is the strike price, $t = 2$ is expiry and $R = (1 + r)$. The following table illustrates the price movement of the stock along the binomial lattice



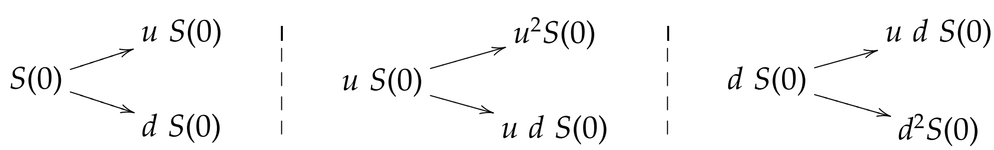

Along the line of the single period case, the table for the call price should be




Here $C_{uu}$, $C_{ud}$ and $C_{dd}$ are pay-offs of the call at the expiry (i.e. at $t = 2$). As per our definition, it is obvious that

$$
\begin{aligned}
C_{uu} &= \mathrm{Max}(u^2 S(0) - K,\ 0)\\
C_{ud} &= \mathrm{Max}(u\ d\ S(0) - K,\ 0)\\
C_{dd} &= \mathrm{Max}(d^2 S(0) - K,\ 0).
\end{aligned}
$$

But what are the values of $C_u$ and $C_d$? For this let us re-look the binomial lattice for the price of the stock. We can visualize this bigger lattice as a combination of three binomial lattices of the following type




Let above lattices be referred as $L(1)$, $L(2)$ and $L(3)$ respectively. Treating lattice $L(3)$ as the single period binomial lattice and using formula (3.7), we obtain

$$
C_d = \frac{1}{R}\left[\hat{p}\ C_{ud} + (1 - \hat{p})\ C_{dd}\right], \tag{3.12}
$$

where as before $\hat{p}$ is the risk neutral probability given by $\hat{p} = \dfrac{R-d}{u-d}$. Similarly using lattice $L(2)$ we obtain

$$C_u = \frac{1}{R}\left[\hat{p}\ C_{uu} + (1-\hat{p})\ C_{ud}\right]. \tag{3.13}$$

Finally using lattice $L(1)$ we obtain

$$C_0 = \frac{1}{R}\left[\hat{p}\ C_u + (1-\hat{p})\ C_d\right]. \tag{3.14}$$

Here the same risk neutral probability $\hat{p}$ has been used for all three lattices $L(1)$, $L(2)$ and $L(3)$. The reason being that for these single period lattices the up-tick probability $p$, the down-tick probability $1-p$ and the factors $u$ and $d$ do not change - they remain same for $t = 0$ as well as for $t = 1$. Now substituting for $C_d$ and $C_u$ from (3.12) and (3.13), equation (3.14) gives

$$C_0 = \frac{1}{R^2}\left[\hat{p}^2 C_{uu} + 2\hat{p}\ (1-\hat{p})\ C_{ud} + (1-\hat{p})^2 C_{dd}\right]. \tag{3.15}$$

Formula (3.15) can also be written as

$$C_0 = \frac{1}{R^2}\left[\sum_{j=0}^{2} \frac{2!}{j!(2-j)!}(\hat{p})^j (1-\hat{p})^{2-j}\left(u^j d^{2-j} S(0) - K\right)^+\right], \tag{3.16}$$

where $\left(u^j d^{2-j} S(0) - K\right)^+ = \text{Max}\left(u^j d^{2-j} S(0) - K,\ 0\right)$.

**Remark 3.6.1** *Here, unlike the single period case,* $C_u$ *and* $C_d$ *are NOT the pay-offs - Max*$(u\ S(0) - K,\ 0)$ *and Max*$(d\ S(0) - K,\ 0)$ *at* $t = 1$, *because the call is being exercised at* $t = 2$ *only.*

In deriving formula (3.15), we have not made use of the concept of replicating portfolio directly. This has been possible because we have visualized the given 2-period binomial lattice as a combination of three 1-period binomial lattices which have already been studied. But it will be interesting to derive the same formula by employing the replicating portfolio strategy as well.

Though we are not giving the complete details here, we are giving enough hint to get the complete solution. In order to have replication at maturity $t = 2$ we start replicating backwards, from the end of lattice. For this at $t = 1$, we consider the upper node of the lattice $L_2$ and determine the scalars $a_1$ and $b_1$ such that

$$a_1\left(u^2S(0)\right) + b_1R = C_{uu},$$

and

$$a_1\,(udS(0)) + b_1R = C_{ud}.$$

This system can be solved to get

$$a_1 = \frac{C_{uu} - C_{ud}}{uS(0)(u-d)},$$

and

$$b_1 = \frac{C_{uu} - a_1(u^2S(0))}{R}.$$

Then $C_u = a_1\,(uS(0)) + b_1$ becomes the value of the replicating portfolio in the upper node at $t = 1$. We can similarly find the value of $C_d$. Then we need to replicate these two values in the first period. Thus we need to find $a_0$ and $b_0$ such that

$$a_0\,(uS(0)) + b_0 = C_u,$$

and

$$a_0\,(dS(0)) + b_0 = C_d.$$

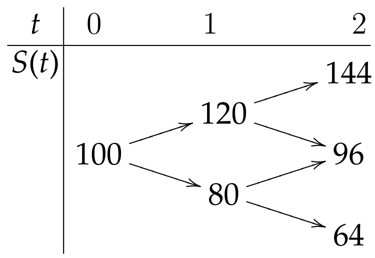
The above system can be solved as before to get the values of $a_0$ and $b_0$. Now a little manipulation and the expression $C(0) = a_0S(0) + b_0R$ will give the formula (3.15).

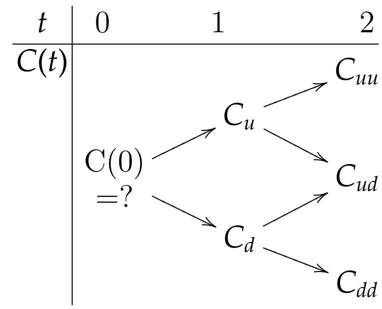


**Remark 3.6.2** *It is to be noted that in this model, the number of shares in the replicating portfolio is always equal to the ratio of $\Delta C$ and $\Delta S$, where $\Delta C$ is the change in the future values of the call and $\Delta S$ is the change in the future values of the stock. This ratio is called the* delta *of the call, and it is different at different nodes of the lattice. Delta is one of the* Greeks *to be studied later for a derivative security defined over a given undertaking. Greeks have been used extensively in derivative pricing for hedging purposes.*

**Example 3.6.1** *Find the price of a European call option with the given data as* $S(0) = 100$, $K = 100$, $u = 1.2$, $d = 0.8$, $r = 10\%$ *per year and time to expiry* $T = 2$ *years.*

**Solution** We have




From the given data $R = 1 + r = 1.1$ and $\hat{p} = \dfrac{R-d}{u-d} = \dfrac{1.1 - 0.8}{1.2 - 0.8} = \dfrac{3}{4}$. Also

$$\begin{aligned} C_{uu} &= \text{Max}(u^2 S(0) - K,\ 0) = \text{Max}(144 - 100, 0) = 44 \\ C_{ud} &= \text{Max}(u\ d\ S(0) - K,\ 0) = \text{Max}(96 - 100, 0) = 0 \\ C_{dd} &= \text{Max}(d^2 S(0) - K,\ 0) = \text{Max}(64 - 100, 0) = 0. \end{aligned}$$

Therefore

$$\begin{aligned} C_u &= \frac{1}{R}[\hat{p}\ C_{uu} + (1-\hat{p})\ C_{ud}] = \frac{1}{1.1}\left[\left(\frac{3}{4} \times 44\right) + \left(\frac{1}{4} \times 0\right)\right] = 30 \\ C_d &= \frac{1}{1.1}\left[\left(\frac{3}{4} \times 0\right) + \left(\frac{1}{4} \times 0\right)\right] = 0, \end{aligned}$$

which gives



$$\begin{aligned} C(0) &= \frac{1}{R}[\hat{p}\ C_u + (1-\hat{p})\ C_d] \\ &= \frac{1}{1.1}\left[\left(\frac{3}{4} \times 30\right) + \left(\frac{1}{4} \times 0\right)\right] \\ &= 20.45. \end{aligned}$$

Alternatively, we can use the formula (3.15) to get

$$\begin{aligned} C(0) &= \frac{1}{R^2}\left[\hat{p}^2 C_{uu} + 2\hat{p}\ (1-\hat{p})\ C_{ud} + (1-\hat{p})^2 C_{dd}\right] \\ &= \frac{1}{1.1^2}\left[\left(\frac{9}{64} \times 44\right)\right] \\ &= 20.45. \end{aligned}$$

□

**Example 3.6.2** *For the data given in Example 3.6.1, find the price of the corresponding European put option.*

**Solution** We have already obtained $\hat{p} = \frac{3}{4}$, $(1 - \hat{p}) = \frac{1}{4}$ and $R = 1.1$. Further

$$\begin{aligned} P_{uu} &= \text{Max}(K - u^2 S(0),\ 0) = \text{Max}(-44, 0) = 0 \\ P_{ud} &= \text{Max}(K - u\ d\ S(0),\ 0) = \text{Max}(4, 0) = 4 \\ P_{dd} &= \text{Max}(K - d^2 S(0),\ 0) = \text{Max}(36, 0) = 36. \end{aligned}$$

Therefore

$$\begin{aligned} P_u &= \frac{1}{1.1}\left[\left(\frac{3}{4} \times 0\right) + \left(\frac{1}{4} \times 4\right)\right] = \frac{1}{1.1} \\ P_d &= \frac{1}{1.1}\left[\left(\frac{3}{4} \times 4\right) + \left(\frac{1}{4} \times 36\right)\right] = \frac{12}{1.1}. \end{aligned}$$

This gives

$$\begin{aligned} P(0) &= \frac{1}{R}\left[\hat{p}\ P_u + (1 - \hat{p})\ P_d\right] \\ &= \frac{1}{1.1}\left[\left(\frac{3}{4} \times \frac{1}{1.1}\right) + \left(\frac{1}{4} \times \frac{12}{1.1}\right)\right] \\ &= 3.10. \end{aligned}$$

Alternatively, we can get

$$\begin{aligned} P(0) &= \frac{1}{R^2}\left[\hat{p}^2 P_{uu} + 2\hat{p}\ (1 - \hat{p})\ P_{ud} + (1 - \hat{p})^2 P_{dd}\right] \\ &= \frac{1}{1.1^2}\left[\left(\left(\frac{3}{4}\right)^2 \times 0\right) + \left(2 \times \frac{3}{4} \times \frac{1}{4} \times 4\right) + \left(\left(\frac{1}{4}\right)^2 \times 36\right)\right] \\ &= 3.10. \end{aligned}$$

We can also use put-call parity to find $P(0)$, once $C(0)$ is known. Specifically

$$C(0) - P(0) + d(0, 2)K = S(0),$$

gives

$$P(0) = C(0) + d(0, 2)K - S(0) = 20.45 + d(0, 2)100 - 100.$$

We can take $d(0, 2) = (1 + 2r)^{-1}$ or $e^{-2r}$ depending upon the nature of the compounding of the interest rate $r$; but it has to be same for the entire calculations.

□

The development of the pricing methodology for the case of multi period binomial lattice model is similar to the one discussed for the two period lattice model. The single period risk free discounting is carried out at every node of the lattice as has been done for the two period case. Obviously we need to start from the final period ($t = N$) and then proceed backward till we reach the initial time period ($t = 0$). For the case of European call option, this process will result in the following formula

$$C(0) = \frac{1}{R^N}\left[\sum_{j=0}^{N} \frac{N!}{j!(N-j)!}(\hat{p})^j(1-\hat{p})^{N-j}\left(u^j d^{N-j} S(0) - K\right)^+\right], \tag{3.17}$$

which reduces to (3.16) for $N = 2$.

## 3.7 Existence of Risk Neutral Probability Measure

We have already seen in Section 3.5 that for the single period binomial pricing model, risk neutral probability measure (RNPM) plays the most fundamental role. We have also exhibited the RNPM $\hat{p}$ and used the same for deriving the option pricing formula (3.7). Here it may be noted that in this scenario the state space $\Omega$ consists of only two points say $\Omega = \{\omega_1, \omega_2\}$ where $\omega_1$ is the up-tick movement of the stock price and $\omega_2$ is the down-tick movement of the stock price. But realistically $\Omega$ may consist of $m$ points $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ or $\Omega$ may be an interval $[a, b]$. Can we still guarantee the existence of RNPM? To answer this question is important if we wish to discuss derivative pricing for more general scenarios.

Let us consider a more general model where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$, $\omega_j$ being the $j^{th}$ state of economy. Thus $\Omega$ represents the finite set of possible values; e.g. these could be $m$ possible values of the stock price $S(1)$ (at $t = 1$), namely $S_1(\omega_1)$, $S_1(\omega_2)$, ..., $S_1(\omega_m)$. For the single period binomial lattice, $\Omega = \{\omega_1, \omega_2\}$ with $S_1(\omega_1) = uS(0)$ and $S_1(\omega_2) = dS(0)$.

Also we could have more than two securities, say $S^{(k)}$ $(k = 0, 1, 2, \ldots, n)$. Let $S_1^{(k)}(\omega_j)$, for $(k = 0, 1, 2, \ldots, n)$ and $(j = 1, 2, \ldots, m)$ denote the price of the $k^{th}$ security at $t = 1$ when the state of the economy is $\omega_j$. We may think of $k$ securities as bond, stock, option, forward contracts, etc having different defining variables but over the same underlying.

Let $S_0^{(k)}$ $(k = 0, 1, 2, \ldots, n)$ denote the current $(t = 0)$ price of the $k^{th}$ security. Here we may note that $S_0^{(k)}$ is deterministic, but $S_1^{(k)}$ is a random variable taking values $\left\{S_1^{(k)}(\omega_j),\ (j = 1, 2, \ldots, m)\right\}$. Since all derivative securities are defined on the same underlying (stock), they all will be random variables taking $m$ possible values.

As a convention, we take $k = 0$ for bond, $k = 1$ for the underlying (stock) and $(k = 2, \ldots, n)$ for other (derivative) securities. Let $r$ be the risk free interest rate for the period $t = 0$ to $t = 1$. It is convenient to assume that $S_0^{(0)} = 1$ and $S_1^{(0)}(\omega_j) = R = (1 + r),\ (j = 1, 2, \ldots, m)$.

**Definition 3.7.1 (Risk Neutral Probability Measure)** *A risk neutral probability measure (RNPM) is a vector $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m)^T$ such that*

*(i)* $\hat{p}_j > 0$ *$(j = 1, 2, \ldots, m)$,*

*(ii)* $\displaystyle\sum_{j=1}^{m} \hat{p}_j = 1$,

*and for every security $k$ $(k = 0, 1, 2, \ldots, n)$, we have*

$$S_0^{(k)} = \frac{1}{R}\left(\sum_{j=1}^{m} \hat{p}_j S_1^{(k)}(\omega_j)\right). \tag{3.18}$$

If we denote by $E_{\hat{p}}\left(S_1^{(k)}\right)$ the expected value of $S_1^{(k)}$ with respect to RNPM $\hat{p}$, then (3.18) can be written as

$$S_0^{(k)} = \frac{1}{R} E_{\hat{p}}\left(S_1^{(k)}\right). \tag{3.19}$$

Here $S_1^{(k)}$ denotes the value of the $k^{th}$ security at $t = 1$ and $S_0^{(k)}$ is the price of the same security at $t = 0$.

Regarding the existence and uniqueness of RNPM we have the following two main theorems

**Theorem 3.7.1 (First Fundamental Theorem of Asset Pricing)** *A risk neutral probability measure $\hat{p}$ exists if and only if no arbitrage principle holds.*

**Theorem 3.7.2 (Second Fundamental Theorem of Asset Pricing)** *The RNPM is unique if and only if the market is complete.*

Thus for an arbitrage free market, there is unique RNPM $\hat{p}$ if and only if the market is *complete*. We have not yet discussed the meaning of *market completeness* but that we postpone for the time being.

We shall now give a linear programming based proof of Theorem 3.7.1. Let us recall the following primal-dual pair of linear programming problem

$$
\begin{array}{lll}
(LP) & \text{Min} & c^T x \\
 & \text{subject to} & \\
 & & Ax \geq b \\
 & & x \geq 0,
\end{array}
$$

and

$$
\begin{array}{lll}
(LD) & \text{Max} & b^T y \\
 & \text{subject to} & \\
 & & A^T y \leq c \\
 & & y \geq 0.
\end{array}
$$

It is well known in duality theory that if (LP) and (LD) both are feasible then both have optimal solutions. The below given theorem, called *strict complementarity theorem*, gives some additional information as well.

**Theorem 3.7.3 (Goldman-Tucker Theorem)** *Let (LP) and (LD) both be feasible. Then they both have optimal solutions $x^*$ (for(LP)) and $y^*$ (for(LD)) satisfying*

$$x^* + (c - A^T y^*) > 0. \tag{3.20}$$

The condition (3.20) is called the strict complementarity condition. Here it may be noted that Theorem 3.7.3 does not tell that (3.20) holds for every pair $(\overline{x}, \overline{y})$ of optimal solution of (LP) and (LD). But rather it guarantees the existence of a pair $(x^*, y^*)$ of optimal solution of (LP) and (LD) for which (3.20) holds. We can refer to Goldman and Tucker [52] for the proof of Theorem 3.7.3.

We now proceed to prove Theorem 3.7.1. For this let us consider a portfolio $P : (x_0, x_1, \ldots, x_n)$. Then the value of this portfolio at $t = 0$ is

$$V_P(0) = \sum_{k=0}^{n} x_k \, S_0^{(k)}.$$

Further its value at $t = 1$ will be one of the $m$ values, namely $V_{P,1}(\omega_j) = \sum_{k=0}^{n} x_k \, S_1^{(k)}(\omega_j)$, $(j = 1, 2, \ldots, m)$, depending upon the state of economy $\omega_j$ where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$. We now consider the following linear programming problem (LPP)

$$
\begin{array}{lll}
\text{Min} & V_P(0) & \\
\text{subject to} & & \\
 & V_{P,1}(\omega_j) \geq 0 & (j = 1, 2, \ldots, m),
\end{array}
$$

i.e.

$$\text{Min} \quad \sum_{k=0}^{n} x_k \; S_0^{(k)}$$

subject to

$$\sum_{k=0}^{n} x_k \; S_1^{(k)}(\omega_j) \geq 0 \qquad (j = 1, 2, \ldots, m). \tag{3.21}$$

The dual of problem (3.21) is

$$\text{Max} \quad \sum_{j=1}^{m} 0.p_j$$

subject to

$$\begin{aligned} \sum_{j=1}^{m} S_1^{(k)}(\omega_j) \; p_j &= S_0^{(k)} \qquad (k = 0, 1, 2, \ldots, n) \\ p_j &\geq 0 \qquad (j = 1, 2, \ldots, m). \end{aligned} \tag{3.22}$$



Next we note that LPP (3.21) is feasible because $\{x_k = 0, \; (k = 0, 1, 2, \ldots, n)\}$ satisfies all its constraints. Also this is optimal because by definition to meet the assumption of solvency, only admissible portfolios are to be considered, i.e. $V_P(0) \geq 0$, i.e. $\sum_{k=0}^{n} x_k \; S_0^{(k)} \geq 0$. But (3.21) is a minimization problem and therefore $\{x_k = 0, \; (k = 0, 1, 2, \ldots, n)\}$ is optimal to (3.21) and the optimal value is zero. Therefore its dual LPP (3.22) also has an optimal solution with optimal value as zero.

We now write the primal-dual pair (3.21) - (3.22) in the form of the pair (LP)-(LD) and then apply the Goldman-Tucker theorem (Theorem 3.7.3) to this pair. This implies that there exists $\overline{x}$ optimal to (3.21) and $\overline{p}$ optimal to (3.22) such that

$$-\sum_{k=0}^{n} S_1^{(k)}(\omega_j)\overline{x}_k \leq 0 \qquad (j = 1, 2, \ldots, m), \tag{3.23}$$

$$\sum_{j=1}^{m} S_1^{(k)}(\omega_j)\ \overline{p}_j = S_0^{(k)} \qquad (k = 0, 1, 2, \ldots, n), \tag{3.24}$$

$$\overline{p}_j \geq 0 \qquad (j = 1, 2, \ldots, m), \tag{3.25}$$

$$\sum_{j=1}^{m} 0.\overline{p}_j = \sum_{k=0}^{n} S_0^{(k)}\ \overline{x}_k = 0, \tag{3.26}$$

and

$$\overline{p}_j + \left(0 - \left(-\sum_{k=0}^{n} S_1^{(k)}(\omega_j)\ \overline{x}_k\right)\right) > 0 \qquad (j = 1, 2, \ldots, m). \tag{3.27}$$

But from (3.26), $V_{\overline{p}}(0) = 0$ for the portfolio $(\overline{x}_k,\ (k = 0, 1, 2, \ldots, n))$. Hence by *no arbitrage principle*, $V_{\overline{p}}(1)$ should be zero with probability 1, i.e. $\sum_{k=0}^{n} S_1^{(k)}(\omega_j)\ \overline{x}_k = 0\ (j = 1, 2, \ldots, m)$. Therefore (3.27) gives $\overline{p}_j > 0\ (j = 1, 2, \ldots, m)$.

Now we recall that $k = 0$ refers to the bond. Taking $S_0^{(0)} = 1$ we get $S_1^{(0)}(\omega_j) = 1 + r = R\ (j = 1, 2, \ldots, m)$. Then (3.24) gives

$$\sum_{j=1}^{m} R\ \overline{p}_j = 1,$$

i.e.

$$\sum_{j=1}^{m} \overline{p}_j = \frac{1}{R}. \tag{3.28}$$

If we now define $\hat{p}$ such that $\hat{p}_j = R\ \overline{p}_j\ (j = 1, 2, \ldots, m)$, then (3.24), (3.25) and (3.28) give $\hat{p}_j > 0\ (j = 1, 2, \ldots, m)$, $\sum_{j=1}^{m} \hat{p}_j = 1$ and

$$\frac{1}{R}\left(\sum_{j=1}^{m} S_1^{(k)}(\omega_j)\ \hat{p}_j\right) = S_0^{(k)} \qquad (k = 0, 1, 2, \ldots, n). \tag{3.29}$$

But then this shows that $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_j, \ldots, \hat{p}_m)$ is a risk neutral probability measure.

We can also write (3.29) in a more familiar notation as

$$S_0^{(k)} = \frac{1}{R} E_{\hat{p}} \left( S_1^{(k)} \right) \qquad (k = 0, 1, 2, \ldots, n), \tag{3.30}$$

where $E_{\hat{p}}$ denotes the expectation under RNPM $\hat{p}$. This proves that if no arbitrage principle holds then RNPM exists. The converse can be proved on similar lines.

□

**Remark 3.7.1** *The formula (3.30) is very general. Apart from the fact that it holds for the stock and bond, it should hold for other securities ($k = 2, 3, \ldots, n$) as well, be it European call, put, forward contract etc. What we simply need to do is to find the expectation of the pay-off of the given security at expiry under RNPM and discount the same for $t = 0$.*

**Remark 3.7.2** *For the case of single period binomial lattice model the linear programming problem to find RNPM is*

$$\begin{aligned}
&\textit{Max} \quad 0\,\hat{p}_1 + 0\,\hat{p}_2 \\
&\textit{subject to} \\
&\qquad (u\ S(0))\,\hat{p}_1 + (d\ S(0))\,\hat{p}_2 = RS(0) \\
&\qquad \hat{p}_1 + \hat{p}_2 = 1 \\
&\qquad \hat{p}_1 \geq 0, \hat{p}_2 \geq 0,
\end{aligned}$$



*such that at optimality $\hat{p}_1 > 0,\ \hat{p}_2 > 0$. It is not difficult to see that the unique optimal solution is* $\left( \hat{p}_1 = \dfrac{R-d}{u-d},\ \hat{p}_2 = \dfrac{R-u}{u-d} \right)$*, which is the same as obtained earlier by replicating portfolio arguments.*

*Also if $k = 2$ and $k = 3$ respectively refer to European call and European put options, then formula (3.30) gives*

$$C(0) = \frac{1}{R} \left[ \hat{p}_1\ C_u + \hat{p}_2\ C_d \right],$$

*and*

$$P(0) = \frac{1}{R} \left[ \hat{p}_1\ P_u + \hat{p}_2\ P_d \right].$$

**Example 3.7.1** *Consider a forward contract with the given data as $S(0) = 100,\ u = 1.2,\ d = 0.8,\ T = 1$ year and $r = 10\%$ per year. Determine the forward price $F$.*

**Solution** We know that $F = RS(0) = (1.1) \times 100 = Rs\ 110$. But here we have to determine $F$ by utilizing the formula (3.30). For this we note that the pay-off of the forward contract is $(S - F)$ if $S > F$ and $-(F - S)$ if $S \leq F$. Also from the given data $(\hat{p}_1 = \frac{3}{4}, \hat{p}_2 = \frac{1}{4})$. Therefore formula (3.30) gives

$$F(0) = \frac{1}{1.1}\left[\frac{3}{4}(120 - F) + \frac{1}{4}(80 - F)\right]. \tag{3.31}$$

But in the case of forward contract $F(0) = 0$, and then (3.31) gives $F = 110$.

□

Let us again look at formula (3.29) and concentrate for the case $k = 1$. We recall that $k = 1$ refers to the stock and therefore if we write $S_0^1 = S(0)$ and $S_1^1 = S(1)$ then

$$\begin{aligned} S(0) &= \frac{1}{R}E_{\hat{p}}(S(1)) \\ &= E_{\hat{p}}\left(\frac{S(1)}{R}\right) \\ &= E_{\hat{p}}\left(\frac{S(1)}{B(1)}\right) \\ &= E_{\hat{p}}(\widetilde{S}(1)). \end{aligned} \tag{3.32}$$



Here we have taken $B(0) = 1$, $\widetilde{S}(1) = (S(1)/B(1))$ and written $E_{\hat{p}}$ to emphasize that expectation has been taken with respect to RNPM $\hat{p}$. Therefore (3.32) gives

$$\widetilde{S}(0) = E_{\hat{p}}(\widetilde{S}(1)) \tag{3.33}$$

The expectation in (3.32) is in fact conditional because it is computed once the stock price $S(0)$ becomes known at $t = 0$. Therefore (3.33) is actually

$$\widetilde{S}(0) = E_{\hat{p}}\left(\widetilde{S}(1))/\widetilde{S}(0)\right),$$

which, in general, is expressed as

$$\widetilde{S}(l) = E_{\hat{p}}\left((\widetilde{S}(l+1))/\widetilde{S}(l)\right). \tag{3.34}$$

The relationship (3.34) is expressed as: *the discounted stock prices* $\widetilde{S}(0), \widetilde{S}(1), \widetilde{S}(2), \ldots$ *form a martingale with respect to RNPM* $\hat{p}$. This holds for other securities as well, i.e.

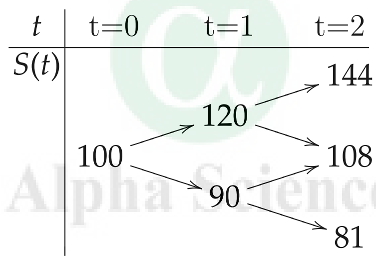
$$\widetilde{S}^{(k)}(l) = E_{\hat{p}}\left(\widetilde{S}^{(k)}(l+1)/\widetilde{S}^{(k)}(l)\right) \quad (k = 0, 1, 2, \ldots, n).$$

Therefore the problem of asset pricing gets translated into the problem of finding a unique RNPM $\hat{p}$ or to be precise $\hat{p}$-*martingale* on the set of scenarios $\Omega$. Moreover $E_{\hat{p}}$ is called the *risk neutral* or *martingle expectation* with respect to $\hat{p}$. These concepts will probably get much clearer once we are familiar with basics of stochastic process and stochastic calculus.

Since there are no goods in the model other than money, it is convenient to pick one of the security as reference and normalize others with respect to it. The security so chosen for normalization purpose is called *numeraire*. In our context, the usual choice of numeraire is the bond price and that is what exactly we have done in defining $\widetilde{S}(0) = S(0)/B(0)$, $\widetilde{S}(1) = S(1)/B(1)$ etc.

**Example 3.7.2** *Consider the data:* $S(0) = 100$, $u = 1.2$, $d = 0.9$, $r = 10\%$ *per year and* $T = 2$ *years. Determine RNPM* $\hat{p}$ *and show that discounted stock prices form a* $\hat{p}$*-martingale.*

**Solution** We have $\hat{p} = \frac{2}{3}$ and $(1-\hat{p}) = \frac{1}{3}$. We have the dynamics of the stock price as




We can check the following

$$\frac{1}{1.1}\left(\frac{2}{3}(144) + \frac{1}{3}(108)\right) = 120$$
$$\frac{1}{1.1}\left(\frac{2}{3}(108) + \frac{1}{3}(81)\right) = 90$$
$$\frac{1}{1.1}\left(\frac{2}{3}(120) + \frac{1}{3}(90)\right) = 100.$$

These give

$$E_{\hat{p}}\left[\left(\frac{S(2)}{(1.1)^2}\right)\Big/\left(\frac{S(1)}{1.1}\right) = \frac{120}{1.1}\right] = \frac{120}{1.1}$$
$$E_{\hat{p}}\left[\left(\frac{S(2)}{(1.1)^2}\right)\Big/\left(\frac{S(1)}{1.1}\right) = \frac{90}{1.1}\right] = \frac{90}{1.1}$$

etc. Thus

$$E_{\hat{p}}\left(\widetilde{S}(2)/\widetilde{S}(1)\right) = \widetilde{S}(1).$$

□

The below given theorem is very general and is valid for any European derivative security.

**Theorem 3.7.4** *Let $D$ be a European derivative security whose pay-off in the $N$-period binomial model is $f(S(N))$. Then*

$$D(0) = \frac{1}{R^N} E_{\hat{p}}\left(f\left(S(N)\right)\right). \tag{3.35}$$

The above theorem essentially tells that the price of a European derivative security $D$ with pay-off $f(S(N))$ in the $N$-period binomial model is the expectation of the discounted pay-off under the risk neutral probably measure.

We know that for a European call, $f(S(N)) = (S(N) - K)^+$; for a European put $f(S(N)) = (K - S(N))^+$ and for a forward contract $f(S(N)) = (S(N) - F)$. Therefore the same formula (3.35) can be used to price each of these derivative securities.

**Example 3.7.3** *Use Theorem 3.7.1 to justify the pricing formula (3.15) for the European call option.*

**Solution** Here $N = 2$ and the three possible value for $f(S(2))$ are

$$C_{uu} = \text{Max}\left(u^2 S(0) - K,\ 0\right)$$
$$C_{ud} = \text{Max}\left(u\ d\ S(0) - K,\ 0\right)$$
$$C_{dd} = \text{Max}\left(d^2 S(0) - K,\ 0\right).$$

Also $\hat{p} = \dfrac{(R-d)}{(u-d)}$. Therefore if we define $\hat{p}_1 = (\hat{p})^2$, $\hat{p}_2 = 2\hat{p}(1-\hat{p})$ and $\hat{p}_3 = (1-\hat{p})^2$ then

(i) $\hat{p}_i > 0 \quad (i = 1, 2, 3)$

(ii) $\displaystyle\sum_{i=1}^{3} \hat{p}_i = 1$ and

(iii) $E_{\hat{p}}\left(S(2)\right) = R^2 S(0)$.

The third assertion can be verified as follows

$$
\begin{aligned}
E_{\hat{p}}\,(S(2)) &= \hat{p}_1\left(u^2 S(0)\right) + \hat{p}_2\,(u\ d\ S(0)) + \hat{p}_3\left(d^2 S(0)\right) \\
&= S(0)\left[u^2(\hat{p})^2 + 2\hat{p}\ (1-\hat{p})\ u\ d\ S(0) + (1-\hat{p})^2 d^2 S(0)\right] \\
&= S(0)\left[u\ \hat{p} + (1-\hat{p})\ d\right]^2 \\
&= S(0)\left[\frac{u(R-d) + d(u-R)}{(u-d)}\right]^2 \\
&= S(0)\left[\frac{R(u-d)}{u-d}\right]^2 = R^2 S(0),
\end{aligned}
$$

i.e.

$$
S(0) = \frac{1}{R^2}\left(E_{\hat{p}}(S(2))\right).
$$

Therefore $\hat{p}_i$ $(i = 1, 2, 3)$ as defined above is in fact RNPM. Hence

$$
\begin{aligned}
C(0) &= \frac{1}{R^2}\left(E_{\hat{p}}\,(f\,(S(2)))\right) \\
&= \frac{1}{R^2}\left[(\hat{p})^2 C_{uu} + 2\hat{p}\ (1-\hat{p})\ C_{ud} + (1-\hat{p})^2 C_{dd}\right],
\end{aligned}
$$

as obtained by formula (3.15).

□

**Remark 3.7.3** *Apparently we have employed two distinct approaches to price a given derivative. These are the replicating portfolio approach and the RNPM approach. Various illustrative examples presented above and also the derivation of single period binomial lattice model suggest that these two approaches are related. In fact for a complete market, the two approaches are equivalent. This is because the replicating portfolio approach uses the law of one price which is essentially a consequence of no arbitrage principle. This is because then the unique RNPM exists which can be used to price any contingent claim. For the existence of RNPM, the market has to be arbitrage free. Therefore if no arbitrage principle does not hold then we cannot price the given derivative even if the corresponding unique replicating portfolio exists. The below given example illustrates the point. We shall further discuss this aspect in Section 3.10.*

**Example 3.7.4** *Let* $B(0) = 100, B(1) = 120,\ S(0) = 100$ *and* $S(1)$ *take values* $120$ *and* $80$ *with probabilities* $0.8$ *and* $0.2$ *respectively. Let* $C$ *be a European call with* $K = 100$ *and* $T = 1$ *year. Find the replicating portfolio* $(x, y)$ *for the call* $C$*. Are we justifying in taking* $C(0) = x\ S(0) + y\ B(0)$*? Determine RNPM if it exists.*

**Solution** The pay-off of the call $C$ is

$$C(1) = \begin{cases} 20, & \text{with probability } 0.8 \\ 0, & \text{with probability } 0.2. \end{cases}$$

Therefore if $(x, y)$ is the replicating portfolio, then we have $x\,S(1) + y\,B(1) = C(1)$. This gives the following two equations

$$120x + 120y = 20$$

$$80x + 120y = 0,$$

having solution as $(x = 1/2, y = -1/3)$. Here we cannot take $C(0) = 1/2S(0) - 1/3B(0)$ because this utilizes the law of one price which is valid only under no arbitrage principle. But as $u = 1.2$ and $R = 1 + r = 1.2$ we do not have the required condition $u > R > d$ for no arbitrage principle to hold.

To determine the RNPM we need to solve the system

$$\hat{p}_1 + \hat{p}_2 = 1$$

$$\frac{1}{1.2}(120\hat{p}_1 + 80\hat{p}_2) = 100$$

$$\hat{p}_1 > 0,\ \hat{p}_2 > 0.$$

The above system does not have a solution because the first two equations give $\hat{p}_1 = 1,\ \hat{p}_2 = 0$. Therefore the RNPM does not exist. This is again because $u = 1.2 = R$ which violates the condition $u > R > d$ and therefore no arbitrage principle does not hold.

□

## 3.8 Pricing American Options: A Binomial Lattice Model

Following earlier notations, let $D$ denote the derivative security under consideration and stock be its underlying. Let $D^A(n)$ denote the price of an American option at the time period $n$, $0 \leq n \leq N$. To be specific, let $N = 2$ and $D$ denote either an American call or an American put. We now wish to analyze the given American option expiring after two time periods.

If the option has not already been exercised then at the expiry it will have the value $D^A(2) = f(S(2))$, where f is the pay-off function of the derivative $D$.

Here we may note that for the two period binomial lattice model, $f(S(2))$ will have three possible values, depending upon the three possible values of $S(2)$. If $D$ denote an American put then these values are denoted by $P^A_{uu}$, $P^A_{ud}$ and $P^A_{dd}$. For the case of an American call these values are denoted by $C^A_{uu}$, $C^A_{ud}$ and $C^A_{dd}$. Since $D$ is a American derivative and the binomial lattice is of two periods, the holder can exercise his/her option at $t = 0$, $t = 1$ or $t = 2$.

At $t = 1$, the option holder has a choice. He/She can exercise immediately or wait until $t = 2$. If he/she exercises his/her option at $t = 1$, the pay-off is $f(S(1))$, otherwise it is $f(S(2))$. But is it worth waiting until $t = 2$ to exercise the option? This question can be answered by computing the *value of waiting*. For this we may treat $f(S(2))$ as a one step European option to be priced at $t = 1$. Let as before $\hat{p} = \dfrac{(R-d)}{(u-d)}$ denote one step RNPM. Then the value of the given derivative at $t = 1$ is given by $\dfrac{1}{R}E_{\hat{p}}(f(S(2)))$, where

$$\frac{1}{R}E_{\hat{p}}(f(S(2))) = \frac{1}{R}\left(\hat{p}\ f(u\ S(1)) + (1-\hat{p})\ f(d\ S(1))\right). \tag{3.36}$$

Therefore the option holder has to choose the higher of two values $f(S(1))$ and $\dfrac{1}{R}E_{\hat{p}}(f(S(2)))$ as given at (3.36). Thus at $t = 1$, the given American option is worth the price $D^A(1)$, where

$$D^A(1) = \text{Max}\left[f(S(1)), \frac{1}{R}(\hat{p}\ f(u\ S(1)) + (1-\hat{p})\ f(d\ S(1)))\right].$$

In a similar manner we can argue to get

$$D^A(0) = \text{Max}\left[f(S(0)), \frac{1}{R}(\hat{p}\ f_1(u\ S(0)) + (1-\hat{p})\ f_1(d\ S(0))\right], \tag{3.37}$$

where

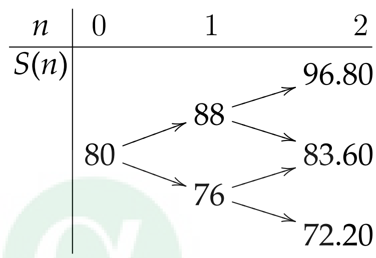
$$f_1(x) = \text{Max}\left[f(x), \frac{1}{R}(\hat{p}\ f(u\ x) + (1-\hat{p})\ f(d\ x))\right]. \tag{3.38}$$

**Remark 3.8.1** *As we have assumed that the stock is non-dividend paying, an American call is the same as a European call and* $C^A(0) = C^E(0)$. *Therefore the formula (3.37) is essentially for the case of an American put, though it can be applied for an American call as well if one so wishes. For a dividend paying stock, an American call will behave differently from a European call.*

**Remark 3.8.2** *Using backward induction, formula (3.37) can be extended to an N-period binomial lattice model in an obvious manner. Unlike the closed form formula (3.35) for the case of European scenario, we do not have any closed form formula for pricing of an American derivative. We do not have any choice except to resort to the backward induction as explained above.*

**Example 3.8.1** *Find the price of an American put option with the given data as* $S(0) = 80,\ K = 80,\ u = 1.8,\ d = 0.95,\ r = 5\%$ *per year and time to expiry* $T = 2$ *years.*

**Solution** We detail the solution in below given steps
**Step 1.** The table for the stock price at different periods is




Also $\hat{p} = \dfrac{R-d}{u-d} = \dfrac{1.05-0.95}{1.8-0.95} = \dfrac{2}{3},\quad (1-\hat{p}) = \dfrac{1}{3}.$

**Step 2.** Let $P^A(n)$, $(n = 0, 1, 2)$, denote the price of given American put. To determine $P^A(n)$, we first need to compute



$$P^f_{uu} = \text{Max}\left(K - u^2 S(0),\ 0\right) = \text{Max}(80 - 96.80, 0) = 0$$
$$P^f_{ud} = \text{Max}\,(K - u\ d\ S(0),\ 0) = \text{Max}\,(80 - 83.60, 0) = 0$$
$$P^f_{dd} = \text{Max}\left(K - d^2 S(0),\ 0\right) = \text{Max}(80 - 72.2, 0) = 7.80.$$

Here $P^f_{uu}$ etc are same as for the European put option. We have used the suffix '$f$' to denote the forward value. The notation $P^b_u$ etc will be used to denote the backward value to be computed in step 3. Thus $f(S(2))$ is a random variable taking values as $P^f_{uu}$, $P^f_{ud}$ and $P^f_{dd}$.

**Step 3.** Compute

$$P^b_u = \frac{1}{R}\left(\hat{p}\ P^f_{uu} + (1-\hat{p})\ P^f_{ud}\right) = \frac{1}{1.05}\left[\left(\frac{2}{3}\times 0\right) + \left(\frac{1}{3}\times 0\right)\right] = 0$$
$$P^b_d = \frac{1}{R}\left(\hat{p}\ P^f_{ud} + (1-\hat{p})\ P^f_{dd}\right) = \frac{1}{1.05}\left[\left(\frac{2}{3}\times 0\right) + \left(\frac{1}{3}\times 7.80\right)\right] = 2.476.$$

**Step 4.** Evaluate

$$P_u^f = \text{Max}\,(K - u\ S(0),\ 0) = \text{Max}(80 - 88, 0) = 0$$
$$P_d^f = \text{Max}\,(K - d\ S(0),\ 0) = \text{Max}(80 - 76, 0) = 4.$$

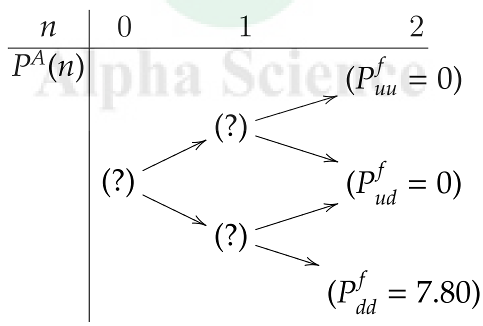
**Step 5.** Evaluate

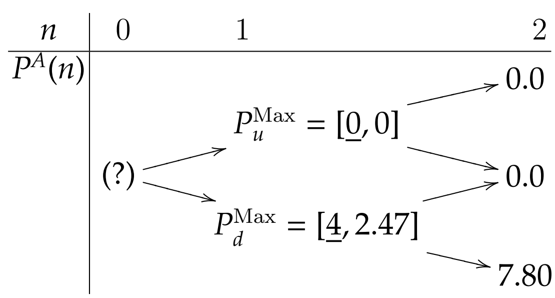
$$P^A(1) = \begin{cases} \text{Max}(P_u^f, P_u^b) = \text{Max}(0,0) = 0 = P_u^{Max} \\ \text{Max}(P_d^f, P_d^b) = \text{Max}(4, 2.476) = 4 = P_d^{\text{Max}}. \end{cases}$$

**Step 6.** Compute

$$\begin{aligned} P^A(0) &= \text{Max}\left(K - S(0),\ \frac{1}{R}\left[\hat{p}\ P_u^{\text{Max}} + (1-\hat{p})\ P_d^{\text{Max}}\right]\right) \\ &= \text{Max}\left((80-80), \frac{1}{1.05}\left[(\frac{2}{3} \times 0) + (\frac{1}{3} \times 4)\right]\right) \\ &= 1.27. \end{aligned}$$

The above calculation are depicted in the following tables which are now self explanatory



| $n$ | 0 | 1 | 2 |
|---|---|---|---|
| $P^A(n)$ | | | 0.0 |
| | | $P_u^{\text{Max}} = [\underline{0}, 0]$ | |
| | $[80 - 80, \underline{1.27}]$ | | 0.0 |
| | | $P_d^{\text{Max}} = [\underline{4}, 2.47]$ | |
| | | | 7.80 |

Here the notation $[a, \underline{b}]$ is used to denote that $b$ is the maximum of $a$ and $b$, i.e. $b = \text{Max}(a, b)$. Therefore $P^A(0) = 1.27$, and the given American put should be exercised early in the down state at $t = 1$. This will give a pay-off of $Rs$ 4.00 which is more than the value of holding it to the expiry for the value of $Rs$ 2.47. We may also compute $P^E(0)$ and get $P^E(0) = 0.79$. Clearly $P^E(0) < P^A(0)$. Hence these are different even for a non-dividend paying stock.

□

**Example 3.8.2** *Find the price of an American call and a European call for the data as* $S(0) = 120$, $K = 120$, $u = 1.2$, $d = 0.9$, $r = 10\%$ *per year and time to expiry* $T = 2$ *years.*

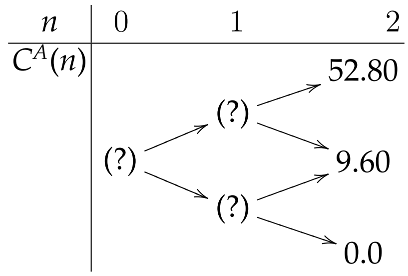


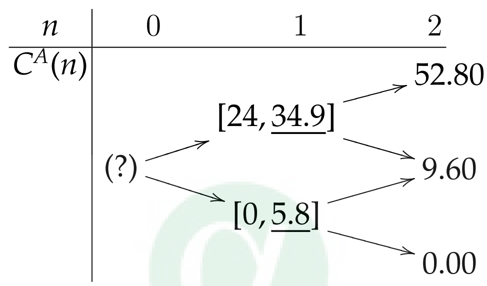
**Solution** As the stock is non-dividend paying, $C^A(0) = C^E(0)$. Computing $C^E(0)$ by formula (3.15) or utilizing the below given tables we get $C^E(0) = 22.92$. Here $\hat{p} = \dfrac{2}{3}$.

| $n$ | 0 | 1 | 2 |
|---|---|---|---|
| $S(n)$ | | | 172.80 |
| | | 144 | |
| | 12 | | 129.60 |
| | | 96 | |
| | | | 97.20 |

| $n$ | 0 | 1 | 2 |
|---|---|---|---|
| $C^E(n)$ | | | 52.80 |
| | | 34.91 | |
| | 22.92 | | 9.60 |
| | | 5.82 | |
| | | | 0.0 |

It may also be interesting to compute $C^A(0)$ as per the procedure outlined in this section and verify that $C^A(0)$ also comes out to be *Rs* 22.92. This calculation is shown in the below given tables




$$
\begin{aligned}
C^A(0) &= \text{Max}\left((120-120), \frac{1}{1.1}\left(\frac{2}{3}(34.9)+\frac{1}{3}(5.80)\right)\right)\\
&= \text{Max}(0, 22.92)\\
&= 22.92.
\end{aligned}
$$

At time $T = 2$, both $C^E$ and $C^A$ have the same pay-off. At $t = 1$, The given American option will not be exercised in the up state because it will bring only *Rs* 24 which is less than the value of holding it up to expiry. In the down state, the American call will be out of money and will not be exercised either.

□

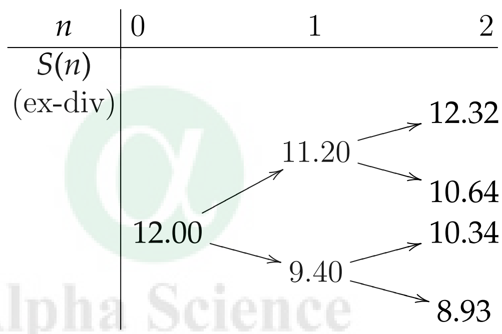
## 3.9 Options on Dividend Paying Stock

It has been remarked in the last section that for a non-dividend paying stock, $C^A(0)$ equals $C^E(0)$. But for a dividend paying stock, in general, $C^A(0)$ and $C^E(0)$ will have different value.

Let the stock pay a dividend of amount $Q$ at time $\tau$, $0 \leq \tau \leq T$. We wish to determine the price of a European call option on this stock, using the multi period binomial model. For this we divide the time interval $[0, T]$ into $N$ time periods. Assuming that the dividend date $\tau$ occurs in the $k^{th}$ period, we form a lattice in the usual way, but subtract the amount $Q$ from the nodes at the $k^{th}$ period. But then the nodes at time period $k$ do not recombine and therefore we get a binomial tree model rather than a binomial lattice model. Except for this change and some obvious modifications, the procedure for finding $P^E(0)$, $C^E(0)$, $P^A(0)$ and $C^A(0)$ remains same as discussed in Section 3.7 and Section 3.8. We illustrate the procedure in the below given exercise.

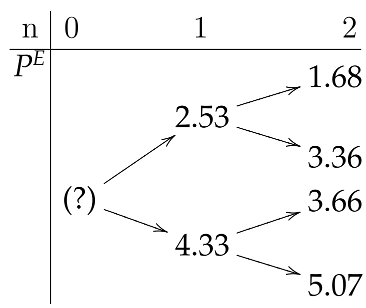
**Example 3.9.1** *Compute the price of the European put with the given data as* $S(0) = 12$, $K = 14$, $u = 1.1$, $d = 0.95$, $r = 2\%$ *per year and time to expiry in two years. Assume that a dividend of* Rs $2.00$ *is paid at time* $t = 1$.

**Solution** The ex-dividend price are as given in the below given table






Then following the methodology discussed in Section 3.8 we get $\hat{p} = \dfrac{7}{15}$ and




Here

$$2.53 = \frac{1}{1.02}\left[\frac{7}{15}(1.68) + \frac{8}{15}(3.36)\right]$$
$$4.33 = \frac{1}{1.02}\left[\frac{7}{15}(3.66) + \frac{8}{15}(5.07)\right].$$

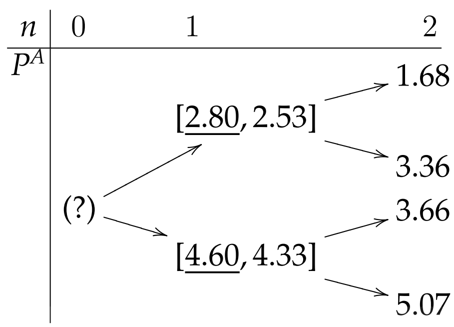
Therefore

$$P^E(0) = \frac{1}{1.02}\left[\frac{7}{15}(2.53) + \frac{8}{15}(4.33)\right] = 3.42.$$

□

**Example 3.9.2** *For the data given in Example 3.9.1, evaluate* $P^A(0)$.

**Solution** Continuing with the data of Example 3.9.1, we have




Therefore

$$\begin{aligned} P^A(0) &= \text{Max}\left(K - S(0), \frac{1}{1.02}\left(\frac{7}{15}(2.80) + \frac{8}{15}(4.60)\right)\right) \\ &= \frac{1}{1.02}\left[\frac{7}{15}(2.80) + \frac{8}{15}(4.60)\right] \\ &= 3.69. \end{aligned}$$

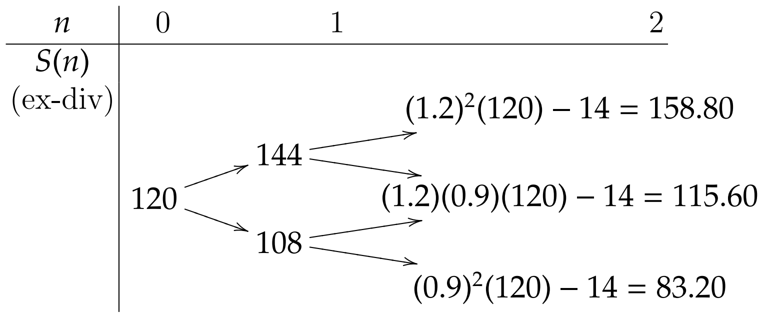


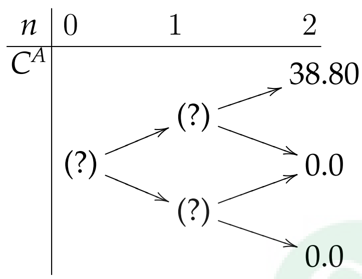
□

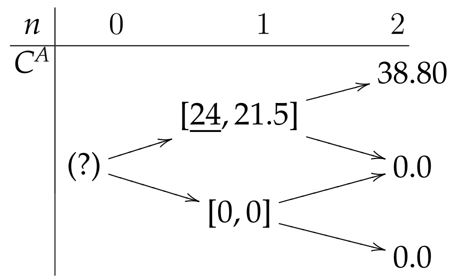
**Example 3.9.3** *Find the price* $C^E(0)$ *and* $C^A(0)$ *for the data,* $S(0) = 120$, $K = 120$, $u = 1.2$, $d = 0.90$, $r = 10\%$ *per year and time to expiry is* $T = 2$ *periods. What are the value of* $C^E(0)$ *and* $C^A(0)$ *if it is additionally given that a dividend of* $Rs$ 14 *is paid at* $T = 2$ *?*

**Solution** If there is no dividend on the stock then $C^A(0) = C^E(0)$. The value of $C^E(0)$ can be easily determined as $Rs$ 22.92. Here $\hat{p} = \dfrac{2}{3}$.

We next consider the case when there is a dividend of $Rs$ 14.00 at time $t = 2$. This gives the following table for the stock price at different periods.



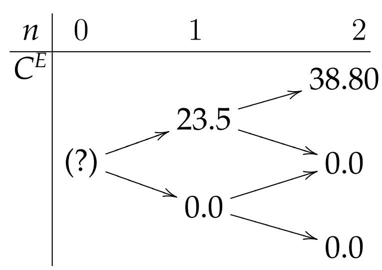




Therefore

$$C^A(0) = \frac{1}{1.1}\left[\frac{2}{3}(24) + \frac{1}{3}(0)\right]$$
$$= 14.54.$$



To compute $C^E(0)$ we follow the usual methodology to get the following table




This gives

$$C^E(0) = \frac{1}{1.1}\left[\frac{2}{3}(23.5) + \frac{1}{3}(0)\right] = 14.24.$$

As expected, here $C^E(0)$ and $C^A(0)$ are not equal because the stock is dividend paying.

□

## 3.10 Notion of Complete Markets

The concept of completeness property for a financial market is very important. This is because, if the market is complete then the unique RNPM exists and therefore every financial contract has unique fair price. In view of Theorems 3.7.1 and 3.7.2 a complete market has certainly to be arbitrage free.

In order to introduce the notion of completeness, we first introduce the notion of *contingent claim*. A *contingent claim* is a financial contract with random pay-off that can be positive or negative. The term 'contingent' reflects the situation that because the pay-off is random, it is contingent (i.e. dependent) on which state of nature is realized at the time of pay-off. All the securities that we have described earlier are contingent claims. Thus stock, bond, call, put, forward contract, futures etc are all contingent claims.

We say that a portfolio *replicates* a contingent claim when the pay-off of the portfolio matches the pay-off the claim in all possible states. Here 'matching' has to be understood in the sense that 'matches with probability one'. Such a portfolio is called a *replicating portfolio*. Here it must be added that we require replicating portfolio strategy to be *self financing*, i.e. the investor can not use funds other than the initial wealth to finance the position in the market and he/she is not allowed to spend money outside market. We say that the market is complete when it is arbitrage free and we are able to uniquely replicate any contingent claim with the existing securities.

Some readers must have realized that single and multi-period binomial lattice models are complete market models as they are capable of pricing any contingent claim. But what about a trinomial model? Here we have three possible outcomes for the price of stock in the next period which are: it goes up, it does not change or it goes down. Thus if $S(0)$ is the price of the stock at $t = 0$ then $S(1)$ takes three values $uS(0)$, $S(0)$ and $dS(0)$ with probabilities $p_1$, $p_2$ and $p_3$ respectively. The trinomial model is certainly arbitrage free because we have again assumed that $u > R > d$. But what about the existence of replicating portfolio? Let $C$ denote European call with strike price $K$ and expiry $t = 1$. Then the portfolio $(x, y)$ will replicate the pay-off of the call if

$$\begin{aligned} xuS(0) + yB(1) &= C_u \\ xS(0) + yB(1) &= C_n \\ xdS(0) + yB(1) &= C_d, \end{aligned}$$

where $C_u = \text{Max}(uS(0) - K, 0)$, $C_n = \text{Max}(S(0) - K, 0)$ and $C_d = \text{Max}(dS(0) - K, 0)$.

We see that the above linear system only permits a solution for a select subset of contingent claims (e.g. European calls here) with pay-offs within a 2-dimensional subspace. Some readers may like to verify that the linear system under consideration is consistent only when

$$(1-d)C_u + (u-1)C_d - (u-d)C_n = 0.$$

Thus not all contigent claims can be priced by the trinomial model. This means that the trinomial model is not complete.

In fact if we try to determine the unique RNPM for the trinomial model we shall again face the problem. Here we need to find $\hat{p}_1 > 0,\ \hat{p}_2 > 0,$ and $\hat{p}_3 > 0$ such that

$$\hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 1,$$

and

$$\hat{p}_1\,(uS(0)) + \hat{p}_2 S(0) + \hat{p}_3\,(dS(0)) = RS(0).$$

Obviously this system is consistent but it cannot have unique solution. This again shows that the trinomial model is not complete.

To tackle such situation we enlarge the set of basic securities. For example, we may take bond, stock and (say) an option whose price $\mathsf{C}(0)$ is known. Now we wish to obtain $\hat{p}_1,\ \hat{p}_2,\ \hat{p}_3\ \ (\hat{p}_i > 0,\ i = 1,2,3)$ such that

$$\begin{aligned} \hat{p}_1 + \hat{p}_2 + \hat{p}_3 &= 1 \\ \hat{p}_1\,(uS(0)) + \hat{p}_2\,(S(0)) + \hat{p}_3\,(dS(0)) &= RS(0) \\ \hat{p}_1 C_u + \hat{p}_2 C_n + \hat{p}_3 C_d &= RC(0). \end{aligned}$$

The above system has unique solution giving the unique RNPM. This unique RNPM can now be used to price other contingent claims. In this way we have artificially made the trinomial model as a complete market model.

**Example 3.10.1** *Let $S(0) = 100$ and $S(1)$ takes three values $120,\ 100$ and $90$ with probabilities $0.4, 0.2$ and $0.4$ respectively. The price of the European call on this stock with the strike price $105$ is $5$ and the price of the European call with strike price $95$ is $10$, with both option expiring at $t = 1$. Replicate the risk-free security that pays Rs $1$ at $t = 1$ regardless of what happens.*

**Solution** Let $x$ denote the number of shares of stock, $z_1$ the number of calls with strike price 100 and $z_2$ the number of calls with strike price $95$. Therefore in order to replicate the risk-free security we need to solve

$$120x + 15z_1 + 25z_2 = 1$$
$$100x + 0z_1 + 5z_2 = 1$$
$$90x + 0z_1 + 0z_2 = 1.$$

The solution of the above system is $x = 1/90$, $z_1 = 2/135$ and $z_3 = -1/45$. Further the cost of this portfolio is $(1/90)100 + (2/135)5 - (1/45)10$, i.e. $0.963$.

Therefore the price of risk-free security is $0.963$. Since its relative return is $(1/0.963) - 1 = 3.85\%$, the risk-free rate in this model is $3.85\%$.

□

## 3.11 Summary and Additional Notes

- Certain basic definitions and preliminaries with regard to European call/put options are presented in Section 3.2. An interesting and useful result of this section is the put- call parity. Here it is also mentioned that there is no such exact put-call parity for American options, but only certain estimates are provided for $C^A(0) - P^A(0)$.
- Section 3.3 is devoted to the analysis of the behavior of option prices with respect to defining variables $K$, $S$ and $T$. Pay-off curves/gain curves of option combinations is the topic of discussion in Section 3.4. In particular, we have introduced bull spread, bear spread, and butterfly spread; and also a general procedure to draw spread of a given option combination.
- The core of this chapter has been discussion on single and multi period binomial lattice models for option pricing. The single period case is discussed in Section 3.5, while multi period case is presented in Section 3.6. These are very important topics on option pricing because they form the basis of the elaborated Black-Scholes formula to be discussed in the next chapter.
- An important question about the existence of RNPM is answered in Section 3.7. Here the readers are able to see the application of the linear programming duality in proving that under no arbitrage principle RNPM certainly exists. Another important aspect of this section is to motivate the readers to study martingales so as to be ready for detailed study in a later chapter.
- Section 3.8 discusses the applicability of binomial lattice approach for pricing American options. Here, unlike the case of European option, we cannot have any closed form pricing formula and backward induction is the only recourse.
- The case of dividend paying stock is considered in Section 3.9 and results of Section 3.5 and 3.6 are restated in this new scenario.

- Most of the material presented in this chapter is motivated by Capinski and Zastawrick [25], and Luenberger [85]. The readers are encouraged to read 'Summary and additional notes' section of the next chapter (Chapter 4) to be familiar with certain recent topics on binomial lattice models for option pricing.
- The globalization (business anywhere in world) and democratization (information advantage) of finance has led to designing more new products which offer a wide array of investment choices to investors. The European and the American calls and puts are the simplest type of investment options, therefore they are sometimes referred to as "plain vanilla options". There are however several other sophisticated options available in the market. They are referred as *exotic options*. In the earlier years, exotic options were traded over the counter (OTC) and not in exchanges; but more recently some of these options can be found on exchange markets too. The prominent among many exotic options are *Asian options, barrier options, basket options, digital (or binary) options, 'as you like it' (or chooser) options, look back options, rainbow options*, to name a few. These options have different characteristic and execution rules. Pricing these options is rather hard and requires advance tools from Econometric and Mathematics. This concept is therefore not touched here in this book. But we sincerely encourage readers to explore the sea of exotic options, pricing, and how they can be used in hedging risk in different situations; one can refer to [65, 87, 145].



## 3.12 Exercises

**Exercise 3.1** *Let $S(T)$ be the price of a given security (stock) at time $T$. All of the following options have exercise time $T$ and (unless stated otherwise) strike price $K$. Give the pay-off at time $T$ that is earned by an investor who*

*(i) owns one call and one put option.*
*(ii) owns two calls and has sold short one share of stock.*
*(iii) owns one share of stock and has sold one call.*
*(iv) owns one call having strike price $K_1$ and has sold one put having strike price $K_2$.*

**Exercise 3.2** *Consider a family of call options on a non-dividend paying stock, each option being identical except for its strike price. The value of the call with strike price $K$ is denoted by $C(K)$. Let $K_1 < K_2 < K_3$. Show that*

$$C(K_2) \leq \left(\frac{K_3 - K_2}{K_3 - K_1}\right) C(K_1) + \left(\frac{K_2 - K_1}{K_3 - K_1}\right) C(K_3).$$

**Exercise 3.3** *In a binomial lattice model, let possible values of* $S(2)$ *be* $32, 28$ *and* $x$*. Find the value of* $x$*.*

**Exercise 3.4** *Draw the pay-off curves of the following option portfolios*

*(i)* $-P_{80} + P_{100} + 2C_{130} - C_{150}$.
*(ii)* $-P_{100} + P_{120} + C_{150}$.

**Exercise 3.5** *Let* $B(0) = 100, B(1) = 110, S(0) = 100$ *and* $S(1)$ *take values* $120$ *and* $80$ *with probabilities* $0.8$ *and* $0.2$ *respectively. Also let* $C^E$ *and* $P^E$ *respectively be a European call and a European put with* $K = 100$ *and* $T = 2$.

*(i) Verify put call pairty.*
*(ii) What can you say if the options are American options* $C^A$ *and* $P^A$*?*
*(iii) What difficulty (if any) will be faced if instead of* $B(1) = 110$ *we have* $B(1) = 120$*? Give mathematical justification to your answer.*

**Exercise 3.6** *Let* $r = 0.2$*. Find the risk neutral conditional expectation of* $S(3)$ *given that* $S(2) = Rs\ 110$.

**Exercise 3.7** *A certain stock is selling for* $Rs\ 50$*. The feeling is that for each month, for the next two months, the stock price will rise by* $10\%$ *or fall by* $10\%$*. Assuming that the risk free rate is* $1\%$*, calculate the price of the European call with the strike price of* $Rs\ 48$.

**Exercise 3.8** *Prove the following*

*(i)* $S(0) - Ke^{-rt} - D(0) \leq C^E(0)$
*(ii)* $Ke^{-rt} + D(0) - S(0) \leq P^E(0)$,

*where* $D(0)$ *is the present value of the dividend paid by the stock. The rest of terminologies stand for usual meanings.*

**Exercise 3.9** *Consider the data* $S(0) = 60, K = 62, u = 1.1, d = 0.95, r = 0.03$ *and* $T = 3$*. Find* $C^E(0), P^E(0), C^A(0)$ *and* $P^A(0)$*. Identify the time instants when* $C^A$ *and* $P^A$ *will be exercised.*

**Exercise 3.10** *Let* $S(0) = 120, u = 1.2, d = 0.9$ *and* $r = 1\%$*. Consider a call option with strike price* $K = 120$ *and* $T = 2$*. Find the option price and the replicating strategy.*

**Exercise 3.11** *For the single period trinomial model, let* $S(0) = 100$ *and* $S(1)$ *take three values* $120$, $100$ *and* $90$ *with probabilities* $0.4, 0.2$ *and* $0.4$ *respectively. Let* $r = 3.85\%$*. Further let the price of the European call on this stock with the strike price* $105$ *be* $5$*. Find the price of the European call with the strike price* $95$*.*

# 4
# Basic Theory of Option Pricing-II

## 4.1 Introduction

The world of options underwent a revolutionary change in 1973 when Fisher Black and Maryon Scholes [15] and Robert C. Merton [94] published their seminal papers on theory of option pricing. The basic idea behind their studies is that the price of an option is determined implicitly by the price of the underlying stock and certain other parameters whose values, except one of them, are easily observable. Black-Scholes option pricing model constructed a continuously hedging strategy to protect the writer's short position in option. Also Merton designed a self-financing and dynamically replicating hedged position containing the options, the underlying risky stock and a riskless asset. Both approaches led to a partial differential equation, now famously known as the Black-Scholes partial differential equation or simply Black-Scholes equation, with appropriate boundary conditions defined according to the contractual specifications of the option. These models will be presented and discussed later in Chapter 10. The price of a particular derivative security or option is obtained by solving the Black-Scholes equation subject to boundary conditions. In recognition to their pioneering contributions, Scholes and Merton were awarded 1997 Nobel prize in Economics. Black's contribution was also recorded by the Swedish Academy though he was found ineligible for prize as he died in 1995.

Despite its richness and simplicity, the Black-Scholes model has one major limitation. It cannot be used to accurately price options with an American-style exercise options as it only calculates the option price at expiration. It does not consider the steps along the way where there could be the possibility of early exercise of an option. To this end, came the alternative proposal of using binomial option pricing model. Both the Black-Scholes model and the binomial model are based on the same theoretical foundations and assumptions but there are also some important differences between the two. The mathematical tools Black, Scholes and Merton employed are quite advanced, like solving a partial differential equation, and this fact does not make the underlying economics more clear. In contrast to the Black-Scholes model the binomial option-pricing model is simple and can easily be understood with undergraduate knowledge. Partly the same has already been experienced in previous chapter wherein the readers have been provided with the glimpse of multi period binomial lattice models for option pricing through simple two period examples.

As already exhibited in the previous chapter, the guiding principle of the binomial model is to break down the time to expiration into a large number of time intervals or steps. Thereafter, at each step, it is assumed that the stock price will move up or down by certain amounts. This produces a binomial lattice of underlying stock prices. Subsequently the option prices at each node of the lattice are calculated, working backward from expiration to the present. For almost all practical purposes, in practice the discrete model of option pricing is preferred.

In this chapter, we aim to present the binomial lattice model proposed by John C. Cox, Stephen A. Ross and Mark Rubinstein [33] in 1979 for option pricing. This model is now famously called the *CRR model* for option pricing. Quoting Cox, Ross and Rubinstein, "Our formulation, by its very construction, leads to an alternative numerical procedure which is both simpler, and for many purposes, computationally more efficient". In fact the Black-Scholes formula turns out to be a particular limiting case of the discrete binomial CRR model. In Section 4.2, we describe the construction of the CRR model. In Section 4.3, we present the case when the CRR model can be matched with the multi-period binomial model. This matching finally leads to the Black-Scholes formula in the limiting case. Section 4.4 presents an extension of the Black-Scholes formula for options on dividend paying underlying stocks. Section 4.5 briefly describes the parameters, called the Greeks, which are vital tools in risk management.

## 4.2 The CRR Model

We illustrate how the knowledge gained in previous chapter on a single period binomial model and the strategy of replication can be used to extend the option pricing theory to multi-period scenario. The assumptions on the financial market made for single period binomial model are carried forward here too. For the sake of completeness, let us recall them as follows.

(i) The underlying stock on which option is written is perfectly divisible.

(ii) The underlying stock pays no dividend.
(iii) There is no transaction costs in buying or selling the option, and no taxes.
(iv) Short selling is allowed.
(v) The riskless interest rate is known and constant till time to expiration of option.
(vi) No arbitrage principle holds.

Let us now divide the time horizon interval $[0, T]$ into $n$ subintervals each of length $\Delta t = T/n$, and assume that in each subinterval the stock price changes like a one period binomial case. Thus in each period the underlying stock price either moves up by a constant factor $u$ or moves down by a constant factor $d$ with positive probability. For $k = 1, \ldots, n$, we also define

$$E_k = \begin{cases} u, & \text{with probability } p \\ d, & \text{with probability } 1 - p, \end{cases}$$

with $0 < p < 1$. Here we note that $E_k$ is a Bernoulli random variable.

At time to expiration $T$, the stock price $S(T)$ is given by

$$\begin{aligned} S(T) &= S(0) E_1 E_2 \cdots E_n \\ &= S(0)\, e^H, \qquad \text{(say)} \end{aligned}$$

where $e^H = E_1 E_2 \cdots E_n$. Thus,

$$ln(S(T)) = ln(S(0)) + H; \qquad H = \sum_{k=1}^{n} ln(E_k).$$

Here, $H$ represents the logarithmic growth of the stock price.

The above expression helps us to visualize it as a discrete approximation to the continuous price process. The same is shown in Figure 4.1.

We recollect that if a random variable $X$, defined as

$$X = \begin{cases} a, & \text{with probability } p \\ b, & \text{with probability } 1 - p, \end{cases}$$

the expectation of $X$ is $E(X) = ap + b(1 - p)$ and the variance of $X$ is $Var(X) = E(X^2) - (E(X))^2 = p(1 - p)(a - b)^2$. Using this we can easily derive that

$$E(ln(E_k)) = \; pln(u) + (1 - p) ln(d)\,, \tag{4.1}$$

$$Var(ln(E_k)) = p(1 - p)(ln(u) - ln(d))^2. \tag{4.2}$$

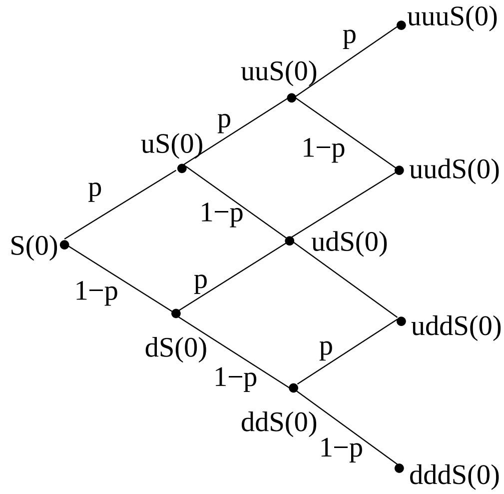



**Fig. 4.1.** The underline binomial lattice model

We now introduce two new parameters $\mu$ and $\sigma^2$ as follows

$$\mu \Delta t = E(ln(E_k)) \quad \text{and} \quad \sigma^2 \triangle t = Var(ln(E_k)). \tag{4.3}$$

The parameter $\mu$ is called the *drift* and the parameter $\sigma$ is called the *volatility*. It may appear a little strange as to how and why we suddenly bring in two new parameters $\mu$ and $\sigma^2$. As of now we urge to take them on face value as we shall be providing their physical interpretation later in this chapter. Now for each $k = 1, \ldots, n$, let us introduce another random variable

$$X_k = \frac{ln(E_k) - E(ln(E_k))}{\sqrt{Var(ln(E_k))}}.$$

Then, by using (4.1), (4.2) and definition of $E_k$, we can easily work out that

$$X_k = \begin{cases} \dfrac{1-p}{\sqrt{p(1-p)}}, & \text{with probability } p \\ \dfrac{-p}{\sqrt{p(1-p)}}, & \text{with probability } 1-p. \end{cases}$$

Clearly, for each $(k = 1, \ldots, n)$, $E(X_k) = 0$ and $Var(X_k) = 1$. Now,

$$\begin{aligned} H &= \sum_{k=1}^{n} ln(E_k) \\ &= \sum_{k=1}^{n} (\mu \triangle t + \sigma \sqrt{\triangle t} X_k) \\ &= \mu T + \sigma \sqrt{\triangle t} Y, \end{aligned}$$

where

$$Y = \sum_{k=1}^{n} X_k, \tag{4.4}$$

is a simple random walk. Though the notion of simple random walk will be explained in details in Chapter 7, here we just wish to get familiar with it for clarity.

**Definition 4.2.1 (Random Walk)** *A random walk is a stochastic process* $\{S_n,\ n = 0, 1, \ldots\}$*, with* $S_0 = 0$*, defined by*

$$S_n = \sum_{k=1}^{n} X_k\,,$$

*where* $\{X_k\}$ *are independent and identically distributed random variables. The random walk is simple if for each* $k = 1, \ldots, n$*,* $X_k$ *takes value from* $\{a, b\}$ *only,* $a$ *and* $b$ *are real constants, with* $P(X_k = a) = p$ *and* $P(X_k = b) = 1 - p$*,* $p \in [0, 1]$.

It is important to observe that the constants $a$, $b$, $p$ are independent of $k$. Generally $a$ and $b$ are taken as 1 and $-1$ respectively. A simple random walk can be visualized as a game involving two persons, which consists of a sequence of independent identically distributed moves, and the sum $S_n$ represents the score of the first person, say, after $n$ moves, with the assumption that the score of the second person is $-S_n$.

In view of the above discussion we have the following lemma.

**Lemma 4.2.1** *For a CRR model with probability of up tick* $u$ *equals* $p$ *and probability of down tick* $d$ *equals* $1 - p$*, life time* $T$ *and time increment* $\triangle t = T/n$*, the stock price is given by*

$$S(T) = S(0)\ \exp\left(\mu T + \sigma \sqrt{\Delta T}\ Y\right),$$

*where* $\mu$ *is the drift and* $\sigma$ *is the volatility described by (4.3) and* $Y$ *is a simple random walk given by (4.4).*

We calculate expectation and variance of the random variable $H$.

$$\begin{aligned} E\Big(ln\Big(\frac{S(T)}{S(0)}\Big)\Big) &= E(H) \\ &= E(\mu T + \sigma\sqrt{\triangle t}Y) \\ &= \mu T + \sigma\sqrt{\triangle t}\sum_{k=1}^{n} E(X_k) \\ &= \mu T \qquad (\text{as } E(X_k) = 0,\ \forall\ k)\,, \end{aligned}$$

and

$$\begin{aligned} Var\Big(ln\Big(\frac{S(T)}{S(0)}\Big)\Big) &= Var(H) \\ &= Var(\mu T + \sigma\sqrt{\triangle t}Y) \\ &= \sigma^2\triangle t\sum_{k=1}^{n} Var(X_k) \\ &= \sigma^2 T \qquad (\text{as } Var(X_k) = 1,\ \forall\ k)\,. \end{aligned}$$

The above two relations made us ready to give interpretations of the two parameters - the drift and the volatility. Suppose we have $n$ quotes of stock prices per year. A typical example is the daily closing stock prices for $T = 1$ year (or $252$ trading days per year), so in this case $n = 252$ and $\triangle t = 1$ day. We consider the random stock prices, initially $S(0) = S_0$ and on subsequent days $S_1, S_2, \ldots, S_{251}$ and finally at the end of one year $S_{252} = S(T) = S(1)$. As an investor we are interested in the yearly return $\frac{S(1)}{S(0)}$, or its logarithm $ln\Big(\frac{S(1)}{S(0)}\Big)$. Note that the logarithmic returns of the daily observed stock prices are $\{ln(S_1) - ln(S_0), ln(S_2) - ln(S_1), \ldots, ln(S_{252}) - ln(S_{251})\}$. The sum of these daily returns yield $(ln(S_{252}) - ln(S_0))$, which is the logarithmic yearly return. This simple characteristic is not available in decimal returns $\{\frac{S_1 - S_0}{S_0}, \frac{S_2 - S_1}{S_1}, \ldots, \frac{S_n - S_{n-1}}{S_{n-1}}\}$. By using the logarithmic prices we can convert an exponential problem to a linear problem; the logarithmic returns can simply be added. For this simple reason the logarithmic yearly returns are often used. What we observe is that the drift $\mu$ is the expectation of the logarithmic return and the volatility $\sigma$ is the standard deviation of the logarithmic return.

The volatility is the most critical of all option trading concepts. It provides investors with an estimate of how much movement a stock can be expected to make over a given time frame. Two types of volatilities are generally talked about - historical and implied. At present we shall not be talking about the implied volatility as it requires other conceptual knowledge which we have not yet acquired. (Please refer to the third point in the notes, Section 4.6).

Also, there are many sophisticated models in financial econometric for computing historical volatility, like ARCH (autoregressive conditional heteroskedasticity), GARCH (generalized ARCH), risk metrics EWMA (equally weighted moving average), threshold autoregressive model, and so on so forth. Again we are not in a position to elaborate on them. However, for the sake of understanding, we briefly explain a very basic model for computing the historical volatility. First calculate the natural log of the ratio of a stock price from the current day $t$ to the previous day $t-1$, that is, $R_t = ln(S_t/S_{t-1})$. Then find the average of all these for the given period $n$ (for example, $n = 20$ days), say, $\overline{R} = (\sum_{t=1}^{n} R_t)/n$. The historical volatility is the standard deviation of $R_t$, $(t = 1, \ldots, n)$, from the mean $\overline{R}$. If a stock price follows a Brownian motion or a geometric Brownian motion then the above result is multiplied by the square root of the average number of trading days in a year to quote the volatility on an annual basis. The reason for the last step lies in the fact that the volatility increases with the square root of the unit of time. Though we agree that one may find this discussion a bit incomplete, but the details have been avoided just to skip the complications involved with volatility. However, we encourage the readers to search the web and find sufficient material/examples on this concept or refer to an excellent text [138].

We shall next attempt to approximate the CRR model by a multi-period binomial model. It is important to point out here that not every CRR model is a multi-period binomial model.



## 4.3 Matching of CRR Model with a Multi-Period Binomial Model

The primary question is to find the values of $u$, $d$, and $p$ for given values of $\mu$ and $\sigma^2$ from (4.1), (4.2), (4.3). It is worth to mention that, in the conclusive analysis, the value of $\mu$ will interestingly not be required. However, it is the actual value of $\sigma$ which is rather hard to capture. Nevertheless, we assume that $\sigma$ is known from some past data of the stock (i.e. say historical volatility).

Now, in lieu of (4.3) we have two relations (4.1) and (4.2) to determine three parameters $u$, $d$, $p$. In order to describe these parameters uniquely we need a third condition consistent with the above two relations. To meet this need, let us denote by $U = ln(u)$, $D = ln(d)$, and assume that $D = -U$. In other words, we shall be matching the CRR model with a particular multi-period binomial model in which $d = \frac{1}{u}$. Now, (4.1) and (4.2), on using (4.3), can be re-expressed as follows

$$\begin{aligned} (2p-1)U &= \mu \triangle t \\ 4p(1-p)U^2 &= \sigma^2 \triangle t. \end{aligned} \tag{4.5}$$

Simplifying the above equations, we get

$$(2p-1)^2U^2 + 4p(1-p)U^2 = \mu^2(\triangle t)^2 + \sigma^2 \triangle t,$$

that is,

$$U = \sqrt{\mu^2(\triangle t)^2 + \sigma^2 \triangle t}, \text{ and } D = -U.$$

Furthermore, (4.5) yields

$$p = \frac{1}{2} + \frac{\mu \triangle t}{2\sqrt{\mu^2(\triangle t)^2 + \sigma^2 \triangle t}}.$$

Let us take $n$ sufficiently large (equivalently, the time steps between successive trading instances approach zero) so that $(\triangle t)^2$ terms can be neglected. We get

$$U = ln(u) = \sigma\sqrt{\triangle t} \quad \text{and} \quad D = -U = -\sigma\sqrt{\triangle t}.$$

Consequently, we have

$$u = e^{\sigma\sqrt{\triangle t}}, \quad d = e^{-\sigma\sqrt{\triangle t}} \quad \text{and } p = \frac{1}{2} + \frac{\mu}{2\sigma}\sqrt{\triangle t}. \tag{4.6}$$

We take a small break here for some examples to illustrate the choice of the parameters.

**Example 4.3.1** *A non-dividend paying stock is currently selling at Rs* 100 *with annual volatility* 20%. *Assume that the continuously compounded risk-free interest rate is* 5%. *Using a two-period CRR binomial option pricing model, find the price of one European call option on this stock with a strike price of Rs* 80 *and time to expiration 4 years.*

**Solution** We are given $S(0) = Rs\ 100$, $K = Rs\ 80$, $T = 4$, $\triangle t = 2$, $r = 0.05$, $\sigma = 0.2$. Applying the CRR model, the up-factor and the down-factor are respectively given by

$$u = e^{\sigma\sqrt{\triangle t}} = 1.3269, \quad d = \frac{1}{u} = 0.7536.$$

Now,

$$S_{uu} = u^2 S(0) = 176.0664,\ S_{ud} = udS(0) = 100.00,\ S_{dd} = 56.7913.$$

Thus,

$$\begin{aligned} C_{uu} &= \text{Max}\{S_{uu} - K, 0\} = 96.0664, \\ C_{ud} &= \text{Max}\{S_{ud} - K, 0\} = 20.00, \\ C_{dd} &= \text{Max}\{S_{dd} - K, 0\} = 0. \end{aligned}$$

Since the annual rate of risk-free interest is continuously compounded (so instead of $1 + r$ or $R$ we shall use $e^{r\Delta t}$), the risk neutral probability is

$$\hat{p} = \frac{e^{r\Delta t} - d}{u - d} = 0.6132.$$

Hence for a two period binomial model,

$$\begin{aligned} C(0) &= e^{-r\Delta t}((\hat{p})^2 C_{uu} + 2\hat{p}(1 - \hat{p})C_{ud} + (1 - \hat{p})^2 C_{dd}) \\ &= e^{-0.1}((0.3760)(96.0664) + (0.4744)(20.00) + 0) \\ &= \text{Rs } 41.27. \end{aligned}$$

□

For computational purpose, we take the annualized risk-free interest rate and volatility. Although we take 365 days for counting a year but the exchanges across the world remain closed and no trading take place on Saturday and Sunday. Moreover, there are some other holidays observed by exchanges. Consequently the number of trading days are less. In practice options are priced taking into account the number of trading days. The following example throw light on this aspect.

**Example 4.3.2** *A non-dividend paying stock is selling at Rs* $1,500$ *on March 1, 2010, with annual volatility of* $22\%$*. Assume that the continuously compounded risk-free interest rate is* $3\%$*. Compute the price of a European call option written on this stock struck at Rs* $1470$ *expiring in April 2010 using a single-period CRR binomial model.*

**Solution** It is given that $S(0) = Rs\ 1,500,\ K = Rs\ 1,470,\ r = 0.03,\ \sigma = 0.22$. The option is written on March 1 and it expires in the business day immediately preceding the last business day of the contract month which is April 29, 2010. The total number of trading days are 44 days. The number of trading days in 2010 is 252. Thus, $T = 44/252 = 0.1746$. For a single-period model, $\Delta t = T = 0.1746$. Then, $u = e^{\sigma\sqrt{\Delta t}} = 1.0963,\ d = u^{-1} = 0.9124$. The risk neutral probability measure is $\hat{p} = \dfrac{e^{r\Delta t} - d}{u - d} = 0.505$. Thus,

$$C(0) = e^{-rT}(\hat{p}C_u + (1-\hat{p})C_d) = (0.9948)((0.505)174.45 + (0.495)0) = Rs\ 87.64.$$

□

However, for text purpose we continue to take 365 days or 12 months in a year. This may lead to calculation differences between the theoretical value, when applied on a real data, and the actual financial quotation (available in newspapers or www) for the option. The readers are cautioned.

Continuing with our previous discussion, we next define a counter on up tick and down tick movements on stock price at time $k\triangle t,\ (k = 1, \ldots, n)$, as a Bernoulli random variable

$$Y_k = \begin{cases} 1, & \text{with probability } p \text{ if stock goes up} \\ 0, & \text{with probability } 1-p \text{ if stock goes down.} \end{cases}$$

Then,

$$\begin{aligned} S(n\triangle t) &= S(T) \\ &= S(0)u^{\sum_{k=1}^{n} Y_k} d^{(n-\sum_{k=1}^{n} Y_k)} \\ &= d^n S(0)\Big(\frac{u}{d}\Big)^{\sum_{k=1}^{n} Y_k}, \end{aligned}$$

giving

$$\frac{S(T)}{S(0)} = d^{\frac{T}{\triangle t}} \Big(\frac{u}{d}\Big)^{\sum_{k=1}^{T/\triangle t} Y_k}.$$

Observe that $Y = \sum_{k=1}^{T/\triangle t} Y_k$ is a simple random walk with $E(Y) = p(T/\triangle t)$ and $Var(Y) = p(1-p)(T/\triangle t)$.

Using the value of $ln(d)$ from (4.6), we have

$$ln\Big(\frac{S(T)}{S(0)}\Big) = \frac{-T\sigma}{\sqrt{\triangle t}} + 2\sigma\sqrt{\triangle t}\sum_{k=1}^{T/\triangle t} Y_k.$$

Hence,

$$\begin{aligned} E\Big(ln\Big(\frac{S(T)}{S(0)}\Big)\Big) &= \frac{-T\sigma}{\sqrt{\triangle t}} + 2\sigma\sqrt{\triangle t}\sum_{k=1}^{T/\triangle t} E(Y_k) \\ &= \frac{-T\sigma}{\sqrt{\triangle t}} + 2\sigma\sqrt{\triangle t}\frac{T}{\triangle t}p \\ &= \mu T, \end{aligned}$$

where the last equality follows on using the value of $p$ from (4.6). Also,

$$Var\Big(\ln\Big(\frac{S(T)}{S(0)}\Big)\Big) = 4\sigma^2 \triangle t \sum_{k=1}^{T/\triangle t} Var(Y_k)$$
$$= 4\sigma^2 Tp(1-p)$$
$$\to \sigma^2 T, \text{ as } p \to \frac{1}{2} \text{ when } n \to \infty.$$

Furthermore, by application of the central limit theorem, we can assume that $Y_k$ follows a normal distribution when time steps approach zero. Summarizing the above discussion we can conclude that

$$ln\Big(\frac{S(T)}{S(0)}\Big) \sim \mathcal{N}(\mu T, \sigma^2 T).$$

Moving ahead, observe that

$$u = e^{\sigma\sqrt{T/n}} \cong 1 + \sigma\sqrt{\frac{T}{n}} + \frac{\sigma^2 T}{2n} \tag{4.7}$$

$$d = e^{-\sigma\sqrt{T/n}} \cong 1 - \sigma\sqrt{\frac{T}{n}} + \frac{\sigma^2 T}{2n}. \tag{4.8}$$

So, the risk neutral probability measure (RNPM) is given by

$$\hat{p} = \frac{R-d}{u-d}$$
$$= \frac{1 + \frac{rT}{n} - d}{u-d}$$
$$\cong \frac{\frac{rT}{n} + \sigma\sqrt{\frac{T}{n}} - \frac{\sigma^2 T}{2n}}{2\sigma\sqrt{\frac{T}{n}}}.$$

Thus,

$$\hat{p} = \frac{1}{2} + \frac{2r - \sigma^2}{4\sigma}\sqrt{\frac{T}{n}}. \tag{4.9}$$

We have seen that the European call or put options are simple to price. We therefore concentrate on pricing the European call option using the above described scheme. As established in previous chapter, the European call option price for an $n$-period binomial lattice model is described as follows.

$$\begin{aligned} C(0) &= \Big(1 + \frac{rT}{n}\Big)^{-n} E_{\hat{p}}((S(T) - K)_+) \\ &= \Big(1 + \frac{rT}{n}\Big)^{-n} E_{\hat{p}}((S(0)u^Y d^{n-Y} - K)_+) \\ &= \Big(1 + \frac{rT}{n}\Big)^{-n} E_{\hat{p}}((S(0)\Big(\frac{u}{d}\Big)^Y d^n - K)_+). \end{aligned}$$

It follows from (4.7) and (4.8), that

$$C(0) = \Big(1 + \frac{rT}{n}\Big)^{-n} E_{\hat{p}}((S(0)e^w - K)_+), \text{ where } w = 2\sigma\sqrt{\frac{T}{n}}Y - \sigma\sqrt{nT}. \tag{4.10}$$

It is to be noted here that

$$\begin{aligned} E(w) &= 2\sigma\sqrt{\frac{T}{n}}E(Y) - \sigma\sqrt{nT} \\ &= 2\sigma\sqrt{\frac{T}{n}}np - \sigma\sqrt{nT} \\ &= (2p-1)\sigma\sqrt{nT} \\ &= \mu\sqrt{\triangle t}\sqrt{nT} \\ &= \mu T, \end{aligned}$$



where the second last equality follows by virtue of first relation in (4.5) and the value of $U = \sigma\sqrt{\triangle t}$. Also, using that $Var(Y) = p(1-p)(T/\triangle t)$ and $n\triangle t = T$, we have

$$\begin{aligned} Var(w) &= 4\sigma^2\frac{T}{n}Var(Y) \\ &= 4p(1-p)\sigma^2 T \\ &\to \sigma^2 T, \end{aligned}$$

where the last relation follows in the limiting case when $n \to \infty$ (or $\triangle t \to 0$), and $p \to \frac{1}{2}$ from (4.6). Moreover it is important to note that

$$\begin{aligned}
E_{\hat{p}}(w) &= 2\sigma\sqrt{\frac{T}{n}}E_{\hat{p}}(Y) - \sigma\sqrt{nT} \\
&= 2\sigma\sqrt{\frac{T}{n}}np^{*} - \sigma\sqrt{nT} \\
&= 2\sigma\sqrt{nT}\Big(\frac{1}{2} + \frac{2r-\sigma^2}{4\sigma}\sqrt{\frac{T}{n}}\Big) - \sigma\sqrt{nT} \qquad \text{(using (4.9))} \\
&= \Big(r - \frac{\sigma^2}{2}\Big)T
\end{aligned}$$

and

$$\begin{aligned}
Var(w)_{\hat{p}} &= 4\sigma^2\hat{p}(1-\hat{p})T \\
&\to \sigma^2 T,
\end{aligned}$$

$$\text{(using the limiting case when n} \to \infty \text{ and p} \to \tfrac{1}{2} \text{ in (4.9)).}$$

As a consequence of above relations, (4.10) gives

$$C(0) = e^{-rT}\int_{-\infty}^{\infty}(S(0)e^{w} - K)_{+}\frac{1}{\sigma\sqrt{2\pi T}}\, e^{-\frac{1}{2}\Big(\frac{w-(r-\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\Big)^2} dw.$$

Now $S(0)e^{w} > K$, implies $w > ln\Big(\frac{K}{S(0)}\Big) = w_1$. Therefore,

$$C(0) = e^{-rT}\int_{w>w_1}(S(0)e^{w} - K)\frac{1}{\sigma\sqrt{2\pi T}}\, e^{-\frac{1}{2}\Big(\frac{w-(r-\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\Big)^2} dw.$$

To evaluate this integral, substitute

$$y = \frac{w - \Big(r - \frac{\sigma^2}{2}\Big)T}{\sigma\sqrt{T}}.$$

Then,

$$w = y\sigma\sqrt{T} + \Big(r - \frac{\sigma^2}{2}\Big)T, \quad \text{so } dw = \sigma\sqrt{T}\,dy.$$

Moreover, $w > w_1$ gives $y > y_1$, where

$$y_1 = \frac{1}{\sigma\sqrt{T}}\Big(ln\Big(\frac{K}{S(0)}\Big) - \Big(r - \frac{\sigma^2}{2}\Big)T\Big).$$

Subsequently,

$$\begin{aligned} C(0) &= \frac{e^{-rT}}{\sqrt{2\pi}}\int_{y>y_1}(S(0)e^{y\sigma\sqrt{T}+(r-\frac{\sigma^2}{2})T} - K)e^{-\frac{y^2}{2}}\,dy \\ &= \frac{e^{-rT}}{\sqrt{2\pi}}\int_{y>y_1} S(0)e^{y\sigma\sqrt{T}+(r-\frac{\sigma^2}{2})T}e^{-\frac{y^2}{2}}\,dy - \frac{1}{\sqrt{2\pi}}Ke^{-rT}\int_{y_1}^{\infty} e^{-\frac{y^2}{2}}dy \\ &\equiv I - Ke^{-rT}\Phi(-y_1), \end{aligned} \tag{4.11}$$

where, $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}dz$ is the distribution function of a standard normal random variable and

$$\begin{aligned} I &= \frac{e^{-rT}}{\sqrt{2\pi}}\int_{y>y_1} S(0)e^{y\sigma\sqrt{T}+(r-\frac{\sigma^2}{2})T}e^{-\frac{y^2}{2}}\,dy \\ &= \frac{S(0)}{\sqrt{2\pi}}\int_{y>y_1} e^{-\frac{(y\ -\ \sigma\sqrt{T})^2}{2}}\,dy. \end{aligned}$$

To evaluate $I$ we define a new variable $s = y - \sigma\sqrt{T}$. Then for $y > y_1$, we have $s > s_1 = y_1 - \sigma\sqrt{T}$. Thus,

$$\begin{aligned} I &= \frac{S(0)}{\sqrt{2\pi}}\int_{s>s_1} e^{-\frac{s^2}{2}}\,ds \\ &= S(0)\Phi(-s_1) \\ &= S(0)\Phi\Big(\sigma\sqrt{T} - \frac{1}{\sigma\sqrt{T}}\Big(ln\Big(\frac{K}{S(0)}\Big) - \Big(r - \frac{\sigma^2}{2}\Big)T\Big)\Big) \\ &= S(0)\Phi(d_1), \end{aligned}$$

where

$$d_1 = \frac{ln\Big(\frac{S(0)}{K}\Big) + \Big(r + \frac{\sigma^2}{2}\Big)T}{\sigma\sqrt{T}}. \tag{4.12}$$

Summarizing the discussion, from (4.11), we have

$$C(0) = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{4.13}$$

where $d_2 = -y_1 = d_1 - \sigma\sqrt{T}$ and $d_1$ is as described in (4.12). The expression in (4.13) is known as the *Black-Scholes formula* for computing the European call option price under the set of assumptions described at the beginning of Section 4.2.

Observe that the expected rate of return $\mu$ on the stock does not appear in the Black-Scholes formula. Moreover all parameters except for $\sigma$, the volatility of the underlying stock, can be observed in the market. Regarding $\sigma$, we assume to estimate it from historical market data (historical volatility) even though it may not present the true picture. We shall briefly be touching the concept of'*implied volatility* at the end of this chapter.

The next example underlines the strength of the Black-Scholes formula.

**Example 4.3.3** *Suppose a non-dividend paying stock is currently selling at Rs* 100 *and the stock's volatility is* 24%. *Assume that the continuously compounded risk-free interest rate is* 5%. *A European call option is offered on this stock with time to maturity 3 months and strike price Rs* 125. *Calculate the price of the block of 100 options in the Black-Scholes framework.*

**Solution** Here, $S(0) = Rs\ 100,\ K = Rs\ 125,\ T = 3/12 = 0.25,\ r = 0.05,\ \sigma = 0.24.$

$$d_1 = \frac{ln(100/125) + (0.05 + \frac{1}{2}(0.24)^2)\,0.25}{0.24\sqrt{0.25}} = -1.695363$$

and

$$d_2 = d_1 - \sigma\sqrt{T} = -1.815363.$$

Because $d_1$ and $d_2$ are negative, we use, $\Phi(d_1) = 1-\Phi(-d_1)$ and $\Phi(d_2) = 1-\Phi(-d_2)$. We round off $d_1$ to 1.70 and $d_2$ to 1.82 before looking up the standard normal distribution table. Thus,

$$\Phi(d_1) = 1 - 0.9554 = 0.0446, \quad \Phi(d_2) = 1 - 0.9656 = 0.0344.$$

Invoking (4.13), we get

$$C(0) = 100(0.0446) - 125e^{(-0.05)(0.25)}(0.0344) = 0.2134.$$

Therefore, the cost of the block of 100 options is Rs 21.34.

□

In case we wish to compute the European call option price at any time $t$, $0 \le t < T$, then we simply need to make a small change in the Black-Scholes formula as follows.

$$C(t) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

$$d_1 = \frac{ln\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T-t},$$

where $S(t)$ is the spot price of the stock at time $t$ and $T - t$ is the remaining time to maturity.

The price of the European put option with the strike price $K$ and time to maturity $T - t,\ 0 \leq t < T$, can either be calculated using the put-call parity or directly by the following Black-Scholes formula.

$$P(t) = Ke^{-r(T-t)}\Phi(-d_2) - S(t)\Phi(-d_1)$$

$$d_1 = \frac{ln\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

Obviously if we enter into the European put option on the day it is offered then the put price $P(0)$ can be obtained by taking $t = 0$ in the above formula. Here, we encouraged the readers to derive the Black-Scholes formula for pricing European put option using the analogous arguments as done above in the derivation of $C(0)$.

So far, we have discussed a European call or a European put options cases. It is well known that, for a non-dividend paying stock, the value of an American call option is same as that of a European call option with the same strike price and time to expiration. Hence (4.13) can be used to compute price of an American call option on a non-dividend paying underlying stock. Notably, the Black-Scholes formula for an American put option is not the same as for European put option as it may pay to exercise them early. An American put price has to be approximated using the binomial method explained in previous chapter wherein we can simulate sufficiently large binomial lattice by taking sufficiently small time steps for more accurate approximation.

## 4.4 Black-Scholes Formula for Dividend Paying Stock

One of the assumptions in the Black-Scholes basic model is that the underlying stock pays no dividend. We can easily relax this by adjusting the initial stock price by the present value of the dividend. To make us understand this, let us first explore how the option gets affected by dividends on the underlying stock.

Dividends are a way by which the companies distribute part of their profits to their investors. The question of who should be paid dividends becomes complex as the composition of shareholders changes each day. To settle this question, companies designate a date known as the record date. Dividends are paid to the list of shareholders who hold stock on the record date. In order to allow time for settling stock purchases, stock exchanges set a date generally two business days prior to the record date known as the ex-dividend date. Someone who purchases the stock on or after the ex-dividend date is not eligible to receive dividends.

Now suppose we own a call option on a dividend paying stock, but we don't own any of the actual underlying stock. Then the bad news is that we are not entitled to get any dividends because dividends are paid only to the actual shareholders. We can only receive the dividend if we own the stock before the ex-dividend date. All what we own is the right to become a shareholder and do so only when, for instance, we exercise the call option. But then all is not lost. The good news is that not getting the dividends while holding only the option is compensated on the price of option. This is due to the fact that on the ex-dividend date the underlying stock price falls by roughly the amount of the dividend. This drop causes call option price becoming cheaper relative to the amount of the dividend and put option price becoming slightly more expensive. In other words, with dividends, the premium we have to pay for call option gets reduced while the put option premium increases.

The Black-Scholes option pricing formula has to be adjusted when the underlying stock pays dividends. For adjusting the price of European options, for known discrete dividends, we merely have to subtract the present value of the dividend from the current price of the underlying asset in calculating the Black-Scholes value. That is, to take into account the drop in stock price on ex-dividend date, instead of $S(0)$ work with $S(0) - div\, e^{-r\, t_{div}}$ in the Black-Scholes formula, where $div$ denotes the dividend amount in proportions to stock, and $t_{div}$ denotes the dividend time. In case if dividends $div_1, \ldots, div_p$ are paid at discrete times $t_{div_1}, \ldots, t_{div_p}$ then take $S(0) - \sum_{j=1}^{p} div_j e^{-r\, t_{div_j}}$ instead of $S(0)$ in the Black-Scholes formula (4.13) and (4.12). A word of caution: be careful to see that $S(0) > \sum_{j=1}^{p} div_j\, e^{-r\, t_{div_j}}$ for we have to take its logarithm while computing $d_1$ in the Black-Scholes formula.

The following examples help us to appreciate the discussed scenario.

**Example 4.4.1** *A stock currently trades for Rs* 100 *per share. The annual continuously compounded risk-free interest rate is* 5% *and the annual price volatility* *relevant for the Black-Scholes formula is* 30%. *Call options are written on this stock with a strike price of Rs* 80 *and time to expiration of 5 years. The stock will pay a dividend of Rs* 20 *in 2 years and another dividend of Rs* 30 *in 3 years. Use the Black-Scholes formula to find the price of one such call option.*

**Solution** Here, $S(0) = Rs\ 100,\ K = Rs\ 80,\ T = 5,\ r = 0.05,\ \sigma = 0.3,\ div = Rs\ 20$ for first 2 years and $Rs\ 30$ for next 3 years. We first compute the adjusted stock price by decreasing the current stock price by the present value of the dividend. The adjusted stock price is denoted by say $S_a$.

$$S_a = S(0) - (div)e^{-r\,t_{div}} = 100 - 20e^{-(0.05)2} - 30e^{-(0.05)3} = 56.082.$$

Now we use the formula

$$\begin{aligned} d_1 &= \frac{ln(S_a/K) + (r + 0.5\sigma^2)T}{\sigma\sqrt{T}} \\ &= \frac{ln(56.08/80) + (0.05 + 0.5(0.3)^2)5}{0.3\sqrt{5}} = 0.1786. \\ d_2 &= d_1 - \sigma\sqrt{T} = -0.4922. \\ \Phi(d_1) &= \Phi(0.18) = 0.5714, \quad \Phi(d_2) = 1 - \Phi(-d_2) = 1 - 0.6879 = 0.3121. \\ C(0) &= S_a\Phi(d_1) - Ke^{-rT}\Phi(d_2). \end{aligned}$$

Thus,

$$C(0) = (56.082)(0.5714) - (62.304)(0.3121) = \text{Rs } 12.60.$$

□

**Remark 4.4.1** *Suppose in the above example the underlying stock pays no dividend in the first two years but thereafter pays a dividend of Rs* 30 *in 3 years. Then,*

$$S_a = 74.1787,\ d_1 = 0.5955,\ d_2 = -0.0753.$$

$$\Phi(d_1) = 0.7123,\ \Phi(d_2) = 1 - \Phi(-d_2) = 0.4721 \text{ and } C(0) = Rs\ 23.42.$$

*Now assume the stock pays no dividend in its entire lifespan of 5 years, that is,* $div = 0$. *Then*

$$d_1 = \frac{ln(100/80) + (0.05 + 0.5(0.3)^2)5}{0.3\sqrt{5}} = 1.0407, \quad d_2 = 0.3699.$$

$$\Phi(d_1) = 0.8508,\ \Phi(d_2) = 0.6443,$$

*and hence* $C(0) = Rs\ 44.94$.

*Thus, the dividend paid by the underlying stock gets reflected in price reduction of the European call option.*

**Remark 4.4.2** *Working with the same data as in Example 4.4.1, we assume that a European put has been offered on the stock. Although we can directly compute the put price using the Black-Scholes formula for European put option, we use the put-call parity with dividends to do so.*

*(i) In case the stock pays dividends exactly in the same manner as described in the Example 4.4.1 then*

$$P(0) = C(0) - (S(0) - (div)e^{-r\,t_{div}} - Ke^{-rT}) = Rs\ 18.82.$$

*(ii) Consider the other case when the stock pays no dividends in the first 2 years and pays a dividend of Rs* 30 *in the next 3 years. Then*

$$P(0) = C(0) - (S(0) - (div)e^{-r\,t_{div}} - Ke^{-rT}) = Rs\ 11.55.$$

*(iii) If stock pays no dividend in 5 years then* $P(0) = C(0) - (S(0) - Ke^{-rT}) = Rs\ 7.24$.

*This clearly illustrates that the European put option become more costly on a dividend paying underlying stock.*

It is worth to note that if the underlying stock pays continuous dividend at a rate $\delta_{div}$, then the price of the European call option on such a stock is computed as follows.

$$C(0) = S(0)e^{-\delta_{div}T}\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

$$d_1 = \frac{ln\Big(\frac{S(0)e^{-\delta_{div}T}}{K}\Big) + \Big(r + \frac{\sigma^2}{2}\Big)T}{\sigma\sqrt{T}},$$

$$d_2 = d_1 - \sigma\sqrt{T}.$$

Similarly the price of a European put option is given by

$$P(0) = Ke^{-rT}\Phi(-d_2) - S(0)e^{-\delta_{div}T}\Phi(-d_1),$$

where $d_1$ and $d_2$ are described immediately above.

**Example 4.4.2** *A stock currently trades for Rs* 100 *per share. The annual continuously compounded risk-free interest rate is* 5% *and the annual price volatility relevant for the Black-Scholes formula is* 20%. *European call option is written on this stock with a strike price of Rs* 80 *and time to expiration of 3 months. The stock pays dividend continuously at the rate of* 2%. *Use the Black-Scholes formula to find the price of one such call option.*

**Solution** The data is $S(0) = Rs\ 100$, $K = Rs\ 80$, $T = 3/12 = 0.25$, $r = 0.05$, $\sigma = 0.2$, and $\delta_{div} = 0.02$.

$$d_1 = \frac{ln(S(0)e^{-\delta_{div}T}/K) + (r + 0.5\sigma^2)T}{\sigma\sqrt{T}} = 2.3564, \quad d_2 = 2.2564.$$

$$\Phi(d_1) = 0.9909, \ \Phi(d_2) = 0.9881.$$

$$C(0) = 100e^{-(0.02)0.25}(0.9909) - 80e^{-(0.05)0.25}(0.9881) = Rs\ 20.53.$$

□

## 4.5 The Greeks

Let us shift our attention to another noteworthy aspect of the Black-Scholes formula that we had already worked out for the equilibrium price of the European call option on a non-dividend paying stock. Recall that

$$C(t) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \tag{4.14}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}, \ \ d_1 = \frac{ln(S(t)/K) + (r + (\sigma^2/2))(T-t)}{\sigma\sqrt{T-t}}. \tag{4.15}$$

Note that, besides the two constants - time to maturity ($T$) and the strike price ($K$), quite a few parameters are involved in the Black-Scholes formula, viz., the price of the underlying asset $S(t)$; time to price $t$; the volatility of the underlying asset $\sigma$; and the interest rate $r$. Even a minor change in one or more of them will affect the price of the call. It is therefore essential to understand the effect of changes in these parameters on the call price.

In calculus, the notion of "derivative" is a pivotal tool used to measure the rate of change of a dependent variable as an independent variable is changed. A similar situation is prevalent in the aforementioned context. Here we will calculate the partial derivatives of option value formula (4.14), and they in turn will allow us to determine how sensitive the value of option is to the changes in the parameters. In this section we shall be learning about *the delta*, *the gamma*, *the theta*, *the vega*, and *the rho*. These famous fives are collectively known as *the Greeks*.

Let us first begin with some simple derivatives. The cumulative distribution function is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy.$$

By the fundamental theorem of calculus, the derivative of $\Phi$ with respect to $x$ is

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \tag{4.16}$$

Also, a basic calculus working in (4.15) yield the following.

$$\frac{\partial d_1}{\partial t} = \frac{1}{2\sigma\sqrt{T-t}}\Big(\frac{ln(S/K)}{T-t} - r - \frac{\sigma^2}{2}\Big). \tag{4.17}$$

$$\frac{\partial d_1}{\partial S} = \frac{1}{S\sigma\sqrt{T-t}}. \tag{4.18}$$

$$\frac{\partial d_1}{\partial r} = \frac{\sqrt{T-t}}{\sigma}. \tag{4.19}$$

$$\frac{\partial d_1}{\partial \sigma} = \sqrt{T-t} - \frac{d_1}{\sigma}. \tag{4.20}$$

We encourage the readers to compute the above expressions.

Since time is the only deterministic independent variable we are certain, we first introduce the Greek related to time; also called the *Theta* denoted by $\Theta$. The theta of the European call option is defined as $\frac{\partial C}{\partial t}$. From (4.14), we have

$$\Theta = \frac{\partial C}{\partial t} = S\Phi'(d_1)\frac{\partial d_1}{\partial t} - rKe^{-r(T-t)}\Phi(d_2) - Ke^{-r(T-t)}\Phi'(d_2)\Big(\frac{\partial d_1}{\partial t} + \frac{\sigma}{2\sqrt{T-t}}\Big).$$



We can plug in the values of the partial derivatives from (4.16) and (4.17) to get the final expression for $\Theta$. Moreover, using the put-call parity, we can easily compute the $\Theta$ of the non-dividend paying European put option.

Perhaps the most significant Greek is *the delta* denoted by $\Delta$. For the European call option, it is defined to be the partial derivative of call price $C$ with respect to the price of the stock (underlying derivative security). From (4.14), and noting that $\partial d_2/\partial S = \partial d_1/\partial S$, we have

$$\Delta = \frac{\partial C}{\partial S} = \Phi(d_1) + (S\Phi'(d_1) - Ke^{-r(T-t)}\Phi'(d_2))\frac{\partial d_1}{\partial S}. \tag{4.21}$$

Now concentrate on the second term in the expression (4.21), we will try to simplify it. Using (4.16) and (4.18), we obtain

$$\begin{aligned}(S\Phi'(d_1) - Ke^{-r(T-t)}\Phi'(d_2))\frac{\partial d_1}{\partial S} &= \frac{1}{S\sigma\sqrt{2\pi(T-t)}}(Se^{-d_1^2/2} - Ke^{-r(T-t)-d_2^2/2})\\ &= \frac{1}{S\sigma\sqrt{2\pi(T-t)}}(Se^{-d_1^2/2} - Ke^{-r(T-t)-\frac{(d_1-\sigma\sqrt{T-t})^2}{2}})\\ &= \frac{e^{-d_1^2/2}}{S\sigma\sqrt{2\pi(T-t)}}(S - Ke^{-(r+\sigma^2/2)(T-t)+\sigma d_1\sqrt{T-t}})\\ &= \frac{e^{-d_1^2/2}}{S\sigma\sqrt{2\pi(T-t)}}(S - Ke^{ln(S/K)})\\ &= 0.\end{aligned}$$

Therefore, from (4.21) we get $\Delta = \Phi(d_1)$ for the non-dividend paying European call. Notice that $\Delta > 0$, and it is less than 1. Using the put-call parity we can easily work out that for a non-dividend paying European put option $\Delta = \Phi(d_1) - 1$.

**Example 4.5.1** *The current price of a stock is Rs 100 and its volatility is 30% per year. The risk-free interest rate is 4% per year. A portfolio is constructed consisting of one 6-month European call option with a strike price of Rs 80 and the cash obtained from shorting $\Delta$ shares of the stock. The portfolio value is non-random. What is $\Delta$?*

**Solution** By assumption that the portfolio value is non-random, we have, $C - \Delta S = 0$. Thus

$$\frac{\partial C}{\partial S} = \Delta = \Phi(d_1),$$

where

$$d_1 = \frac{ln(100/80) + (0.04 + (0.3)^2/2)(0.5 - 0)}{0.3\sqrt{0.5 - 0}} = 1.2522.$$

Consequently, $\Delta = 0.894752$.

□

Before we proceed in describing the other Greeks, we would like to get more comfortable as to what roles these Greeks play in option theory. We present a situation below in this context.

A company has sold a 3-month European call option on 1,000 shares of a non-dividend paying stock at a strike price of *Rs* 100. The share is currently trading at *Rs* 95 per unit, the interest free rate of return is 5% per annum, and the market volatility is estimated at 20% per annum. Using (4.14) and (4.15) we can easily compute the value of 1,000 call options $C(0) =$ *Rs* 2269.30. The company is a writer of call, so it is exposed to a risk on the call options which is the difference between the value of the call premium (for 1000 calls) invested at a risk free rate of interest for 3 months and the value of the call at the exercise time. Thus the risk profile of the company is $C(0)e^{rT} - N\text{Max}\{S(T) - K, 0\} = Rs\ (2297.84 - 1000\text{Max}\{S(\frac{1}{4}) - 100, 0\})$. *A natural question is that how does the company hedge this risk?* If it decides not to take any action (called the naked position) then the company is exposed to lose if the stock price rises from present $Rs\ 95$. See, if $S(\frac{1}{4}) = Rs\ 103$, then the company loses $Rs\ 2,702.16$ (the amount may appears to be meagre but the numeric data in the example can be changed to make this figure large). On the other hand, if the company decided to buy say $1,000$ shares of the underlying stock at $t = 0$ (called the covered position), then it is exposed to lose if stock decreases. See, if $S(\frac{1}{4}) = Rs\ 90$, the company will lose $Rs\ 2,702.16$. Both positions expose the company to risk! One can easily think of designing some other strategy, like wait and buy $1,000$ shares as soon as price of the stock reaches $Rs\ 100$ and sell $1,000$ shares as soon as the price of the stock declines below $Rs\ 100$. But simple hedging strategy sometime does not work well.

What we are looking for is a more concrete and workable strategy to hedge the risk positions in options created due to changes in the stock price. It is here when the delta, $\Delta$, becomes a handy tool. In the immediate aforementioned situation, the company has $S(0) = Rs\ 95$, $C(0) = Rs\ 2,269.30$, and $\Delta = \Phi(d_1) = \Phi(-0.338) = 0.3685$. By linear approximation from the classical calculus, we have, $dC = \Delta\, dS$. So, if stock price goes up by $Rs$ 1 (i.e. it becomes $Rs$ 96) then the call price increases approximately by $Rs\ 0.3685$ or $Rs 368.50$ for $1,000$ call options. Now, since the company is the writer of $1,000$ call, it must own 368.50 shares of stock so that a change of one unit in the stock price is offset by the change in the short call.

Generally speaking, a writer (short position) of any derivative security can take a hedge position against risk in the security whose price depends on the underlying stock. An important point is that the writer should frequently re-balance the hedge position by constantly monitoring the underlying stock price movements. A portfolio embedding the derivative security can be constructed such that it becomes delta-neutral where by the *delta-neutral portfolio* we mean the portfolio having delta equal to zero. Traders usually ensure that their portfolios are delta-neutral at least once a day. A trading strategy that dynamically maintains a delta-neutral portfolio is called *the delta hedging* which is based on the simple trading rule of “buy high, sell low”.

Continuing with our discussion on Greeks, another important Greek is *the gamma* denoted by $\Gamma$. The gamma for the non-dividend paying European call option is the second order partial derivative of $C$ with respect to $S$, thus, $\Gamma = \dfrac{\partial^2 C}{\partial S^2}$. Obviously,

$$\Gamma = \frac{\partial^2 C}{\partial S^2} = \Phi'(d_1)\frac{\partial d_1}{\partial S} = \frac{e^{-d_1^2/2}}{S\sigma\sqrt{2\pi(T-t)}}.$$

Note that since $\Gamma > 0$ for the European call option, thus $C$ is a convex function of $S$, keeping other parameters constants. Again, using the put-call parity one can work out the gamma of the European put option. We left this for the readers to complete.

Just like delta hedging, there is another hedging strategy called the *delta-gamma hedging*. A delta-gamma hedge is a delta hedge that maintains zero portfolio gamma. The additional gamma neutrality is maintained by constructing a portfolio comprising of a short position in the derivative security, a long position in the underlying stock, and a long position in another hedging call such that the delta and the gamma of the portfolio become zero. Note that the delta hedge is based on the first-order approximation; and when change in $S$ is not small enough, the second-order approximation (or gamma) helps. For this reason, a *delta-gamma neutral portfolio* generally offers better protection against the stock price changes than a simple delta neutral portfolio.

Also, we shall see later in Chapter 10 that the Black-Scholes PDE is

$$\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2 (S(t))^2 \frac{\partial^2 f}{\partial S^2} + rS(t)\frac{\partial f}{\partial S} = rf.$$

Since $\dfrac{\partial f}{\partial t} = \Theta$, $\dfrac{\partial f}{\partial S} = \Delta$, and $\dfrac{\partial^2 f}{\partial S^2} = \Gamma$, the Black-Scholes PDE can be thought of as

$$\Theta + rS(t)\Delta + \frac{1}{2}\sigma^2 S^2 \Gamma = rf. \tag{4.22}$$

Here, $f$ in (4.22) can be price of any derivative security which is governed by the Black-Scholes formulas (4.14) and (4.15). Also, (4.22) provides a relationship between the three Greeks $\Theta$, $\Delta$ and $\Gamma$. The readers can skip this discussion for the time being and can come back to it after going through Chapter 10. It will provide better understanding then. But we have included this text here for the sake of completeness.

The fourth Greek that we would like to get familiar with is called *the vega* which is the partial derivative with respect to volatility $\sigma$. It is denoted by $\mathcal{V}$. For the European call option it is $\dfrac{\partial C}{\partial \sigma}$. From (4.14), we have

$$\mathcal{V} = \frac{\partial C}{\partial \sigma} = S\Phi'(d_1)\frac{\partial d_1}{\partial \sigma} - Ke^{-r(T-t)}\Phi'(d_2)(\frac{\partial d_1}{\partial \sigma} - \sqrt{T-t}).$$

Using (4.20), we get

$$\begin{aligned}\mathcal{V} &= \frac{\sqrt{T-t}}{\sqrt{2\pi}}Se^{-d_1^2/2} + \frac{d_1}{\sigma}(Ke^{-r(T-t)}\Phi'(d_2) - S\Phi'(d_1)) \\ &= \frac{\sqrt{T-t}}{\sqrt{2\pi}}Se^{-d_1^2/2}.\end{aligned}$$

The latter equation follows on account of $Ke^{-r(T-t)}\Phi'(d_2) - S\Phi'(d_1) = 0$. For detail working of this point, we urge the readers to take a look back at the analogous instance while computing $\Delta = \Phi(d_1)$ for the European call option. It is immediate from the put-call parity that the vega of the European put option is identical to that of the European call option.

**Example 4.5.2** *Consider a* 3*-month European put option on a stock whose current value is Rs* 100 *and whose volatility is* 30% *per annum. The option has a strike price of Rs* 95 *and the risk-free interest rate is* 3.25% *per annum. Find the vega of the option. If the volatility of the stock increases to* 31%, *approximate the change in the value of the put.*

**Solution** We provide the hints to solve; the gaps can be filled by the readers. First compute $d_1$ using (4.15), thereafter compute $\mathcal{V} = \dfrac{\sqrt{T-t}}{\sqrt{2\pi}}Se^{-d_1^2/2}$ for the put option. Use the linear approximation $dP = \mathcal{V}d\sigma$, to find approximate change in the value of the put.

□

If volatility changes then the delta-gamma hedge may not work well. An enhancement is *the delta-gamma-vega hedge*, which also maintains vega zero portfolio. This is accomplished by bringing in one more security in the portfolio.

The final Greek to be introduced herein is *the rho*. It is denoted by $\rho$ and is the partial derivative with respect to the risk-free interest rate $r$. For the European call option it is described as follows.

$$\begin{aligned}\rho = \frac{\partial C}{\partial r} &= S\Phi'(d_1)\frac{\partial d_1}{\partial r} + K(T-t)e^{-r(T-t)}\Phi(d_2) - Ke^{-r(T-t)}\Phi'(d_2)\frac{\partial d_2}{\partial r} \\ &= K(T-t)e^{-r(T-t)}\Phi(d_2).\end{aligned}$$

Again the last equation follows on similar argument as the one given above along with (4.19) and the fact that $\frac{\partial d_2}{\partial r} = \frac{\partial d_1}{\partial r}$. We encourage you to find the $\rho$ of the European put.

**Example 4.5.3** *Consider a 3-month European put option on a stock whose current value is Rs* 100 *and whose volatility is* 30% *per annum. The option has a strike price of Rs* 90 *and the risk-free interest rate is* 3.25% *per annum. Find the rho of the option. If the interest rate increases to* 4%, *then approximate the change in the value of the put option.*

**Solution** Given that $K = \text{Rs } 90,\ S(0) = \text{Rs } 100,\ r = 0.0325,\ \sigma = 0.3,\ T = \frac{3}{12} = 0.25,\ t = 0$. We first compute $d_1$ as follows

$$d_1 = \frac{ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} = 0.83157.$$

The rho of the European put is given by

$$\begin{aligned}\rho &= -K(T-t)e^{-r(T-t)}\Phi(\sigma\sqrt{T-t} - d_1)\\ &= (-22.13733)\Phi(-0.68157) = -5.48463.\end{aligned}$$

Using the linear approximation from the classical calculus, the change in value of put option is given by



$$dP = \rho dr = (-5.48463)(0.0075) = 0.041134.$$

□

## 4.6 Summary and Additional Notes

- American options are much harder to deal with than European ones. Except for American call option on non-dividend paying underlying stock, there is no explicit formula as Black-Scholes for American put options (on dividend or non-dividend paying stock) or American call options on dividend paying stock. There is only one known analytical formula for American call options due to Roll-Geske-Whaley [65] on a stock that pays a known dividend. Also refer to http://www.bus.lsu.edu/academics/finance/faculty/dchance/Instructional/TN98-01.pdf. Otherwise, the CRR binomial model can be applied to compute prices of such options. This has already been explained in previous chapter. For more accuracy in pricing these options, one can take sufficiently small time steps in the CRR binomial model. Some numerical methods for solving partial differential equations can also be applied for computing American option prices. For more details on these tools, we refer to [57, 70].

- Another interesting area where the Black-Scholes theory has been extended is the foreign currency options. The potential of market of foreign currency options has been fully trapped in the United States of America, Europe and some other parts of the world. These markets are used to hedge foreign currency risks. For example, with currency option, one can insure against the adverse affects of changes in foreign exchange rates. A modified Black-Scholes formula have been developed by Garman and Kohlhagen [48]. The formula involves both the foreign and domestic interest rates. The shortcomings of the proposed model has also been rectified in some subsequent studies. One can take a look at research articles [48] for further insight into this concept.
- The volatility $\sigma$ is the only parameter in the Black-Scholes formula which is not explicitly observable in the market. Although we stated that we can use historical volatility yet it has been observed that the theoretical value of the option computed from the Black-Scholes formula fails to match the actual option price quoted in financial circuits for that option. There could be more than one cause for this difference. However, non availability of the actual market value of $\sigma$ is a key reason for it. This forces us to have a deeper look at this issue. We then ask a reverse question: knowing the option price, say $C(0)$, along with other parameters $K,\ T,\ r$ in the Black-Scholes formula, can we compute $\sigma$ for that stock? The $\sigma$ so obtained is called *implied volatility*, and should be distinguished from the historical volatility. Notably, there is no direct formula to obtain $\sigma$ from the Black-Scholes formula as we do not have an explicit formula for inverting the $\Phi$ function. In other words, the Black-Scholes formula can not be inverted to get the explicit expression for $\sigma$ in terms of $C(0),\ K,\ r, T$. Again numerical techniques such as the Newton Raphson method for computing root of an equation can come to our rescue.
- One of the critical limitation of the CRR binomial model is that the volatility is assumed to be constant on all the nodes of the binomial lattice. To acknowledge the critical role of implied volatility and to model option prices consistent with the market, many new approaches have been proposed. Rubeinstein [115] proposed the concept of *implied binomial tree* (called IBT in short) which has been extended by Derman, Kani and Chriss [38]. The principal idea in these studies is to compute local volatility at each node of the tree. This is done by using the Arrow-Debreu prices. Consequently, unlike in the CRR model, the

risk neutral probability also see a change at each node. For further details on IBT, please refer to [31].

- In recent years, innumerable contributions to lattice approaches have been published. We would like to share some of them with our readers. Jarrow and Rudd [68] constructed a binomial model where the first two moments, mean and variance, of the discrete and continuous models coincide. Boyle [20] constructed a trinomial lattice. Tian [136] proposed binomial and trinomial models where the model parameters are derived as unique solutions to some equation systems derived from the first three moments. Leisen and Reimer [82] change the formulas to determine the constant up and down factors, and proposed a new binomial model which converges to the Black-Scholes formula with a high order of convergence than the previously quoted methods. The main limitation of almost all lattice models are their relatively slow speed. They do not offer a practical solution for the calculation of thousands of prices in a few seconds. Rapid calculations are the need of today's market. In fact the issue of rate of convergence of different lattice methods is crucial and has attracted attention off late.
- The Black-Scholes no-arbitrage argument fails to capture the correct picture of financial market involving proportional transaction costs. In such a market, there is no portfolio that can replicate the European call option. Thus the argument of replication of portfolio, as explained in previous chapter, can no longer be applicable. This makes it harder to relax the assumption of "no transaction cost" in binomial models, and consequently in the Black-Scholes theory. Some alternative approaches have been suggested in literature for option pricing with transactions costs. These studies are beyond the scope of present discussion. Interested readers can refer to the works of Boyle and Vorst [20], Leland [81], Sonar et al. [125] and Perrakis and Lefoll [108].
- While discussing the matching of CRR model with a multi period binomial model in Section 4.3, it has been assumed that $ud = 1$, i.e. $S(0) = \sqrt{(uS(0))(dS(0))}$. In other words $S(0)$ is the geometric mean of $uS(0)$ and $dS(0)$. This opens the possibility of using arithmetic mean and harmonic mean of $uS(0)$ and $dS(0)$ as well. We may refer to Chawala [26] for further details in this regard.

## 4.7 Exercises

**Exercise 4.1** *Consider the following data:* $S(0) = Rs\ 51$, $K = Rs\ 50$, $\sigma = 30\%$, $r = 8\%$. *Assuming the Black-Scholes framework, and that the stock pays* *no dividend, compute* 3*-months European call price and* 3*-months European put price using the Black-Scholes formula. Also compute the put price using the put-call parity. Are the two values same?*

**Exercise 4.2** *The price of a stock is* Rs 260*. A 6-month European call option on the stock with strike price* Rs 256 *is priced using Black-Scholes formula. It is given that the continuously compounded risk-free rate is* 4%*; the stock pays no dividend; the volatility of the stock is* 25%*. Determine the price of the call option.*

**Exercise 4.3** *You own 100 shares of a stock whose current price is* Rs 42*. You would like to hedge your downside exposure by buying 6-month European put option with a strike price of* Rs 40*. It is given that the continuously compounded risk-free rate is* 5%*; the stock pays no dividends; the stock volatility is* 22%*. Assuming the Black-Scholes framework determine the cost of the put option.*

**Exercise 4.4** *Consider purchase of 100 units of 3-month* Rs 25*-strike European call option. It s given that the stock is currently selling for* Rs 20*; the continuous compounding risk free interest is* 5%*; the stocks volatility is* 24% *per annum. If the stock pays dividends continuously at the rate of* 3% *per annum, determine the price of block of 100 call options, assuming the Black-Scholes framework.*

**Exercise 4.5** *For European call and put options on a stock having the same expiry and strike price, it is given that the stock price is* Rs 85*; the strike price is* Rs 90*; the continuously compounded risk free rate is* 4%*; the continuously compounded dividend rate on the stock is* 2%*. If the call option has premium* Rs 9.91 *and a put option has premium* Rs 12.63*, then determine the time to expiry for the options.*

**Exercise 4.6** *Consider a 1-year European call option on a non-dividend paying stock which is currently priced at* Rs 40*. The call strike price is* Rs 45 *and the continuously compounded risk-free rate is* 5%*. It has been observed that if the stock price increases* Rs 0.50*, the price of the option increases by* Rs 0.25*. Assuming the price of the stock follows the Black-Scholes framework, determine the implied volatility of the stock.*

**Exercise 4.7** *Consider a stock which is currently trading at* Rs 100*.*

*(i) Assuming the stock pays no dividend, compute the Black-Scholes European call price for* $T$ *years to maturity with a strike price* Rs 120*, stock price volatility* 30%*, continuous compounding return* 8%*.*
*(ii) What happens to the option price as* $T \to \infty$*?*
*(iii) Suppose the same stock pays a dividend of* 0.1%*. Repeat (i).*

*(iv) Now what happens to the option price in (iii) when* $T \to \infty$*?*
*(v) What accounts for the difference in (ii) and (iv)?*

# 5
# Portfolio Optimization-I

## 5.1 Introduction

The readers have already come across certain terms like *portfolio, portfolio return*, and *portfolio risk* in Chapter 1. There, while giving example of a simple *portfolio optimization problem*, it was remarked that we need to minimize the portfolio risk for given aspiration level of portfolio returns and thereby do a trade-off between the two. This has precisely been the approach of the celebrated *mean-variance theory* of Markowitz [90] for the single period portfolio optimization problem. The aim of this chapter is to continue this discussion in greater detail for a general $n$-asset problem and discuss other related results, in particular the *capital asset pricing model (CAPM)*.



## 5.2 Risk and Return of an Asset

Let us recollect the usual definition of *asset return* and *asset risk* from Chapter 1.

**Definition 5.2.1 (Return of an Asset)** *The return of an asset is an indicator of gain/loss in the investment of an asset in the financial market. It is determined by the following formula*

$$return = \frac{amount\ \ received\ - amount\ \ invested}{amount\ invested}\,.$$

In the context of a single period model, let at the initial time $t = 0$ the price of an asset be $B(0)$, and at the final time $t = T$ its price be $A(T)$. Then its return $r$ is given by

$$r = \frac{B(T) - B(0)}{B(0)}\,,$$

and is normally expressed in percentage.

The positive value of return on an asset signifies gain while the negative value of return signifies loss. The zero value of return means neither gain nor loss from the investment.

**Remark 5.2.1** *It is important to mention here that Definition 5.2.1 is actually the definition of the rate of return on an asset. However, we shall continue to call it as return on an asset for consistency with the financial market glossary.*

**Definition 5.2.2 (Risk of an Asset)** *The risk is often defined as the degree of uncertainty of return on an asset. It signifies the possibility of loss in the investment.*

The risk can either be zero, implying that the asset is risk-free, or positive, implying the asset is risky. If the asset is risk-free then the future value of the asset is known with certainty otherwise the future value of the risky asset is uncertain. Any financial asset can thus be classified either as *risk-free* asset (like a fixed deposit) or as a risky asset (like share of a stock or share of a mutual fund).

Unfortunately there is no unique measure of risk of an asset. In practice we most often measure risk of an asset in terms of variance of returns on that asset, but that may not be the best thing to do. However, unless otherwise stated, the variance will always be used as a risk measure. But we shall have a much deeper discussion on various risk measures in the next chapter.

In order to obtain the mathematical model of a typical portfolio optimization problem we make certain assumptions. These are as follows

(i) The prices of all assets at any time are strictly positive.
(ii) The return $r$ of an asset is a random variable.
(iii) An investor can own a fraction of an asset. This assumption is known as *divisibility*. Also, we assume that this fraction can be positive or negative, i.e. short selling is allowed. We shall mention specifically if short selling is not allowed.
(iv) An asset can be bought or sold on demand in any quantity at the market price. This assumption is known as *liquidity*.
(v) No arbitrage principle holds.
(vi) Unless stated otherwise, there are no commissions/transaction costs.

## 5.3 Portfolio Optimization Problem

We are now in a position to move towards our main goal, i.e. to study the *portfolio optimization problem.* We first define a portfolio and other associated terms.

**Definition 5.3.1 (Portfolio)** *A portfolio is a collection of two or more assets, say, $a_1, \ldots, a_n$, represented by an ordered $n$-tuple $\Theta = (x_1, \ldots, x_n)$, where $x_i \in \mathbf{R}$ is the number of units of the asset $a_i$ $(i = 1, \ldots, n)$ owned by the investor.*

We consider only a single period model, that is, in between the initial time taken as $t = 0$ and the final transaction time taken as $t = T$, no transaction ever takes place.

Let $V_i(0)$ and $V_i(T)$ be the values of the $i$-th asset at $t = 0$ and $t = T$, respectively. Let $V_\Theta(0)$ and $V_\Theta(T)$ denote the values of the portfolio $\Theta = (x_1, \ldots, x_n)$ at $t = 0$ and $t = T$, respectively. Then, we have

$$V_\Theta(0) = \sum_{i=1}^{n} x_i V_i(0), \quad \text{and} \quad V_\Theta(T) = \sum_{i=1}^{n} x_i V_i(T).$$

The quantity $r_\Theta(T) = \dfrac{V_\Theta(T) - V_\Theta(0)}{V_\Theta(0)}$ is referred as the *return of the portfolio* $\Theta$.

**Definition 5.3.2 (Asset Weights)** *The weight $w_i$ of the asset $a_i$ is the proportion of the value of the asset in the portfolio at $t = 0$, i.e.*



$$w_i = \frac{x_i V_i(0)}{\displaystyle\sum_{j=1}^{n} x_j V_j(0)} \qquad (i = 1, \ldots, n).$$

It can be observed that $w_1 + \cdots + w_n = 1$.

In view of Definition 5.3.2, a portfolio can now be represented by the weights $w_i$, $(i = 1, 2, \ldots, n)$ such that $w_1 + \cdots + w_n = 1$. Thus we can imagine that we are having one unit of money (say $Rs$ 1) and if its allocation in the $i^{th}$ asset is $w_i, (i = 1, 2, \ldots, n)$ then the resulting portfolio is $(w_1, w_2, \ldots, w_n)$.

**Remark 5.3.1** *In a portfolio, if $w_i < 0$, for some $i$, it indicates that the investor has taken a short position on the $i$-th asset $a_i$.*

Let $r_i$ be the return on the $i$-th asset. Then

$$r_i = \frac{V_i(T) - V_i(0)}{V_i(0)} \qquad (i = 1, \ldots, n).$$

**Definition 5.3.3 (Mean of the Portfolio Return)** *Let* $(w_1, w_2, \ldots, w_n)$ *be a portfolio of* $n$ *assets* $a_1, a_2, \ldots, a_n$*. Let* $r_i$ $(i = 1, 2, \ldots, n)$*, be the return on the* $i^{th}$ *asset* $a_i$ *and* $E(r_i) = \mu_i$ $(i = 1, 2, \ldots, n)$*, be its expected value. Then the mean of the portfolio return is defined as*

$$\mu = E(\sum_{i=1}^{n} w_i r_i) = \sum_{i=1}^{n} w_i E(r_i) = \sum_{i=1}^{n} w_i \mu_i.$$

**Definition 5.3.4 (Variance of the Portfolio)** *Let* $(w_1, w_2, \ldots, w_n)$ *be a portfolio of* $n$ *assets* $a_1, a_2, \ldots, a_n$*. Let* $r_i$ *be the return on the* $i^{th}$ *asset* $a_i$ *and* $E(r_i) = \mu_i$*,* $(i = 1, 2, \ldots, n)$*. Let* $Cov\ (r_i, r_j) = \sigma_{ij}$*, for* $i, j = 1, 2, \ldots n$*. Then the variance of the portfolio return is defined as*

$$\sigma^2 = Var(\sum_{i=1}^{n} w_i r_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}.$$

*Here the covariance term* $\sigma_{ij}$ *equals* $\rho_{ij}\sigma_i\sigma_j$*, where* $Var(r_i) = \sigma_i^2$*,* $Var(r_j) = \sigma_j^2$ *and* $\rho_{ij}$ *is the correlation coefficient between* $r_i$ *and* $r_j$*.*

In practice, the mean of the portfolio return is simply referred as the *return of the portfolio*. Also the variance or rather the standard deviation of the portfolio return is referred as the *risk of the portfolio*. Therefore given a portfolio $A$ : $(w_1, w_2, \ldots, w_n)$, we can compute its mean $\mu_A$ and standard deviation $\sigma_A$ and therefore get the point $A : (\sigma_A, \mu_A)$ in $(\sigma, \mu)$-plane. Thus irrespective of the number of assets, a portfolio can always be identified as a point in the $(\sigma, \mu)$-plane. This representation is called $(\sigma, \mu)$-*diagram* or $(\sigma, \mu)$-graph (see Fig. 5.1) and is very convenient for further discussion.



The *portfolio optimization problem* refers to the problem of determining weights $w_i$ $(i = 1, 2, \ldots, n)$, such that the return of the portfolio is maximum and the risk of the portfolio is minimum. Thus we aim to solve the following optimization problem

Min $\sum_{i,j=1}^{n} w_i w_j \sigma_{ij}$
and
Max $\sum_{i=1}^{n} w_i \mu_i$
subject to

$$w_1 + w_2 + \ldots + w_n = 1.$$

The above problem, is a bi-criteria optimization problem which has been studied extensively in the mathematical programming literature. As in general, the two

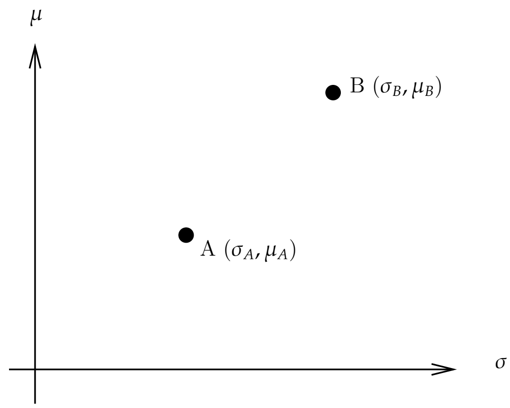



**Fig. 5.1.** $(\sigma, \mu)$-diagram

objective functions may not attain their optimum at the same point, it is not possible to define the meaning of optimization in the usual way. Therefore we consider either (i) to minimize risk for a given level of return or (ii) to maximize return for a given level of risk. From the algorithmic point of view the first approach is more convenient because it results into a linearly constrained quadratic programming problem. We shall first discuss the two asset case and then move to the multi-asset case.



## 5.4 Two Assets Portfolio Optimization

Consider a portfolio with two assets, say, $a_1$, $a_2$ with weights $w_1$, $w_2$, returns $r_1$, $r_2$ and standard deviations $\sigma_1$, $\sigma_2$, respectively. Then the portfolio expected return $\mu$ and portfolio variance $\sigma^2$ are respectively given by

$$\mu = E(w_1 r_1 + w_2 r_2) = w_1 \mu_1 + w_2 \mu_2, \tag{5.1}$$
$$\sigma^2 = Var(w_1 r_1 + w_2 r_2) = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2. \tag{5.2}$$

Here $\rho$ is the coefficient of correlation between $r_1$ and $r_2$, and the value of $\rho$ lies in $[-1, 1]$. We pause here to analyze the effect of $\rho$ on the risk involved in a portfolio. What we shall be observing is that the value of $\rho$ provides a measure of the extent of diversification of portfolio so as to reduce risk. The more negative the value of $\rho$, the greater are the benefits of the portfolio diversification.

As $w_1$ and $w_2$ are weights representing the proportions of total investment in two assets $a_1$ and $a_2$, respectively, we have $w_1 + w_2 = 1$. Moreover, in case of short selling, the weights can be negative. Subsequently, we write $w_1 = 1 - s$, and so, $w_2 = s,\ s \in \mathbf{R}$. Now, it follows from relations (5.1) and (5.2) that

$$\mu = (1-s)\mu_1 + s\mu_2, \tag{5.3}$$
$$\sigma^2 = (1-s)^2\sigma_1^2 + s^2\sigma_2^2 + 2\rho(1-s)s\sigma_1\sigma_2. \tag{5.4}$$

Relation (5.4) can be simplified as

$$\sigma^2 = (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)s^2 - 2\sigma_1(\sigma_1 - \rho\sigma_2)s + \sigma_1^2. \tag{5.5}$$

Without loss of generality we assume that $0 < \sigma_1 \leq \sigma_2$. We discuss the following two independent cases

$$(i) \quad \rho = \pm 1 \qquad\qquad (ii) \quad -1 < \rho < 1.$$

**Case (i)** $\rho = \pm 1$. From relations (5.3) and (5.5), the portfolio expected return and standard deviation are respectively given by

$$\mu = (1-s)\mu_1 + s\mu_2$$
$$\sigma = |(1-s)\sigma_1 \pm s\sigma_2|.$$

For $s \in [0,1]$ both weights are non-negative and thus the portfolio has no short positions. If $s > 1$ then $w_1 < 0$, and therefore that asset $a_1$ is held short. If $s < 0$ then $w_2 < 0$ and therefore it indicates that asset $a_2$ is held short. It may be noted that an investor can not take short position on both the assets simultaneously.

For $\mu = (1-s)\mu_1 + s\mu_2$, $\tilde{\sigma} = (1-s)\sigma_1 + s\sigma_2$, we plot the points $(\tilde{\sigma}, \mu)$ in $(\tilde{\sigma}, \mu)$-plane, and get the first graph of Fig. 5.2. The second graph in Fig. 5.2 corresponds to the case when $\mu = (1-s)\mu_1 + s\mu_2$, $\tilde{\sigma} = (1-s)\sigma_1 - s\sigma_2$. These graphs are essentially straight lines with the bold parts corresponding to $s \in [0,1]$.

Subsequently, we plot the standard deviation-mean diagram of the portfolio, i.e. the $(\sigma, \mu)$-graph, for $\rho = \pm 1$. Observe that $\sigma = |\tilde{\sigma}|$ and therefore to get the required graphics we simply need to flip the portion of the line that lies in the left half-plane over the $\mu$ axis. The two graphs so obtained are depicted in Fig. 5.3. From Fig. 5.3 we have the following observations to share.

**Remark 5.4.1** *When the returns of the two assets are perfectly positively or negatively correlated, i.e.* $\rho = \pm 1$*, the expected portfolio return increases with the gradual decrease in the overall risk of the portfolio (point A to point P Fig. 5.3) till the risk is completely eliminated (point P). After that the higher return comes with higher risk as weight of riskier asset increases. For the case* $\rho = +1$*, the point P corresponds to a portfolio in which a short position is taken on asset* $a_2$*; where as for the case* $\rho = -1$*, the point P corresponds to a portfolio in which weights of both assets are non-negative.*

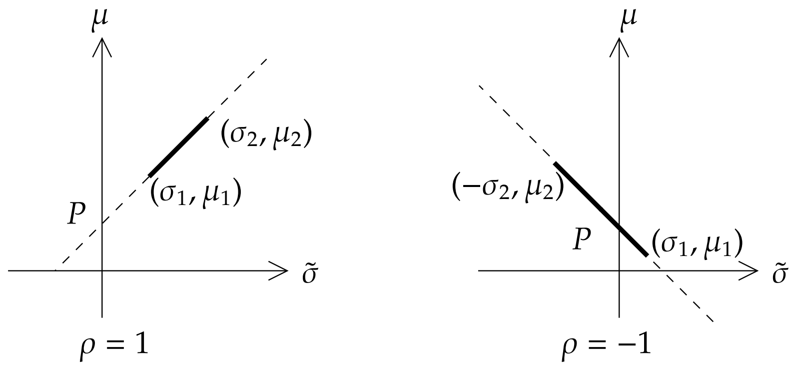



**Fig. 5.2.** $(\widetilde{\sigma}, \mu)$-graph for$\rho = \pm 1$.

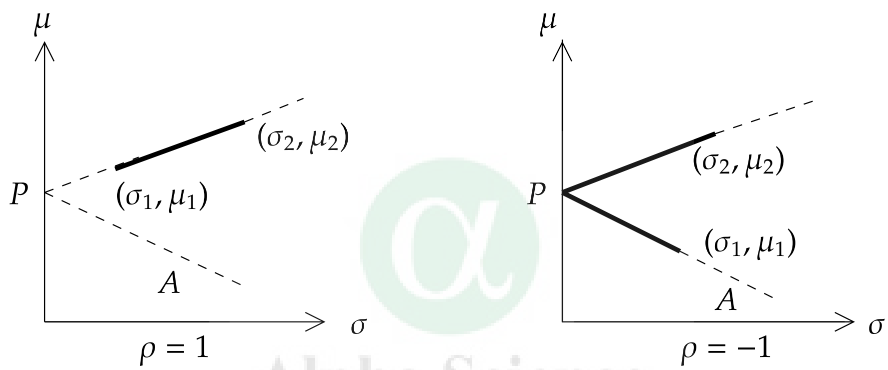





**Fig. 5.3.** $(\sigma, \mu)$-graph for$\rho = \pm 1$.

Below we provide a mathematical justification of the above observations. For the case when $\rho = 1$ and $\sigma_1 = \sigma_2$, then $\sigma_{\min} = \sigma_1$ for all $s \in \mathbf{R}$. Therefore we need to consider following two cases only.

(a) $\rho = 1$ and $\sigma_1 < \sigma_2$. In this case we have

$$\sigma^2 = (1-s)^2\sigma_1^2 + s^2\sigma_2^2 + 2(1-s)s\sigma_1\sigma_2.$$

Our aim is to minimize $\sigma$, or equivalently minimize $\sigma^2$. Now

$$\frac{d\sigma^2}{ds} = 2[s(\sigma_1 - \sigma_2)^2 - \sigma_1(\sigma_1 - \sigma_2)]\,, \tag{5.6}$$

and

$$\frac{d^2\sigma^2}{ds^2} = 2(\sigma_1 - \sigma_2)^2 > 0. \tag{5.7}$$

Thus, to minimize the risk $\sigma$, we must choose the weight $s$ such that $\dfrac{d\sigma^2}{ds} = 0$. Thereby (5.6) yields $s_{min}$ (the value of $s$ for which $\sigma^2$ is minimum) as

$$s_{min} = \frac{\sigma_1}{\sigma_1 - \sigma_2} < 0.$$

Hence $(1 - s_{min}) = \dfrac{\sigma_2}{\sigma_2 - \sigma_1} > 0$. Let $\mu_{min}$ and $\sigma^2_{min}$ respectively denote the expected return and variance of the portfolio with $(w_1 = 1 - s_{min}, w_2 = s_{min})$ then

$$\mu_{\min} = \frac{\sigma_1\mu_2 - \sigma_2\mu_1}{\sigma_1 - \sigma_2} \quad \text{and} \quad \sigma^2_{min} = 0.$$

Since $s_{min} < 0$ (i.e. $w_2 < 0$), an investor can eliminate risk in the portfolio by taking a short position with asset $a_2$.

(b) When $\rho = -1$ and $\sigma_1 \leq \sigma_2$. In this case we have

$$\sigma^2 = (1 - s)^2\sigma_1^2 + s^2\sigma_2^2 - 2(1 - s)s\sigma_1\sigma_2.$$

Thus, we have

$$s_{min} = \frac{\sigma_1}{\sigma_1 + \sigma_2} > 0$$

$$1 - s_{min} = \frac{\sigma_2}{\sigma_1 + \sigma_2} > 0$$



$$\mu_{\min} = \frac{\sigma_1\mu_2 + \sigma_2\mu_1}{\sigma_1 + \sigma_2}$$

$$\sigma^2_{min} = 0.$$

Since $s_{min} > 0$ and $1 - s_{min} > 0$ hence the investor can eliminate the risk in the portfolio without restoring to short selling.

**Case (ii)** We now consider the second case when $-1 < \rho < 1$. Recalling relations (5.3) and (5.5), we have

$$\mu = (1 - s)\mu_1 + s\mu_2,$$

and

$$\sigma^2 = (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)s^2 - 2\sigma_1(\sigma_1 - \rho\sigma_2)s + \sigma_1^2\,,$$

which represents the parametric equation of a parabola in $(\sigma^2, \mu)$-plane. Now

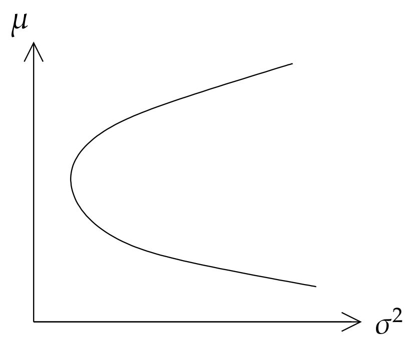



**Fig. 5.4.** $(\sigma^2, \mu)$-graph for $-1 < \rho < 1$.

$$\frac{d\sigma^2}{ds} = 0 \;\Rightarrow\; s = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Also

$$\frac{d^2\sigma^2}{ds^2} = 2((\sigma_1 - \rho\sigma_2)^2 + \sigma_2^2(1 - \rho^2)) > 0.$$

Consequently $s_{min} = \dfrac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$, and the minimum value of $\sigma^2$ is given by

$$\sigma_{\min}^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Further, the corresponding expected portfolio return equals

$$\mu_{\min} = (\mu_2 - \mu_1)s_{\min} + \mu_1.$$

**Remark 5.4.2** *It is important to take note of the following points.*

*(i) The condition* $-1 \leq \rho < \dfrac{\sigma_1}{\sigma_2}$ *is equivalent to* $0 < s_{\min} < 1$. *Thus the minimum risk can be achieved without short selling. Also, in this case*

$$\sigma_{\min}^2 = 0 \;\Leftrightarrow\; \rho = -1.$$

*(ii)* $\rho = \dfrac{\sigma_1}{\sigma_2} \;\Leftrightarrow\; s_{\min} = 0 \;\Leftrightarrow\; \sigma_{\min}^2 = \sigma_1^2.$

*(iii) The condition* $\dfrac{\sigma_1}{\sigma_2} < \rho \leq 1$ *is equivalent to* $s_{\min} < 0$. *In this case the investor has taken a short position on asset* $a_2$ *in order to minimize the portfolio risk. Further, in this case*

$$\sigma^2_{\min} = 0 \iff \rho = 1.$$

The entire theory of this section is summarized in Fig. 5.5.

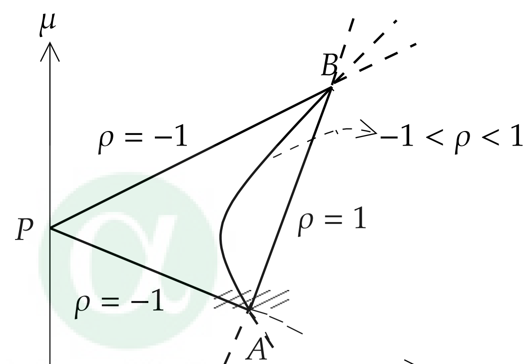





**Fig. 5.5.** Feasible region for two asset problem.

The risk-return relation of two assets for various values of $\rho$ provides us with a triangle $APB$. The points A and B signify undiversified portfolios. Since $-1 \leq \rho \leq 1$, $\triangle APB$ specifies the limit of diversification. The risk-return relation for all values of $\rho$ except $\pm 1$ lie within this triangle. Here the bold portions of the graphs represent the case $0 \leq s \leq 1$. We verify the two-assets portfolio theory by considering the below given example.

Let A and B be two assets with expected returns 12% and 16% and standard deviation 16% and 20%, respectively. Let $x_A$ and $x_B$ denote the number of units of asset $A$ and asset $B$, respectively, in a portfolio. Then we have the following table

| $x_A$ | $x_B$ | $\mu$ | | | $\sigma$ | | |
|---|---|---|---|---|---|---|---|
| | | | $\rho = 1$ | $\rho = 0 \cdot 5$ | $\rho = 0$ | $\rho = -0 \cdot 5$ | $\rho = -1$ |
| 100 | 0 | 12·00 | 16·00 | 16·00 | 16·00 | 16·00 | 16·00 |
| 90 | 10 | 12·40 | 16·40 | 15·50 | 14·54 | 13·51 | 12·40 |
| 80 | 20 | 12·80 | 16·80 | 15·20 | 13·41 | 11·34 | 8·80 |
| 70 | 30 | 13·20 | 17·20 | 15·12 | 12·71 | 9·71 | 5·20 |
| 60 | 40 | 13·60 | 17·60 | 15·26 | 12·50 | 8·91 | 1·60 |
| 50 | 50 | 14·00 | 18·00 | 15·62 | 12·81 | 9·4 | 2·00 |
| 40 | 60 | 14·40 | 18·40 | 16·18 | 13·60 | 10·40 | 5·60 |
| 30 | 70 | 14·80 | 18·80 | 16·92 | 14·80 | 12·32 | 9·20 |
| 20 | 80 | 15·20 | 19·20 | 17·82 | 16·32 | 14·66 | 12·80 |
| 10 | 90 | 15·60 | 19·60 | 18·85 | 18·07 | 17·26 | 16·40 |
| 0 | 100 | 16·00 | 20·00 | 20·00 | 20·00 | 20·00 | 20·00 |

From the above table we observe the following.

(i) The two assets can be combined in such a way that the portfolio risk is less than the individual risks. For instance, when the assets are taken in the ratio 80 : 20 and $\rho = 0{\cdot}5$ then the risk in the portfolio is 15.20% whereas if the entire investment is put in asset $A$ only or in asset $B$ only then the risks involved are 16% and 20%, respectively. While if $\rho = -0 \cdot 5$ then 60 : 40 ratio can bring down the risk to 8.91%. It signifies that in many circumstances diversification of the portfolio is advisable for reducing the risk. The underline principle is thus that 'do not put all the eggs in one basket'.

(ii) For a given weight combination the risk reduces as $\rho$ moves from 1 to -1.

(iii) When $\rho < 1$ then certain combinations of the assets are better than the others. For example, for $\rho = 0$ the 60:40 combination with $\mu = 13.60\%$ and $\sigma = 12.50\%$ is better than the 70:30 combination with $\mu = 13.20\%$ and $\sigma = 12.71\%$. For $\rho = 0.5$, the 70 : 30 combination yields return $13 \cdot 20\%$ and risk 15.12% which is better than the 80 : 20 ratio that gives lower return 12.80% with higher risk 15.20%

From the above discussion we thus conclude that by choosing appropriate ratio of investment between two assets, the risk can be reduced considerably.

## 5.5 Multi Asset Portfolio Optimization

The weights of the various assets $a_1, \ldots, a_n$ in the portfolio are written in the vector form $w^T = (w_1, \ldots, w_n)$. Let $e^T = (1, \ldots, 1) \in \mathbf{R}^n$. Then $w_1 + \cdots + w_n = 1$

can be expressed as $e^T w = 1$. Let $m^T = (\mu_1, \ldots, \mu_n)$ be the expected return vector of the portfolio, where, $\mu_i = E(r_i)$ $(i = 1, \ldots, n)$, and $C = [c_{ij}]$ denotes the $n \times n$ variance-covariance matrix with entries $c_{ij} = Cov(r_i, r_j)$ $(i, j = 1, \ldots, n)$. Note that $c_{ii} = \sigma_i^2$ $(i = 1, \ldots, n)$. Obviously C is a symmetric matrix. Now the expected return $\mu$ of the portfolio is given by

$$\mu = E\left(\sum_{i=1}^{n} w_i r_i\right) = \sum_{i=1}^{n} w_i \mu_i = m^T w\,,$$

and the risk $\sigma^2$ of the portfolio is

$$\sigma^2 = Var\left(\sum_{i=1}^{n} w_i r_i\right) = \sum_{i,j=1}^{n} c_{ij} w_i w_j = w^T C w. \tag{5.8}$$

Here $C$ is certainly positive semidefinite. In practice, it is also assumed to be positive definite (and hence invertible) because the minimum risk of a general $n$-asset portfolio is rarely zero.

During a portfolio selection, every investor is faced with a choice of either minimizing a risk with respect to certain value of return or maximizing a return with respect to certain value of risk. Now, from (5.8), we observe that the portfolio risk $\sigma^2$ depends on three factors, viz.,

(i) risk of each individual asset;
(ii) coefficient of correlation between assets returns;
(iii) weights of the assets.

Out of these contributing factors, the only factor that an investor can control is the weights of the assets. Our main aim is to examine the optimal choice of these weights.

### The Feasible Region of a Portfolio Problem

Let $W = \{w \in \mathbf{R}^n : e^T w = 1\}$ be the collection of all portfolios. We have earlier observed that every portfolio $w \in W$ corresponds to a point in the $(\sigma, \mu)$-plane, say $(\sigma^{(w)}, \mu^{(w)})$. The set $\{(\sigma^{(w)}, \mu^{(w)}) \mid w \in W\}$ is called the *feasible region* or *feasible set* of the given portfolio optimization problem. It should be of interest to know the geometry of the feasible region and that is the point of our discussion now.

Consider the $n$-dimensional hyperplane $e^T w = 1$ in which the weight vector $w$ resides. Let $f$ be the mapping that takes each weight vector in the weight hyperplane to the corresponding portfolio point in the $(\sigma, \mu)$-plane. We try and find the image of any straight line in the weight hyperplane $e^T w = 1$ under the mapping $f$. For this we note that the parametric equation of any line in the weight hyperplane is of the form

$$\begin{aligned} l(\xi) &= (s_1\xi + b_1, \ldots, s_n\xi + b_n)^T \\ &= \xi s + b, \qquad\qquad -\infty < \xi < \infty, \end{aligned}$$

where $s = (s_1, \ldots, s_n)^T$ and $b = (b_1, \ldots, b_n)^T$. Let $w$ be any point on this line. Then

$$\begin{aligned} \mu &= m^T w \\ &= m^T(\xi s + b) \\ &= \xi(m^T s) + (m^T b). \end{aligned}$$

Let $\alpha = (m^T s)^{-1}$, $\beta = -(m^T b)(m^T s)^{-1}$. Then, $\xi = \alpha\mu + \beta$. Moreover,

$$\begin{aligned} \sigma^2 &= w^T C w \\ &= (\xi s + b)^T C(\xi s + b) \\ &= (s^T C s)\xi^2 + (s^T C b + b^T C s)\xi + b^T C b \\ &\equiv \gamma\xi^2 + \delta\xi + \eta. \end{aligned}$$

Substituting the value of $\xi$ we get

$$\sigma^2 = \gamma(\alpha\mu + \beta)^2 + \delta(\alpha\mu + \beta) + \eta. \tag{5.9}$$

As $\xi$ varies from $-\infty$ to $\infty$, the ordered pair $(\sigma^2, \mu)$ traces out a parabola given by (5.9) which lies in $(\sigma, \mu)$-plane with axis parallel to $\sigma$-axis and sides open on the right.
We are actually interested in $(\sigma, \mu)$-graph. Taking the square root of $\sigma^2$, the resulting curve is

$$\sigma = \sqrt{\gamma(\alpha\mu + \beta)^2 + \delta(\alpha\mu + \beta) + \eta}. \tag{5.10}$$

This curve is called a *Markowitz curve*. Thus, each line in the weight hyperplane is mapped onto a Markowitz curve. This phenomena is depicted in Fig. 5.6.

**Remark 5.5.1** *Here it is important to note that the Markowitz curve (5.10) is not a parabola. In fact the main difference between the parabola (5.9) and the Markowitz curve (5.10) in $(\sigma, \mu)$-graph is that a tangent can be drawn to the parabola (5.9) from any point on the $\mu$-axis, whereas the Markowitz curve behaves almost as a straight line as $\mu \to \infty$, thereby, it is not possible to draw a tangent to*

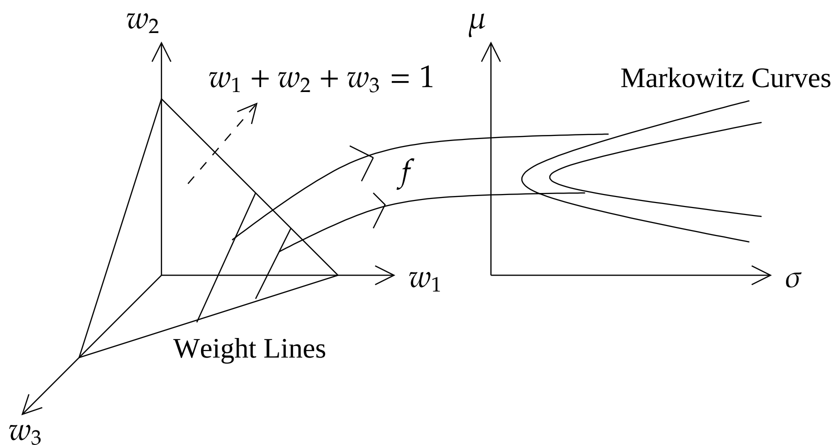



**Fig. 5.6.** Markowitz curves.

*the Markowitz curve as $\mu \to \infty$. This difference may not sound significant right now but it plays a vital role when the portfolio consists of one risk-free asset as well. We shall be addressing to this type of portfolio in the next section. For the current discussion we have assumed that all the assets in the portfolio are risky.*

**Remark 5.5.2** *As we cover the weight hyperplane by taking weight lines, we trace a family of Markowitz curves in the $(\sigma, \mu)$-plane. It is not difficult to get convinced that this region in the $(\sigma, \mu)$-plane is going to be a solid region and its shape will be like a bullet, which is appropriately called the Markowitz bullet.*

There is another much simpler way to get convinced that the feasible set in the $(\sigma, \mu)$ plane should be of the type as described above. Suppose we have three assets $a_1$, $a_2$ and $a_3$. Let these be represented by three points $A : (\sigma_{a_1}, \mu_{a_1})$, $B : (\sigma_{a_2}, \mu_{a_2})$ and $C : (\sigma_{a_3}, \mu_{a_3})$ in the $(\sigma, \mu)$-plane. As already seen in the case of two asset portfolio diagram, any two assets when combined to form portfolios give rise to a curved line or a straight line between them. Whether the line will be straight or curved depends upon the correlation coefficient between them. The three (curved) lines between the possible three pairs are shown in the left hand side figure of Fig. 5.7. Now if we take any asset $D$ formed by the combination of $B$ and $C$, then again combination of assets $A$ and $D$ will give us a (curved) line connecting $A$ and $D$ and the process continues. Eventually we get the feasible region which is a solid two dimensional region in the $(\sigma, \mu)$-plane. This region is depicted in the right hand side figure of Fig. 5.7. Further the feasible region is convex to the left, i.e. given any two points in the region, the straight line connecting them does not cross the left boundary.

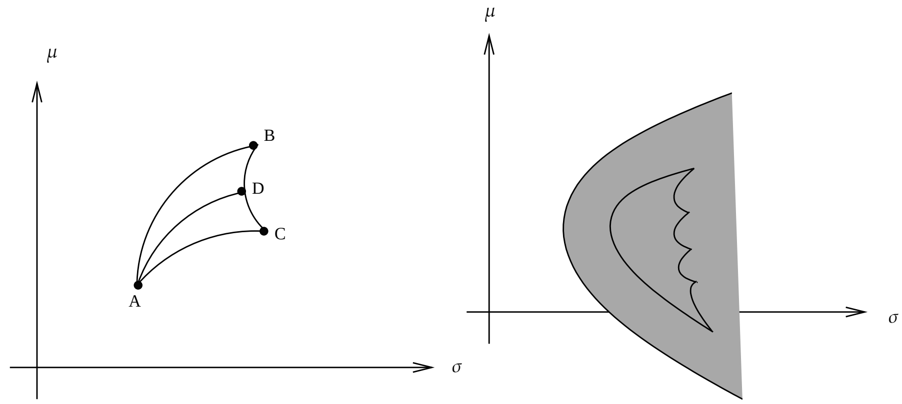



**Fig. 5.7.** Feasible region for three asset problem.

In the feasible region as depicted in Fig. 5.7, the outer region corresponds to the case when shorting is allowed and the inner region corresponds to the case when shorting is not allowed.



## 5.6 The Minimum Variance Set, The Minimum Variance Point and the Efficient Frontier

Consider the feasible set of a given portfolio optimization problem as depicted in $(\sigma, \mu)$-diagram (Fig. 5.8). Its left boundary is called the *minimum variance set* because for any level of the mean rate return, the feasible point with the smallest variance (or standard deviation) is the corresponding left boundary point. This situation is illustrated in Fig. 5.8.

Here when the return is at level $\mu_0$, the feasible point with the smallest variance is $P_0$. Similarly for return levels $\mu_1$ and $\mu_2$ we get the respective points $P_1$ and $P_2$. Amongst all such points $(\sigma, \mu)$ lying on the left boundary of the feasible region, there is a point $P_{min}$ which has the least variance. This point is called the *minimum variance point*.

Mathematically, to find the minimum variance point we need to solve the following risk minimization problem

$$\begin{aligned} \text{Min} \quad & \sigma^2 = w^T C w \\ \text{subject to} \quad & \\ & e^T w = 1. \end{aligned}$$

The following theorem gives a closed form solution of the above problem

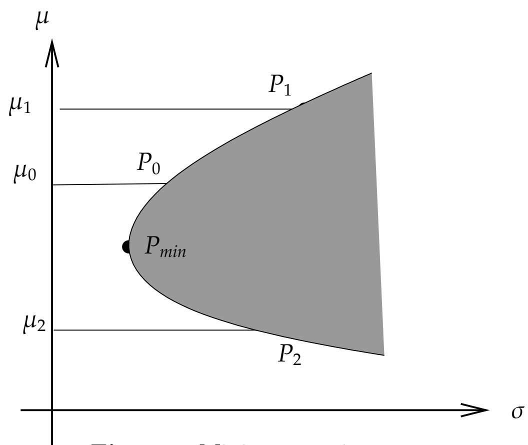



**Fig. 5.8.** Minimum variance set.

**Theorem 5.6.1** *A portfolio with minimum risk has weights given by*

$$w = \frac{C^{-1}e}{e^T C^{-1} e}.$$

**Proof.** We desire to solve the following optimization problem

$$\begin{aligned} &\text{Min} \quad \sigma^2 = w^T C w \\ &\text{subject to} \\ &\qquad\quad e^T w = 1. \end{aligned} \tag{5.11}$$

Using the Lagrange multiplier $\lambda \in \mathbf{R}$, we minimize the Lagrangian

$$L(w, \lambda) = w^T C w + \lambda(1 - e^T w). \tag{5.12}$$

Note that $\lambda$ is unrestricted in sign because the constraint in the risk minimization problem is an equation $e^T w = 1$. Now, differentiating (5.12) with respect to $w$, we obtain

$$2w^T C - \lambda e^T = 0 \implies w = \frac{\lambda}{2} C^{-1} e.$$

Using (5.11), we get

$$e^T \left( \frac{\lambda}{2} C^{-1} e \right) = 1 \implies \frac{\lambda}{2} = \frac{1}{e^T C^{-1} e}.$$

Thus the requisite result follows. □

**Markowitz Efficient Frontier**
Looking at the minimum variance set in the $(\sigma, \mu)$-diagram, we observe that for a given level of risk, (say $\sigma_0$), there are two values of returns ($\mu_0^{(L)}$ and $\mu_0^{(U)}$ with $\mu_0^{(U)} > \mu_0^{(L)}$). Since our aim in portfolio optimization is to maximize return for a given level of risk, we shall obviously choose the larger of these two returns, i.e $\mu_0^{(U)}$. Therefore in the minimum variance set, it is only the upper half which is of importance for investment. This upper half portion of the minimum variance set is called the *Markowitz efficient frontier*. A typical efficient frontier is illustrated in Fig. 5.9.

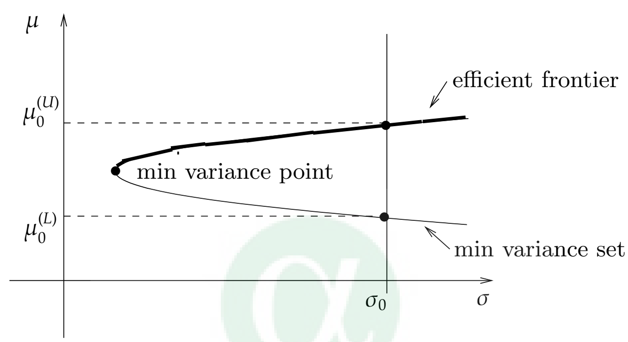



**Fig. 5.9.** Efficient frontier.

We now formalize the definition of efficient frontier and outline a procedure to determine the same.

**Definition 5.6.1 (Dominating Point and Efficient Frontier)** *Let $P : (\sigma_1, \mu_1)$ and $Q : (\sigma_2, \mu_2)$ be two points in the feasible region of a given portfolio optimization problem. Then the point P is said to dominate Q if $\sigma_1 \leq \sigma_2$ and $\mu_1 \geq \mu_2$.*

**Definition 5.6.2 (Non-dominated Point and Efficient Frontier)** *Let A be a point in the feasible region. The point A is said to be a non-dominated point if there is no point P in the feasible region which dominates A. The set of all non-dominated points in the feasible region is called its efficient frontier.*

Geometrically the efficient frontier is the upper portion of the minimum variance set as depicted in Fig. 5.9. Obviously for an investor it is only the efficient frontier which will be of interest and therefore we should explore methodologies to determine the same.

Before we proceed in that direction we notice that in many cases, it is more likely that an investor provides a fixed value of the expected return, say $\mu$, that he/she desires to achieve. Therefore the investor's problem is to decide the right investment strategy to obtain the return $\mu$ with the minimum risk. We look at this scenario in the result to follow.

**Theorem 5.6.2** *For a given expected return $\mu$, the portfolio with minimum risk has weights given by*

$$w = \frac{det\begin{pmatrix} \mu & m^TC^{-1}e \\ 1 & e^TC^{-1}e \end{pmatrix}C^{-1}m + det\begin{pmatrix} m^TC^{-1}m & \mu \\ e^TC^{-1}m & 1 \end{pmatrix}C^{-1}e}{det\begin{pmatrix} m^TC^{-1}m & m^TC^{-1}e \\ e^TC^{-1}m & e^TC^{-1}e \end{pmatrix}}. \tag{5.13}$$

**Proof.** We wish to solve the following quadratic programming problem

$$\text{Min} \qquad \sigma^2 = \frac{1}{2}w^TCw$$

subject to

$$\begin{aligned} m^Tw &= \mu \\ e^Tw &= 1. \end{aligned} \tag{5.14}$$

This is a convex quadratic programming problem with unrestricted variable vector $w$. Therefore we define the Lagrangian



$$L(w, \alpha, \beta) = \frac{1}{2}w^TCw + \alpha(\mu - m^Tw) + \beta(1 - e^Tw)$$

where $\alpha, \beta \in \mathbf{R}$ are Lagrange multipliers. Now solving $\frac{\partial L}{\partial w} = 0$ gives

$$w^TC - \alpha m^T - \beta e^T = 0$$

i.e.

$$w = C^{-1}(\alpha m + \beta e). \tag{5.15}$$

Substituting the value of $w$ in (5.14), we get

$$\begin{aligned} (m^TC^{-1}m)\alpha + (m^TC^{-1}e)\beta &= \mu \\ (e^TC^{-1}m)\alpha + (e^TC^{-1}e)\beta &= 1. \end{aligned}$$

Solving the above two equations for $\alpha$ and $\beta$, we obtain

$$\alpha = \frac{det\begin{pmatrix} \mu & m^TC^{-1}e \\ 1 & e^TC^{-1}m \end{pmatrix}}{det\begin{pmatrix} m^TC^{-1}m & m^TC^{-1}e \\ e^TC^{-1}m & e^TC^{-1}e \end{pmatrix}}, \quad \beta = \frac{det\begin{pmatrix} m^TC^{-1}m & \mu \\ e^TC^{-1}m & 1. \end{pmatrix}}{det\begin{pmatrix} m^TC^{-1}m & m^TC^{-1}e \\ e^TC^{-1}m & e^TC^{-1}e \end{pmatrix}}.$$

Substituting these values in the expression (5.15) for $w$, we get the required expression (5.13).

□

Now to generate the entire efficient frontier we need to solve problems of type (5.14) for all values of $\mu \in \mathbf{R}$. This is almost impossible. But then an extremely interesting observation is made here. Recall from (5.14) and (5.15) that, for a given value of return $\mu$, the points of minimum variance must satisfy the following system of $(n+2)$ linear equations in $(n+2)$ unknowns $w \in \mathbf{R}^n,\ \alpha \in \mathbf{R},\ \beta \in \mathbf{R}$

$$\begin{aligned} w^TC - \alpha m^T - \beta e^T &= 0 \\ m^Tw &= \mu \\ e^Tw &= 1. \end{aligned} \tag{5.16}$$

Suppose we solve the system (5.16) for two distinct values of expected return $\mu$, say $\bar{\mu}^{(1)}$ and $\bar{\mu}^{(2)}$. Let the two solutions be $(w^{(1)})^T = \left((w_1^{(1)}, \ldots, w_n^{(1)}), \alpha^{(1)}, \beta^{(1)}\right)^T$ and $(w^{(2)})^T = \left((w_1^{(2)}, \ldots, w_n^{(2)}), \alpha^{(2)}, \beta^{(2)}\right)^T$, respectively. Then it is simple to verify that the combination portfolio, $\lambda(w^{(1)}, \alpha^{(1)}, \beta^{(1)})^T + (1-\lambda)(w^{(2)}, \alpha^{(2)}, \beta^{(2)})^T,\ \lambda \in \mathbf{R}$, is also a solution of the system (5.16) corresponding to the return $\lambda\bar{\mu}^{(1)} + (1-\lambda)\bar{\mu}^{(2)}$. Therefore, in order to solve (5.16) for every value of $\mu$, one is only required to solve it for two distinct values of $\mu$ and then form the combination of the two solutions. Thus, the knowledge of two distinct portfolios yielding the minimum variances is sufficient to generate the entire minimum variance set. This result is significant from investor's point of view. Also, it demonstrates a very good application of Karush-Kuhn-Tucker optimality conditions. The result is known as the *two fund theorem.*

**Theorem 5.6.3 (Two Fund Theorem)** *Two efficient portfolios can be established so that any other efficient portfolio can be duplicated, in terms of mean and variance, as a linear combination of these two assets. In other words, it says that, an investor seeking an efficient portfolio need to invest only in the combination of these two assets.*

The most convenient way to get two solutions of (5.16) is to assign two distinct values to $\alpha$ and $\beta$, and then work out the solutions. The most convenient choices are $\alpha = 1,\ \beta = 0$ and $\alpha = 0,\ \beta = 1$. The above discussion is illustrated through the below given example.

**Example 5.6.1** *Consider three risky assets with the variance-covariance matrix and expected returns as follows.*

| variance - covariance matrix(C) | | | return(M) |
|---|---|---|---|
| *2* | *1* | *0* | *0·4* |
| *1* | *2* | *1* | *0·8* |
| *0* | *1* | *2* | *0·8* |

*Find two portfolios yielding the minimum variance. Also, determine the expected returns from these two portfolios. Using the two fund theorem, construct the portfolio giving the return of* $33 \cdot 4\%$ *with minimum risk.*

**Solution** Taking $\alpha = 0,\ \beta = 1$ in (5.16), we need to solve: $\sum_{j=1}^{3} \sigma_{ij} v_j^{(1)} = 1\ (i = 1, 2, 3)$, resulting in the following system of linear equations

$$\begin{aligned} 2v_1^{(1)} + v_2^{(1)} &= 1 \\ v_1^{(1)} + 2v_2^{(1)} + v_3^{(1)} &= 1 \\ v_2^{(1)} + 2v_3^{(1)} &= 1. \end{aligned}$$



The solution is $V^{(1)} = (0 \cdot 5,\ 0,\ 0 \cdot 5)^T$. We next take $\alpha = 1,\ \beta = 0$ in (5.16), to solve $\sum_{j=1}^{3} \sigma_{ij} v_j^{(2)} = \mu_i\ (i = 1, 2, 3)$, i.e.

$$\begin{aligned} 2v_1^{(2)} + v_2^{(2)} &= 0 \cdot 4 \\ v_1^{(2)} + 2v_2^{(2)} + v_3^{(2)} &= 0 \cdot 8 \\ v_2^{(2)} + 2v_3^{(2)} &= 0 \cdot 4. \end{aligned}$$

The solution of the above system is $V^{(2)} = (0 \cdot 1,\ 0 \cdot 2,\ 0 \cdot 3)^T$.

Note that $\sum_{j=1}^{3} v_j^{(1)} = 1$, thus we take $w^{(1)} = V^{(1)} = (0.5,\ 0,\ 0.5)$. Normalizing $V^{(2)}$, we get, $w^{(2)} = (1/6,\ 1/3,\ 1/2)^T$ (so that, $\sum_{j=1}^{3} w_j^{(2)} = 1$). The corresponding returns from the two portfolios with weights $w^{(1)}$ and $w^{(2)}$ are $\bar{\mu}^{(1)} = m^T w^{(1)} = 0 \cdot 6$ and $\bar{\mu}^{(2)} = m^T w^{(2)} = 0 \cdot 733$, respectively.

Next, we consider the case when the investor desired a return of $\mu = 0 \cdot 334$ at minimum risk. It is simple to check that for $\lambda = 3$, $\lambda\bar{\mu}^{(1)} + (1 - \lambda)\bar{\mu}^{(2)} = 0 \cdot 334$.

Thus, by the two fund theorem the requisite portfolio is given by $w = \lambda w^{(1)} + (1-\lambda)w^{(2)} = (7/6, -2/3, 1/2)$. Observe that the second asset has a short position in this portfolio. The variance corresponding to this portfolio is

$$w^T C w = \begin{pmatrix} 7/6 & -2/3 & 1/2 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 7/6 \\ -2/3 \\ 1/2 \end{pmatrix} = 2/9.$$

□

## 5.7 Capital Asset Pricing Model (CAPM)

So far we have assumed that all assets in the portfolio are risky assets. So it is natural to query as to what would be the scenario if one risk-free asset is included in the portfolio? In this section we make an attempt to study this aspect of portfolio selection.

Consider a portfolio with $n$ risky assets, $a_1, \ldots, a_n$ with weights $w_1, \ldots, w_n$ and one risk-free asset $a_{\text{rf}}$ with weight $w_{\text{rf}}$. Then

$$w_{\text{risky}} + w_{\text{rf}} = \sum_{i=1}^{n} w_i + w_{\text{rf}} = 1. \tag{5.17}$$

Also, the expected return and the variance associated with this portfolio are respectively given by



$$\mu = \sum_{i=1}^{n} w_i \mu_i + w_{\text{rf}} \mu_{\text{rf}} = \mu_{\text{risky}} + w_{\text{rf}} \mu_{\text{rf}},$$

and

$$\sigma^2 = Var\left(\sum_{i=1}^{n} w_i r_i + w_{\text{rf}} r_{\text{rf}}\right) = Var\left(\sum_{i=1}^{n} w_i r_i\right) = \sigma^2_{\text{risky}}.$$

If we remove the risk-free asset from the portfolio and readjust the weights of the risky assets so that their sum remain 1, the resultant portfolio so obtained is referred to as the *derived risky portfolio*. We use $\mu_{\text{der}}$ and $\sigma^2_{\text{der}}$ to denote the derived risky portfolio expected return and risk, respectively. Then,

$$\begin{aligned}\mu &= \sum_{i=1}^{n} w_i\mu_i + w_{\rm rf}\mu_{\rm rf}\\ &= w_{\rm risky}\left(\sum_{i=1}^{n}\frac{w_i}{w_{\rm risky}}\mu_i\right) + w_{\rm rf}\mu_{\rm rf}\\ &= w_{\rm risky}\mu_{\rm der} + w_{\rm rf}\mu_{\rm rf}\\ &= w_{\rm risky}\mu_{\rm der} + (1-w_{\rm risky})\mu_{\rm rf}\\ &= w_{\rm risky}(\mu_{\rm der}-\mu_{\rm rf}) + \mu_{\rm rf}.\end{aligned} \tag{5.18}$$

Also,

$$\begin{aligned}\sigma^2 &= Var\left(\sum_{i=1}^{n} w_i r_i\right)\\ &= w_{\rm risky}^2 Var\left(\sum_{i=1}^{n}\frac{w_i}{w_{\rm risky}} r_i\right)\\ &= w_{\rm risky}^2\ \sigma_{\rm der}^2,\end{aligned} \tag{5.19}$$

which gives $w_{\rm risky} = \dfrac{\sigma}{\sigma_{der}}$. From (5.18) and (5.19) we get

$$\mu = \mu_{\rm rf} + \left(\frac{\mu_{\rm der}-\mu_{\rm rf}}{\sigma_{der}}\right)\sigma, \tag{5.20}$$



which is an equation of the line joining $(0, \mu_{\rm rf})$ and $(\sigma_{\rm der}, \mu_{\rm der})$ in the $(\sigma, \mu)$-graph.

Now, for a given risk $\sigma$, if we choose various weight combinations of risk-free asset and risky assets satisfying (5.17), we generate different lines represented by (5.20) in $(\sigma, \mu)$-graph. Obviously, among all such lines, the line that produces the point with highest expected return for a given risk is tangent to the upper portion of the Markowitz bullet. This is illustrated in Fig. 5.10.

**Definition 5.7.1 (Capital Market Line)** *Among all the lines (5.20) for various weight combinations of risk-free asset and risky assets, the line giving the highest return for a given risk is called the capital market line.*

**Definition 5.7.2 (Market Portfolio)** *The point on the Markowitz bullet where the capital market line is tangential is said to represent the market portfolio.*

Theoretically, the market portfolio must contain all risky assets, for if some asset is not in it then it will wither and die. Since the market portfolio contains all risky assets, it is a completely diversified portfolio with no unsystematic risk.

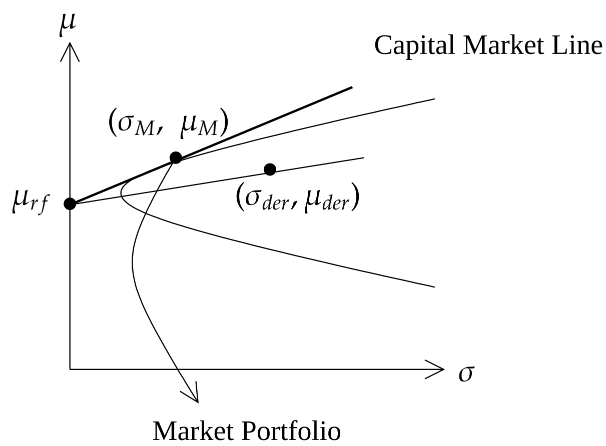



**Fig. 5.10.** Capital market line.

The basic idea of the capital asset pricing model (CAPM) is that an investor can improve the risk-expected return balance by investing partially in a portfolio of risky assets and partially in a risk-free asset. All investors will end up with portfolios along the capital market line as all *efficient portfolios* lie along this line while any other combination of risk-free asset and risky assets, except those which are efficient, lies below the capital market line. It is thus important to observe that all investors will hold combinations of only two assets, viz. the market portfolio $M$ and a risk-free asset. This fund scenario is summarized in the following theorem.

**Theorem 5.7.1 (One Fund Theorem)** *There exists a single portfolio, namely the market portfolio $M$, of risky assets such that any efficient portfolio can be constructed as a linear combination of the market portfolio $M$ and the risk-free asset.*

Unlike with the two fund theorem where any two efficient portfolios are sufficient, in this case, the tangent portfolio is a specific portfolio.

**Theorem 5.7.2** *For any expected risk-free return $\mu_{\rm rf}$, the weight vector $w_M$ of the market portfolio is given by*

$$w_M = \frac{C^{-1}(m - \mu_{\rm rf}\, e)}{e^T C^{-1}(m - \mu_{\rm rf}\, e)}\,.$$

**Proof.** From Fig. 5.10, we observe that for any point $(\sigma, \mu)$ in the Markowitz bullet, the slope of the line joining $(0, \mu_{\rm rf})$ and $(\sigma, \mu)$ is

$$s = \frac{\mu - \mu_{\rm rf}}{\sigma} = \frac{\sum_{i=1}^{n} \mu_i w_i - \mu_{\rm rf}}{\left(\sum_{i,j=1}^{n} c_{ij} w_i w_j\right)^{\frac{1}{2}}}.$$

For the line joining $(0, \mu_{\rm rf})$ to $(\sigma, \mu)$ to be a tangent line to the Markowitz bullet, we need to solve the following optimization problem

$$\begin{aligned} &\text{Max} && \frac{m^T w - \mu_{\rm rf}}{(w^T C w)^{1/2}} \\ &\text{subject to} && e^T w = 1. \end{aligned} \tag{5.21}$$

The Lagrange function $L : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$ for the problem (5.21) is described as

$$L(w, \lambda) = \frac{m^T w - \mu_{\rm rf}}{(w^T C w)^{1/2}} + \lambda(1 - e^T w).$$

Now, solving (5.21) is same as maximizing $L(w, \lambda)$. So, $\nabla_w L(w, \lambda) = 0$, giving

$$\frac{1}{w^T C w}\Big((w^T C w)^{1/2} m - (m^T w - \mu_{\rm rf})(w^T C w)^{-1/2} C w\Big) = \lambda e.$$

The above expression can be rewritten as

$$\sigma m - (\mu - \mu_{\rm rf})\frac{Cw}{\sigma} = \lambda \sigma^2 e.$$



Multiplying by $\sigma$, we obtain

$$\sigma^2 m - (\mu - \mu_{\rm rf}) C w = \lambda \sigma^3 e, \tag{5.22}$$

which in turn yields

$$\sigma^2 w^T m - (\mu - \mu_{\rm rf}) w^T C w = \lambda \sigma^3 w^T e.$$

Since $e^T w = 1,\ \mu = w^T m$, and $\sigma = w^T C w$, we get

$$\lambda = \frac{\mu_{\rm rf}}{\sigma}. \tag{5.23}$$

The requisite value of weight vector $w_M$ now follows from (5.22) and (5.23).

□

**Remark 5.7.1** *Suppose the market portfolio* $(\sigma_M, \mu_M)$ *is known. Then, from (5.20), the equation of the capital market line is given by*

$$\mu = \mu_{rf} + \left(\frac{\mu_M - \mu_{rf}}{\sigma_M}\right)\sigma.$$

*If the investor is willing to take a positive risk* $\sigma$*, he/she can earn an additional return* $\left(\frac{\mu_M - \mu_{rf}}{\sigma_M}\right)\sigma$ *over and above the risk-free return* $\mu_{rf}$ *to compensate the risk taken by him/her. Therefore sometimes the quantity* $\left(\frac{\mu_M - \mu_{rf}}{\sigma_M}\right)$ *is called the price of risk.*

**Example 5.7.1** *Suppose a portfolio comprises of one risk-free asset with return* $0.5$*, and three mutually independent risky assets with expected returns* $1, 2, 3$ *and variances* $1, 1, 1$*, respectively. Determine the equation of the capital market line.*

**Solution** The given information gives, $m^T = (\mu_1, \mu_2, \mu_3) = (1, 2, 3),\ \mu_{rf} = 0.5,\ C = [\sigma_{ij}] = I_{3\times 3},\ e^T = (1, 1, 1)$. Therefore, the weight vector of the market portfolio is given by

$$w_M = \frac{C^{-1}(m - \mu_{rf}e)}{e^T C^{-1}(m - \mu_{rf}e)} = \begin{pmatrix} 1/9 \\ 1/3 \\ 5/9 \end{pmatrix}.$$



Consequently, the expected return and variance of the market portfolio are

$$\mu_M = m^T w_M = \frac{22}{9}, \quad \sigma_M = ((w_M)^T C w_M)^{1/2} = \frac{\sqrt{35}}{9}.$$

Thus, the equation of the capital market line is

$$\begin{aligned} \mu &= \mu_{rf} + \left(\frac{\mu_M - \mu_{rf}}{\sigma_M}\right)\sigma \\ &= \frac{1}{2} + \frac{\sqrt{35}}{2}\sigma. \end{aligned}$$

□

In practice there are certain assets which are listed in the stock called *index stocks*. These limited assets are significant ones that can capture the pulse of the whole market. The most regularly quoted market indices are broad-base indices comprising of the stocks of large companies listed on a nation's largest stock exchanges, such as the American Dow Jones Industrial Average and S&P 500 Index, the British FTSE 100, the French CAC 40, the Japanese Nikkei 225. The

Bombay Stock Exchange is the largest in India, with over 6000 stocks listed and it accounts for over two thirds of the total trading volume in the country. The index stocks finally help us to compute the market portfolio $(\sigma_{\mathrm{M}},\ \mu_{\mathrm{M}})$. The knowledge of the market portfolio yields the equation of capital market line, see Remark 5.7.1. Now suppose an investor $P$ is willing to take risk $\sigma_P$. Then for this risk, the expected return $\mu_P$ is maximum if the point $(\sigma_P, \mu_P)$ lies on the capital market line. Thus,

$$\mu_P = \mu_{\mathrm{rf}} + \left(\frac{\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}}{\sigma_{\mathrm{M}}}\right)\sigma_P.$$

If we let $w_P = \dfrac{\sigma_P}{\sigma_{\mathrm{M}}}$ then

$$\mu_P = w_P\mu_{\mathrm{M}} + (1 - w_P)\mu_{\mathrm{rf}}.$$

**Remark 5.7.2** *The above relation suggests that if an investor is willing to take a risk* $\sigma_P$*, then he/she should invest* $w_P = \dfrac{\sigma_P}{\sigma_{\mathrm{M}}}$ *proportion of investment in index fund and* $(1 - w_P)$ *proportion of investment in the risk-free investment schemes.*

We now aim to examine how an individual asset behaves with respect to the market portfolio. For this, we attempt to build a relationship between the expected return along with the risk of an individual asset with the market portfolio. This gives the CAPM formula (5.24).

**Theorem 5.7.3** *Suppose the market portfolio is* $(\sigma_{\mathrm{M}}, \mu_{\mathrm{M}})$*. The expected return of an asset* $a_i$ *is given by*

$$\mu_i = \mu_{\mathrm{rf}} + \beta_i\,(\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}), \quad \text{where} \quad \beta_{\mathrm{i}} = \frac{\mathrm{Cov}\,(\mathrm{r_i}, \mathrm{r_M})}{\sigma_{\mathrm{M}}^2}. \tag{5.24}$$

**Proof.** Suppose an investor portfolio comprises of asset $a_i$ with weight $w$ and the market portfolio $M$ with weight $1 - w$. Then the expected return and risk of the investor portfolio are respectively given by

$$\begin{aligned} \mu &= w\mu_i + (1 - w)\mu_{\mathrm{M}} \\ \sigma^2 &= w^2\sigma_i^2 + (1 - w)^2\sigma_{\mathrm{M}}^2 + 2\rho w(1 - w)\sigma_i\sigma_{\mathrm{M}} \end{aligned} \tag{5.25}$$

where $\rho$ is the coefficient of correlation between the returns of asset $a_i$ and the market portfolio $M$.

As $w$ varies, these values trace out a curve in the $(\sigma, \mu)$-graph. It can be observed from Fig. 5.11 that as $w$ passes through zero, the capital market line becomes tangent to the curve at $M$. This tangency condition can be translated into the

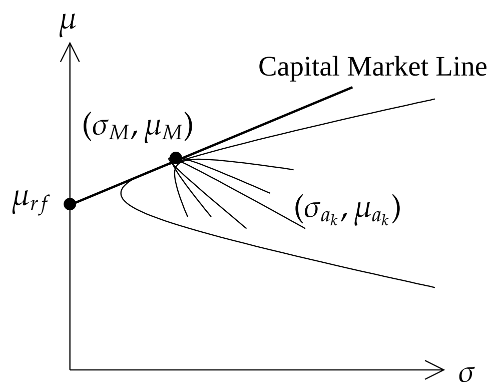



**Fig. 5.11.** Market portfolio.

condition that the slope of the curve is equal to the slope of the capital market line at $M$ (corresponding to $w = 0$).

Now the slope of the curve at $M$ is given by

$$\left.\frac{d\mu}{d\sigma}\right|_{(w=0)} = \left.\frac{d\mu}{dw}\frac{dw}{d\sigma}\right|_{(w=0)} = (\mu_i - \mu_{\mathrm{M}})\left.\frac{dw}{d\sigma}\right|_{(w=0)}.$$



Differentiating (5.25) with respect to $w$ and computing its value at $w = 0$, we get

$$\left.\frac{d\sigma}{dw}\right|_{(w=0)} = \left.\frac{w\sigma_i^2 - (1-w)\sigma_{\mathrm{M}}^2 + \rho\sigma_i\sigma_{\mathrm{M}}(1-2w)}{\sigma}\right|_{(w=0)} = \frac{\sigma_{\mathrm{iM}} - \sigma_{\mathrm{M}}^2}{\sigma_{\mathrm{M}}}, \qquad \sigma_{\mathrm{iM}} = \rho\sigma_i\sigma_{\mathrm{M}}.$$

Consequently,

$$\left.\frac{d\mu}{d\sigma}\right|_{(w=0)} = \frac{(\mu_i - \mu_{\mathrm{M}})\sigma_{\mathrm{M}}}{\sigma_{\mathrm{iM}} - \sigma_{\mathrm{M}}^2}. \tag{5.26}$$

As discussed above, the slope of the curve needs to be equal to the slope of the capital market line at $M$, thereby yielding that

$$\frac{\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}}{\sigma_{\mathrm{M}}} = \frac{d\mu}{d\sigma}\Big|_{(w=0)}.$$

The above relation along with (5.26), on simplification, yields

$$\mu_i = \mu_{\mathrm{rf}} + \left(\frac{\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}}{\sigma_{\mathrm{M}}^2}\right)\sigma_{\mathrm{iM}}$$
$$= \mu_{\mathrm{rf}} + \beta_i\,(\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}).$$

□

**Remark 5.7.3** *Here,* $\beta_i = \dfrac{\sigma_{\mathrm{iM}}}{\sigma_{\mathrm{M}}^2}$ *is called the beta of an asset. Note that, for the market portfolio,* $\beta_M = 1$. *Beta is generally calculated for individual assets using regression analysis. As can be observed, beta measures an asset's volatility or risk in relation to the rest of the market. It is thus appropriately referred to as financial elasticity or correlated relative volatility, and it is all what is required to be known about the asset's risk characteristics in CAPM formula. In other words, an investor ready to bear some systematic risk gets rewarded for it. For instance, if* $\beta_i = 2$, *it indicates that the* $i^{th}$ *asset return is expected to increase (decrease) by* 2% *when the market increases (decreases) by* 1%. *Equivalently, if the market return fluctuates over a specific range of values, the asset returns will fluctuate over a larger range of values. Thus, the market risk is magnified in the asset risk.*

**Remark 5.7.4** *If we take* $\mu_{\mathrm{rf}}$ *as a base, then the above theorem states that the expected return of a particular asset over the risk-free asset (namely* $(\mu_i - \mu_{\mathrm{rf}})$*) is proportional to that of the market (namely* $(\mu_{\mathrm{M}} - \mu_{\mathrm{rf}})$*) and the constant of proportionality is* $\beta_i$. *The CAPM essentially tells that the expected excess rate of return of an assets is directly proportional to its covariance with the market. So it is important to understand this covariance or equivalently the beta of the asset.*

*If* $\beta_i = 0$, *the given asset is completely uncorrelated with the market and therefore CAPM gives* $\mu_i = \mu_{\mathrm{rf}}$. *At first this looks surprising because it tells that no matter how risky the given asset is (even if* $\sigma_i$ *is very large) the expected rate of return of this asset cannot be improved over the base point* $\mu_{\mathrm{rf}}$ *and therefore there is no premium for risk. The main reason being that as the given asset is uncorrelated with the market its risk can be diversified away by having a small amount of this asset in the portfolio.*

*If* $\beta_i < 0$ *then CAPM gives* $\mu_i < \mu_{\mathrm{rf}}$. *This again looks odd because it says that even though asset risk* $\sigma_i$ *may be large, its expected rate of return is less than the base* $\mu_{\mathrm{rf}}$. *But then if, this asset is combined with the market it reduces the overall portfolio risk. Therefore the investor may be willing to have this asset in the portfolio as it has the capability of reducing its overall risk. Such an asset is expected to do well even when other assets are not doing that well and therefore it is called an insurance.*

The above discussion suggests that though for a portfolio an appropriate measure of risk is $\sigma$ but for an individual asset the proper measure of risk is its beta. Thus there is a paradigm shift in understanding the risk of an asset.

**Example 5.7.2** *Let the risk-free rate $\mu_{rf}$ be 8% and the market has $\mu_M = 12\%$ and $\sigma_M = 15\%$. Let an asset a be given which has covariance of 0.045 with the market. Determine the expected rate of return of the given asset.*

**Solution** From the given data we have $\beta_a = \dfrac{0 \cdot 045}{(0 \cdot 015)^2} = 2$. Then CAPM gives

$$\mu_a = \mu_{rf} + \beta(\mu_M - \mu_{rf}) = 0.08 + 2(0.12 - 0.08) = 0 \cdot 16.$$

Therefore the expected rate of return of the given asset is 16%.

□

**Definition 5.7.3 (Beta of the Portfolio)** *The overall $\beta$ of the portfolio is the weighted average of the betas of the individual assets in the portfolio, with the weights being those that define the portfolio, i.e.* $\beta = \sum_{i=1}^{n} w_i \beta_i$.

The above formula is immediate because for the portfolio $w = (w_1, w_2, \ldots, w_n)$, the rate of return is $\sum_{i=1}^{n} w_i r_i$ and $Cov(r, r_{\mathrm{M}}) = \sum_{i=1}^{n} w_i Cov(r_i, r_{\mathrm{M}}) = \sum_{i=1}^{n} w_i \sigma_{\mathrm{iM}} = \sum_{i=1}^{n} \rho \sigma_i \sigma_{\mathrm{M}}$.



**Definition 5.7.4 (Security Market Line)** *A linear equation*

$$\mu = \mu_{\mathrm{rf}} + \beta\,(\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}), \quad \text{where} \quad \beta = \frac{\mathrm{Cov}(\mathrm{r}, \mathrm{r_M})}{\sigma_{\mathrm{M}}^2}$$

*that describes the expected return for all assets in the market is called the security market line.*

The security market line highlights the essence of CAPM formula. It says that under the equilibrium conditions assumed by CAPM, all portfolio investments lie along the security market line in the beta-return space. It emphasizes that the risk of an asset is a function of its covariance with the market, or equivalently a function of its beta. The security market line is depicted in bold in Fig. 5.12

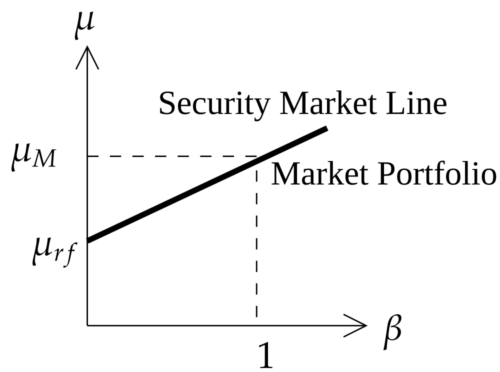



**Fig. 5.12.** Security market line.

### CAPM as a Pricing Formula

We now give another interpretation of CAPM as a pricing formula. Let an asset be purchased at price $P$ and later sold at price $Q$. Then the rate of return is $r = \dfrac{(Q-P)}{P}$. Here $P$ is known but $Q$ is random. If we write $E(Q) = \bar{Q}$, then the CAPM formula gives

$$\frac{\bar{Q}-P}{P} = \mu_{r_f} + \beta(\mu_M - \mu_{r_f})$$

i.e.

$$P = \frac{\bar{Q}}{1 + \mu_{r_f} + \beta(\mu_M - \mu_{r_f})}. \tag{5.27}$$

Here $\beta$ is the beta of the given asset.

The formula (5.27) can be viewed as a pricing formula. Here for the random pay-off $Q$, $\bar{Q}$ is known and the aim is to determine the price $P$.

For the deterministic scenario, it is well known that $P = \dfrac{Q}{1+\mu_{r_f}}$. Therefore, the quantity $\mu_{r_f} + \beta(\mu_M - \mu_{r_f})$ can be interpreted as risk adjusted interest rate. This gives another interpretation of formula (5.27).

**Example 5.7.3** *Consider a mutual fund that invests 10% it funds at a risk free rate of 7%. The remaining 90% is invested in a widely diversified portfolio (resembling the market portfolio) which is expected to give a return of 15%. Further it is known that the beta of the fund is 0.90, and one share of the mutual fund costs Rs 100. Is this price of a share of the mutual fund fair? Justify your answer.*

Solution: The value of a share after one year will be $(10\times1.07)+(90\times1.15) = 114.20$. Thus $\bar{Q} = 114.2$. Therefore

$$P = \frac{114.20}{(1.07) + (0.90)(0.15 - 0.07)} = Rs\ 100.$$

This shows that the price of the share, namely $Rs$ 100, represents $Rs$ 100 of assets in the fund, and therefore CAPM tells that the *price is right*.

□

**The CAPM as a Factor Model**

The CAPM can be derived as a special case of a single factor model. Let us assume that the asset return $r_i$ and market return $r_{\mathrm{M}}$ (taken as a factor) are related as follows

$$(r_i - \mu_{\mathrm{rf}}) = \alpha_i + \beta_i (r_{\mathrm{M}} - \mu_{\mathrm{rf}}) + \epsilon_i. \quad (5.28)$$

Here $\mu_{\mathrm{rf}}$ is the risk-free interest rate and $E(\epsilon_i) = 0$. Also $\epsilon_i$ is uncorrelated with the market return $r_{\mathrm{M}}$ and also with other $\epsilon_j$'s. Further $\alpha_i$ and $\beta_i$ are the usual coefficients appearing in a single factor model.

Taking expectation in (5.27) gives

$$(\mu_i - \mu_{\mathrm{rf}}) = \alpha_i + \beta_i (\mu_{\mathrm{M}} - \mu_{\mathrm{rf}}). \quad (5.29)$$

Here we note that (5.29) is identical with CAPM except that in CAPM, $\alpha_i = 0$. If we further take the covariance of both sides in (5.29) we get

$$\sigma_{iM} = \beta_i \sigma_M^2 .$$

Hence,

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2},$$

which is the same expression as used in CAPM. The equation (5.29) represents a line between the quantities $(\mu_M - \mu_{rf})$ and $(\mu_i - \mu_{rf})$. This line is called the *characteristic equation or characteristic line*.

The characteristic line in a sense is more general than CAPM because here $\alpha_i$ need not be zero. In fact $\alpha_i$ can have a very nice economic interpretation. A stock with non zero $\alpha_i$ can be regarded as mispriced. If $\alpha_i > 0$ then in view of CAPM, the asset is performing better than it should. Similarly if $\alpha_i < 0$, then it is performing worse than it should.

Though we have tried to explain CAPM as a single factor model, we must note that the two are not equivalent. In CAPM we assume that the market is efficient, but in a single factor model we have taken arbitrary covariance matrix $\sigma_{iM}$ and made no assumption on market efficiency.

The CAPM and $\beta_i$ can be understood from a different angle if we take the following model

$$r_i = r_{rf} + \beta_i(r_M - r_{rf}) + \epsilon_i,$$

where $\epsilon_i$ is a random variable. Taking expectation in the above equation and using CAPM we get $E(\epsilon_i) = 0$. Further taking the correlation with $r_M$ in the above equation we get $Cov(\epsilon_i, r_M) = 0$. Therefore we have

$$\sigma_i^2 = \beta_i^2\sigma_M^2 + Var(\epsilon_i).$$

This equation tells that $\sigma_i^2$ is the sum of two expressions. The first expression $\beta_i^2\sigma_M^2$ is called the *systematic risk*. This is the risk associated with the market as a whole. There is no chance of reducing this risk by diversification because all assets with nonzero beta have this risk. The second expression $Var(\epsilon_i)$ is uncorrelated with the market and therefore can be reduced by diversification. The quantity $Var(\epsilon_i)$ is called *unsystematic* risk of the asset. Therefore the systematic risk measured by $\beta$ becomes more important because it directly combines with the systematic risk of other assets.



## 5.8 Summary and Additional Notes

- In this chapter we analyzed the advantages of diversifying the total investment among several assets optimally so as to get a 'decent' return with minimum risk. The theory described here is mainly based on the work of Markowitz [90].
- The requisite terminologies were introduced in Section 5.2, followed by a simple case of two assets portfolio optimization in Section 5.3, to get the feel of the subject. The ideas of Section 5.3 lead us to extend the analysis to the multi asset scenario in Section 5.6.
- Section 5.7 continues with the multi asset portfolio optimization with the difference that one risk-free asset is included in the portfolio. This inclusion results in the Capital Asset Pricing model (CAPM) and a new concept of market portfolio. It was shown that the market portfolio can guide the investor to determine the advantage of taking more risk with his investment.
- The derivation of CAPM assumes the *efficient market hypothesis*. This hypothesis tells that in an efficient market, asset prices fully reflect all available

information. Therefore no relevant information is ignored, and systematic errors are not made. As a consequence, prices are always at levels consistent with fundamentals.

- The contents of the chapter are kept very simple with the intention to familiarize the readers with the basics behind portfolio optimization. However things can be very complex in financial circuits. One needs to realize that any study related to financial problems requires extra skills and high level of understanding. Several interesting books covering various aspects of portfolio theory can be referred to, like, Bartholomew-Biggs [7], Capinski and Zastwawniak [25], Cornuejols and Tütüncü [32], Elton et al. [41], Luenberger [85], Roman [113], Ross [114], to name a few.
- A prominent area where portfolio optimization has gained momentum in recent years is the asset-liability management of the institutional investors, like, insurance companies, pension funds or mutual funds. The institutional investors make huge investments in the markets and simultaneously repay the maturity amounts to the other investors who had invested with them. For this reasons, they need to constantly rebalance their portfolio after every time frame, which is generally very small. An institutional investor will get some inflow of money at $t$ instance of time as the return from various investments that had been made in the market earlier and which had subsequently matured at $t$ time, and also the institutional investor needs to pay the maturity amount to all those investors who had invested with him and whose funds have matured at the end of $t-1$ time. The remaining amount is reinvested in the market. The time scale involved in such asset-liability problems has been captured by using stochastic linear programming models (see Kall [74] for stochastic LPP). Number of research papers can be found on asset-liability management, like Sodhi [125], Yu et al. [148] and Steuer et al. [130] and references therein.
- There are several commercial packages, for instance, CPLEX, LINGO, MATLAB, SAS that provide lot of inbuilt functions for Portfolio Analysis. The major disappointment with all the commercial packages is that they can best generate only the approximate piecewise linear representation of the efficient frontier in portfolio optimization. With large number of assets involved, say 600-800, the performance of these software in computing the efficient frontier deteriorates. The MPQ (multi-parametric quadratic programming), programmed in Java and available in public domain, performs exceptionally well on large-scale applications in a reasonable time and yields the exact efficient frontier. For more on MPQ, we refer to Steuer et al. [130].
- The effect of introduction of transaction costs and/or different lending and

borrowing rates in portfolio optimization theory has also been analyzed in literature. Some of the books mentioned above contain subject matter on this issue.

- Some researchers have re-visited Markowitz's mean- variance model so as to simplify the analysis and computations in the determination of efficient frontier. The interested readers may refer to Chawla [27] and Steinbach [129].

## 5.9 Exercises

**Exercise 5.1** *Suppose there are three financial market scenarios* $\Omega = \{w_1, w_2, w_3\}$ *with different probabilities of occurrence. Consider the following table showing the returns on two different stocks in these three scenarios*

| scenario | prob | return $k_1$% | return $k_2$% |
|---|---|---|---|
| $w_1$ | $0 \cdot 2$ | −10 | −30 |
| $w_2$ | $0 \cdot 5$ | 0 | 20 |
| $w_3$ | $0 \cdot 3$ | 20 | 15 |

*(a) What are the expected returns on the stocks?*
*(b) Suppose* 60% *of the available fund is invested in stock 1 and the remaining is invested in stock 2, then what is the expected return of the portfolio?*
*(c) Compute the weights if the expected return on a portfolio is* 20%.



**Exercise 5.2** *Consider the following data for two different stocks*

| scenario | prob | return $k_1$% | return $k_2$% |
|---|---|---|---|
| $w_1$ | $0 \cdot 4$ | −10 | 20 |
| $w_2$ | $0 \cdot 2$ | 0 | 20 |
| $w_3$ | $0 \cdot 4$ | 20 | 10 |

*Suppose a portfolio comprises of* 40% *of total investment in stock 1 and* 60% *in stock 2. Compare the risk of the portfolio with the risks of its individual components. What will be the risk situation if a portfolio is designed with investment of* 80% *in stock 1 and the remaining in stock 2.*

**Exercise 5.3** *Prove that if short sales are not allowed then the risk of the portfolio can not exceed the greater of the risks of the individual components of the portfolio.*

**Exercise 5.4** *Let a portfolio be designed with investment of* 50% *in stock 1 and the remaning* 50% *in stock 2. Further let short sale be allowed in stock 1 and all the other data being the same as in Exercise 5.2. Does the conclusion of Exercise 5.3 hold.*

**Exercise 5.5** *Suppose the portfolios are constructed using three securities* $a_1$, $a_2$, $a_3$ *with expected returns,* $\mu_1 = 20\%$, $\mu_2 = 13\%$, $\mu_3 = 4\%$, *standard deviations of returns,* $\sigma_1 = 25\%$, $\sigma_2 = 28\%$, $\sigma_3 = 20\%$, *and the correlation between returns,* $\rho_{12} = 0 \cdot 3$, $\rho_{13} = 0 \cdot 15$ *and* $\rho_{23} = 0 \cdot 4$. *Among all the attainable portfolios, find the one with minimum variance. What are the weights of the three securities in this portfolio? Also compute the expected return and standard deviation of this portfolio.*

**Exercise 5.6** *Among all attainable portfolios with expected return* 20% *constructed using the data provided in Exercise* **5.5**, *find the portfolio with minimum variance. Compute the weights of individual assets in this portfolio.*

**Exercise 5.7** *Consider the following data*

| | $\mu$ | $\sigma$ |
|---|---|---|
| asset 1 | 10% | 5% |
| asset 2 | 8% | 2% |

*For each correlation coefficient* $\rho = -1, -0 \cdot 5, 0, 0 \cdot 5, 1$, *what is the combination of the two assets that yields the minimum standard deviation and what is the minimum value of the standard deviation?*

**Exercise 5.8** *Compute the minimum risk portfolio for the following rate return (%) data*

| | Jan | Feb | Mar | Apr | May | June |
|---|---|---|---|---|---|---|
| asset 1 | 12 | 10 | 5 | 7 | 15 | 12 |
| asset 2 | 7 | 12 | 10 | 10 | 12 | 15 |

*Also compute the expected return for the optimal portfolio.*

**Exercise 5.9** *Consider three risky assets with the variance-covariance matrix and expected returns (all data in %) as follows.*

| variance - covariance matrix(C) | | | return(M) |
|---|---|---|---|
| 10 | 4 | 0 | 5 |
| 4 | 12 | 6 | 6 |
| 0 | 6 | 10 | 1 |

*Find two efficient portfolios. Also construct the portfolio giving the return of* $2 \cdot 8\%$ *with minimum risk. Will this portfolio be also efficient?*
*(Hint : use two fund theorem).*

**Exercise 5.10** *Suppose an investor is interested in constructing a portfolio with one risk-free asset* $a_1$*, and three risky assets* $a_2$*,* $a_3$ *and* $a_4$*. Let the expected returns of* $a_1$*,* $a_2$*,* $a_3$ *and* $a_4$ *be* $6\%$, $10\%$, $12\%$ *and* $18\%$ *respectively. Let the variance-covariance matrix* $C$ *of the three risky assets be*

$$C = \begin{pmatrix} 4 & 20 & 40 \\ 20 & 10 & 70 \\ 40 & 70 & 14 \end{pmatrix}.$$

*Determine all efficient portfolios for the investor.*

**Exercise 5.11** *Consider the data of two risky assets* $a_1$*,* $a_2$ *with* $\mu_1 = 12 \cdot 5\%$, $\mu_2 = 10 \cdot 5\%$, $\sigma_1 = 14 \cdot 9\%$, $\sigma_2 = 14\%$, $\rho = 0 \cdot 33$.

(a) *Is it advisable to diversify the investment? If so then what composition of the assets will minimize the risk?*
(b) *What is the minimum value of the risk?*
(c) *If the risk-free rate of return is* $5\%$ *then derive the equation of the capital market line?*



**Exercise 5.12** *Given the following information about the one risk-free asset and three risky assets, find the expected return and standard deviation of the market portfolio. Also determine the equation of the capital market line.*

$$\mu_{rf} = 5\%, \ \mu_1 = 14\%, \ \mu_2 = 8\%, \ \mu_3 = 20\%\,;$$
$$\sigma_1 = 6\%, \ \sigma_2 = 3\%, \ \sigma_3 = 15\%\,; \ \ \sigma_{12} = 0 \cdot 5\%, \ \sigma_{13} = 0 \cdot 2\%, \ \sigma_{23} = 0 \cdot 4\%.$$

**Exercise 5.13** *Assume that the following assets are correctly priced according to the security market line. Derive the security market line.*

$$\mu_1 = 6\%, \ \ \beta_1 = 0 \cdot 5; \quad \mu_2 = 12\%, \ \ \beta_2 = 1 \cdot 5.$$

*What is the expected return on an asset with* $\beta = 2$*?*

**Exercise 5.14** *If the following two assets are correctly priced according to the security market line, what is the return of the market portfolio? What is the risk-free return?*

$$\mu_1 = 9 \cdot 5\%, \ \ \beta_1 = 0 \cdot 8; \quad \mu_2 = 13 \cdot 5\%, \ \ \beta_2 = 1 \cdot 3.$$

**Exercise 5.15** *Let the expected rate of return on the market portfolio be* 23% *and that of the risk free asset be* 7%. *Also let the standard deviation of the market portfolio be* 32% *and let us assume that the market is efficient.*

*(a) What is the equation of capital market line?*
*(b) If Rs 300 is invested in the risk free asset, and Rs 700 in the market portfolio then what is the expected return at the end of the year?*
*(c) If an investor has Rs 1000 to invest and he/she desires a return of* 39%, *then what should be his/her portfolio?*

# 6
# Portfolio Optimization-II

## 6.1 Introduction

We have presented Markowitz's mean variance model for portfolio optimization in the last chapter. But contrary to its theoretical reputation, this model in its original form has not found much favor with the practitioners to construct large scale portfolios. There are several theoretical and practical reasons for not using this model extensively in practice - particularly when the number of assets in the portfolio is large. This chapter aims to understand these reasons and then present some other models which have been developed to improve Markowitz's model both theoretically and computationally.



## 6.2 Markowitz's Model: Some Theoretical and Computational Issues

Theoretically, Markowitz's model is known to be valid if returns $r_i'$s are multivariate normally distributed and the investor is *risk averse* in the sense that he/she prefers less standard deviation of the portfolio to more. But one is not fully convinced of the validity of the standard deviation as a measure of risk. An investor is certainly unhappy to have small or negative profit, but feels happy to have larger profit. In other words, this means that the investor's perception about risk is not symmetric about the mean. There are several empirical studies, which reveal that most $r_i$ are not normally or even symmetrically distributed. In this scenario, one possible approach seems to be to consider the skewness and kurtosis of the distribution in addition to the mean and variance and extend the Markowitz's model to generate the efficient frontier in (mean, variance, skewness, kurtosis)-space. This has been the approach of some of the models, e.g. Konno and Suzuki [79], and

Joro and Na [71], but has not found much favor because the resulting optimization problem is not easy to handle. An alternative and popular approach is to introduce certain new measures of risk which carry information about the possible portfolio losses implied by the tail of the return distribution, even in the case when the distribution is not symmetric. This takes care of those situations where the return distribution is heavily tailed. These risk measures are called *downside or safety-first risk measures* which aim to maximize the probability that the portfolio loss is below a certain acceptable level, commonly referred as the benchmark or the disaster level. Thus these risk measures are *quantile based risk measures* and are different from standard deviation or other *moment based risk measures*. Some of the most popular quantile based risk measures are value at risk (VaR) and conditional value at risk (CVaR). Since downside risk measures of individual securities cannot be easily aggregated into portfolio downside risk measures (we need the entire joint distribution of security returns), their application in practice requires computationally intensive non parametric estimation, simulation and optimization techniques.

There is another major problem associated with the classical Markowitz's model. This model gives us an optimal portfolio assuming that we have perfect information about $\mu_i$'s and $\sigma_{ij}$'s for the assets that we are considering. Therefore an important practical issue is the estimation of the $\mu_i$'s and $\sigma_{ij}$'s. A reasonable approach for estimating these data is to use time series of past returns $r_{it}$ which represents the return of $i^{th}$ asset from time $(t-1)$ to $t$, where $t = 1, 2, \ldots, T$. However, it has been observed that small changes in the time series $r_{it}$ lead to changes in the $\mu_i$'s and $\sigma_{ij}$'s that often lead to significant changes in the optimal portfolio. This is a fundamental weakness of the Markowitz model, no matter how cleverly $\mu_i$'s and $\sigma_{ij}$'s are computed. This is because the optimal portfolio construction is very sensitive to small changes in the data. Only one small change in one $\mu_i$ may produce a totally different portfolio. In fact recent research (Chopra and Ziemba [30]) has revealed that errors in the estimation of means $\mu_i$ can be more damaging than errors in other parameters. This has motivated researchers to employ *robust optimization techniques* in Markowitz's model, e.g. Ben-Tal [9], and Tütüncü and Koenig [139]. A much simpler approach is to consider portfolio optimization under a minimax rule (Cai et al. [23]) and provide some flexibility by allowing $\mu_i$ to lie in some interval $a_i \leq \mu_i \leq b_i$ (Deng et al. [36]).

The mean-variance model of Markowitz, in general, results in a dense quadratic programming problem. If the number of assets in the portfolio is large then it becomes very difficult to obtain an optimal solution of such large-scale dense quadratic programming problem on a real time basis. This has motivated researchers to consider mean-absolute deviation of the portfolio as a measure of risk (Konno and Yamazaki [80]) so that the resulting optimization problem reduces to a linear programming problem. These models are called $L_1$-*risk models* because these are based on $L_1$ metric on $\mathbf{R}^n$. In this terminology, Markowitz's model can be termed as a $L_2$-*risk model* since it uses variance as a risk measure which is based on the notion of $L_2$ metric. In the same spirit, the models based on the minimax rule portfolio selection strategies can be termed as $L_\infty$-*risk models*. It is simple to note that $L_1$ and $L_\infty$ risk models will result in a linear programming formulation. This is in contrast to an $L_2$- risk model which results in a quadratic programming formulation. Therefore computationally, $L_1$ and $L_\infty$ risk models are easier to handle in comparison to $L_2$ risk model. The $L_1$, $L_2$ and $L_\infty$ are three standard moment based risk models which have been studied in the literature.

The above discussion suggests that the choice of a proper risk measure is very crucial to study portfolio selection problems. But then what would be the guiding principles for this choice? The concept of *coherent risk measures* (Artzner et al. [5]) is an important contribution in this regard. This stipulates an axiomatic approach to the study of risk measure by presenting certain desirable properties.

Some other issues associated with Markowitz's model are very natural and these have been incorporated in the original model. These are with regard to its extension for the multi-period scenario and also to incorporate transaction costs. We shall discuss some of the models in the subsequent sections which addresses the above issues.



## 6.3 Mean Absolute Deviation Based Portfolio Optimization: A $L_1$-Risk Model

We first introduce the $L_1$-risk measure and formulate the corresponding portfolio optimization problem with this risk measure. Let an investor has initial wealth $M_0$ which is to be invested in n assets $a_i$ (i=1,2,...,n). Let $r_i$ be the return of the asset $a_i$ which is a random variable. Also let $x_i$ be the amount of money to be invested in the asset $a_i$ out of the total fund $M_0$.

We now define $\mu_i = E(r_i)$ and $q_i = E(|r_i - \mu_i|)$, i=1,2,...,n. Then $\mu_i$ denotes the expected return rate of the asset $a_i$ and $q_i$ denotes the expected absolute deviation of $r_i$ from its mean. Then the expected return of a portfolio $(x_1, x_2, \ldots, x_n)$ is given by

$$\mu = E\left(\sum_{i=1}^{n} r_i x_i\right) = \sum_{i=1}^{n} \mu_i x_i.$$

We now define $L_1$-risk measure or the mean absolute deviation of the portfolio $(x_1, x_2, \ldots, x_n)$.

**Definition 6.3.1 ($L_1$-Risk Measure of a Portfolio)** *Let $(x_1, x_2, \ldots, x_n)$ be the given portfolio. Then its $L_1$-risk measure or mean absolute deviation is defined as*

$$w_{L_1}(x_1, x_2, \ldots, x_n) = E\left[\left|\sum_{i=1}^{n} r_i x_i - E\left(\sum_{i=1}^{n} r_i x_i\right)\right|\right].$$

In terms of the $L_1$-risk measure $w_{L_1}(x_1, x_2, \ldots, x_n)$, the $L_1$-risk model of the portfolio optimization problem is formulated as

$$\text{Min} \qquad w_{L_1}(x_1, x_2, \ldots, x_n) = E\left[\left|\sum_{i=1}^{n} (r_i - \mu_i) x_i\right|\right]$$

subject to

$$\begin{aligned} \sum_{i=1}^{n} \mu_i x_i &\geq \alpha M_0 \\ \sum_{i=1}^{n} x_i &= M_0 \\ 0 \leq x_i &\leq u_i \qquad (i = 1, 2, \ldots, n), \end{aligned} \tag{6.1}$$

where $(\alpha > 1)$ is a parameter representing the minimum rate of return required by the investor. Also $u_i$ is the maximum amount of the money which can be invested in the asset $a_i$.

In (6.1) it may be noted that by defining $w_i = x_i / M_0$, $(i = 1, 2, \ldots, n)$, the portfolio $(x_1, x_2, \ldots, x_n)$ can also be represented in terms of its weight vector $(w_1, w_2, \ldots, w_n)$. Traditionally in the portfolio optimization literature, a portfolio has been represented by its weight vector $(w_1, w_2, \ldots, w_n)$ but it can also be equivalently represented in terms of its amount allocation vector $(x_1, x_2, \ldots, x_n)$.

In practice, the historical data is used to estimate the parameters in the optimization problem (6.1). Let $r_{it}$ be the realization of the random variable $r_i$ during the period t $(t = 1, 2, \ldots, T)$. We assume that $r_{it}$ is available through the historical data and the expected value of $r_i$ can be approximated by the average derived from the data. This gives

$$\mu_i = E(r_i) = \frac{1}{T} \sum_{t=1}^{T} r_{it}. \tag{6.2}$$

Also $w_L(x_1, x_2, \ldots, x_n)$ can be approximated by

$$E\left(\left|\sum_{i=1}^{n}(r_i-\mu_i)x_i\right|\right)=\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{n}(r_{it}-\mu_i)x_i\right|. \tag{6.3}$$

In (6.3) it may be noted that, due to the absolute value function, the expression on the right hand side becomes a nonlinear and non smooth function of $(x_1, x_2, \ldots, x_n)$.

Using (6.2) and (6.3), problem (6.2) can be reformulated as

$$\begin{aligned}
\text{Min} \quad & \frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{n}(r_{it}-\mu_i)x_i\right| \\
\text{subject to} \quad & \\
& \sum_{i=1}^{n}\mu_i x_i \geq \alpha M_0 \\
& \sum_{i=1}^{n} x_i = M_0 \\
& 0 \leq x_i \leq u_i \qquad (i = 1, 2, \ldots, n).
\end{aligned} \tag{6.4}$$

If we now denote $\left|\sum_{i=1}^{n}(r_{it}-\mu_i)x_i\right|$ by $y_t$ and employ the definition of the absolute function then we get

$$\begin{aligned}
y_t &= \left|\sum_{i=1}^{n}(r_{it}-\mu_i)x_i\right| \\
&= Max\left(\sum_{i=1}^{n}(r_{it}-\mu_i)x_i, \quad -\sum_{i=1}^{n}(r_{it}-\mu_i)x_i\right).
\end{aligned} \tag{6.5}$$

Therefore problem (6.4) gets transformed to

$$
\begin{aligned}
&\text{Min} \quad \frac{1}{T}\sum_{t=1}^{T} y_t \\
&\text{subject to} \\
&\qquad \sum_{i=1}^{n}(r_{it}-\mu_i)x_i \le y_t \\
&\qquad -\sum_{i=1}^{n}(r_{it}-\mu_i)x_i \le y_t \\
&\qquad \sum_{i=1}^{n}\mu_i x_i \ge \alpha M_0 \\
&\qquad \sum_{i=1}^{n} x_i = M_0 \\
&\qquad 0 \le x_i \le u_i \qquad (i = 1,2,\ldots,n),
\end{aligned}
\tag{6.6}
$$

where the first two constraints in (6.6) follow from (6.5) and the definition of maximum.
Denoting $(r_{it} - \mu_i)$ by $c_{it}$ $(i = 1,2,\ldots,n;\ t = 1,2,\ldots,T)$, we can rewrite (6.6) as

$$
\begin{aligned}
&\text{Min} \quad \frac{1}{T}\sum_{t=1}^{T} y_t \\
&\text{subject to} \\
&\qquad y_t - \sum_{i=1}^{n} c_{it}x_i \ge 0 \qquad (t = 1,2,\ldots,T) \\
&\qquad y_t + \sum_{i=1}^{n} c_{it}x_i \ge 0 \qquad (t = 1,2,\ldots,T) \\
&\qquad \sum_{i=1}^{n}\mu_i x_i \ge \alpha M_0 \\
&\qquad \sum_{i=1}^{n} x_i = M_0 \\
&\qquad 0 \le x_i \le u_i \qquad (i = 1,2,\ldots,n)\ .
\end{aligned}
\tag{6.7}
$$

**Remark 6.3.1** *The formulation (6.4) is essentially nonlinear and non smooth. But since this nonlinearity and non smoothness occurs due to the presence of absolute value function only, it can be handled in a reasonably simple manner* *as explained above. The resulting linear programming problem (6.7) can be solved efficiently even when n is large.*

One obvious question at this stage is to enquire if there is any relationship between $L_1$ and $L_2$-risk models. The below given theorem answers this question.

**Theorem 6.3.1** *Let $(r_1, r_2, \ldots, r_n)$ be multivariate normally distributed. Then for a given portfolio $x = (x_1, x_2, \ldots, x_n)$*

$$w_{L_1}(x) = \sqrt{\frac{2}{\pi}}\sigma(x),$$

*where the standard deviation $\sigma(x)$ is given by*

$$\sigma(x_1, x_2, \ldots, x_n) = \sqrt{E\left[\left\{\sum_{i=1}^{n} r_i x_i - E\left(\sum_{i=1}^{n} r_i x_i\right)\right\}^2\right]}.$$

**Proof.** Let $(\sigma_{ij}) \in \mathbf{R}^{n\times n}$ be the variance-covariance matrix of $(r_1, r_2, \ldots, r_n)$. Then under the given hypothesis, $\sum_{i=1}^{n} r_i x_i$ is normally distributed with mean $\sum_{i=1}^{n} r_i \mu_i$ and standard deviation

$$\sigma(x) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j}.$$



Therefore

$$\begin{aligned} w_{L_1}(x) &= \frac{1}{\sqrt{2\pi}\sigma(x)} \int_{-\infty}^{\infty} |\, u \,| \exp\left(-\frac{u^2}{2\sigma^2(x)}\right) du \\ &= \frac{2}{\sqrt{2\pi}\sigma(x)} \int_{0}^{\infty} u \, \exp\left(-\frac{u^2}{2\sigma^2(x)}\right) du\,, \end{aligned}$$

which on substitution $(u^2/2\sigma^2) = s$ gives

$$w_{L_1}(x) = \sqrt{\frac{2}{\pi}}\, \frac{\sigma^2(x)}{\sigma(x)} \int_0^{\infty} e^{-s} ds = \sqrt{\frac{2}{\pi}}\sigma(x).$$

□

In view of Theorem 6.3.1, for the multivariate normal case, both $L_1$ and $L_2$-risk models are equivalent. In other words, Markowitz's mean-variance portfolio selection strategy will be the same as the one given by mean-absolute deviation selection strategy. Even in the case when normality assumption does not hold, through certain case studies, it has been shown that minimizing the $L_1$-risk produces portfolios which are comparable to Markowitz's mean-variance model which minimizes $L_2$-risk, i.e. standard deviation. However as one expects the variance of mean-absolute deviation portfolio is always at least as large as the corresponding mean-variance portfolio. But in actual applications, this difference is small.

In fact Konno and Yamazaki [80] applied both $L_1$ and $L_2$-risk models in Tokyo Stock Market by using historical data of 224 stocks in NIKKEL 225 index. They generated efficient frontiers and observed that the difference of the standard deviation of the optimal portfolio generated by $L_2$ and $L_1$-risk models is at most $10\%$ for what ever value of $\alpha$. Of course two frontiers will coincide if $r_i$'s are multivariate normally distributed. Thus this difference can be largely attributed to the non normality of the data. Therefore irrespective of the distribution scenario, $L_1$-risk model provides a good alternative to Markowitz's $L_2$-risk model.

### Advantages of the Formulation (6.7)

We now list some of the advantages of $L_1$-risk model (6.7) over the classical Markowitz's model which minimizes the $L_2$-risk.

(i) The formulation (6.7) is a linear programming problem and hence can be solved much more efficiently in comparison to quadratic programming problem which is obtained in Markowitz's model. This is particulary significant for portfolios having large number of assets.

(ii) We do not need to calculate the variance-covariance matrix to set up the $L_1$-risk model.

(iii) In the formulation (6.7) there are always (2T+2) constraints regardless of the number of assets included in the model. This allows to handle very large portfolios on a real time basis.

(iv) An optimal solution $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ of problem (6.7) contains at most (2T+2) positive components if $u_i = +\infty$ for $i = 1, 2, \dots, n$. This means that an optimal portfolio consists of at most (2T+2) assets regardless of the size of $n$. Therefore we can use $T$ as a control variable when we wish to restrict the number of assets in the portfolio.

## 6.4 Minimax Rule Based Portfolio Optimization: An $L_\infty$-Risk Model

We discuss the *minimax rule based portfolio optimization model* due to Cai et al. [23]. We first introduce the $L_\infty$-risk measure which is based on $L_\infty$-metric on $\mathbf{R}^n$. We then discuss the corresponding portfolio optimization model, namely the $L_\infty$-risk model, and compare the same with Markowitz's mean-variance model.

We continue with the notations introduced in Section 6.3. Thus $M_0$ denotes the initial wealth to be invested in $n$ assets $a_i$ $(i = 1, 2, \ldots, n)$. Further for $(i = 1, 2, \ldots, n)$, $r_i$ denotes the return of the asset $a_i$ and $x_i$ denotes the allocation from $M_0$ in this asset. Then $\mu_i = E(r_i)$ and $q_i = E(|r_i - \mu_i|)$ respectively denote the expected return rate of the asset $a_i$ and the expected absolute deviation of $r_i$ from its mean $\mu_i$.

Let us now introduce the $L_\infty$-risk measure. For this let

$$F = \{x = (x_1, x_2, \ldots, x_n)\ :\ \sum_{i=1}^{n} x_i = M_0,\ x_i \geq 0\ (i = 1, 2, \ldots, n)\}. \tag{6.8}$$

Then $x \in F$ is a portfolio of the assets $a_1, a_2, \ldots, a_n$ and $F$ is called the *feasible region* of the portfolio optimization problem. As explained earlier, the expected return of the portfolio $x : (x_1, x_2, \ldots, x_n)$ is given by $\mu = \sum_{i=1}^{n} \mu_i x_i$. We now have the following definition.

**Definition 6.4.1** **($L_\infty$-Risk Measure)** *Let $x : (x_1, x_2, \ldots, x_n)$ be the given portfolio. Then its $L_\infty$-risk measure is defined as*

$$w_{L_\infty}(x_1, x_2, \ldots, x_n) = \underset{1 \leq i \leq n}{Max}\ E(\ |r_i x_i - \mu_i x_i|\ ) = \underset{1 \leq i \leq n}{Max}\ (q_i x_i).$$

In terms of the $L_\infty$-risk measure $w_{L_\infty}(x_1, x_2, \ldots, x_n)$, the $L_\infty$-risk model of the portfolio optimization problem, denoted by $(POL_\infty)$ is formulated as

$$\underset{x \in F}{\text{Min}}\ \left( \underset{1 \leq i \leq n}{\text{Max}}\ (q_i x_i),\ \ -\sum_{i=1}^{n} \mu_i x_i \right). \tag{6.9}$$

The problem (6.9) is a bi-criteria optimization problem as there are two objectives involved in the optimization. These two objectives are the portfolio risk and the portfolio return. This scenario is again same as for the $L_1$ and $L_2$-risk models studied earlier. There we had minimized the portfolio risk for a given level of portfolio return. Following the same strategy, we get the following problem

$$\begin{aligned}&\operatorname*{Min}_{x\in F}\ \left(\operatorname*{Max}_{1\le i\le n}\ (q_i x_i)\right)\\ &\text{subject to}\\ &\qquad \sum_{i=1}^{n} \mu_i x_i \ge \alpha M_0.\end{aligned} \tag{6.10}$$

Writing $y = \operatorname*{Max}_{1\le i\le n} (q_i x_i)$, we rewrite (6.11) as

$$\begin{aligned}&\operatorname*{Min}_{x\in F}\ \ y\\ &\text{subject to}\\ &\qquad \sum_{i=1}^{n} \mu_i x_i \ge \alpha M_0\\ &\qquad\qquad q_i x_i \le y \qquad (i = 1, 2, \ldots, n).\end{aligned} \tag{6.11}$$

For each $\alpha$, (6.11) is a linear programming problem and hence can be solved efficiently. However to generate the entire efficient frontier we need to solve (6.11) for every $\alpha$, which is practically not possible. In the case of Markowitz's $L_2$-risk model, the famous two fund theorem comes to our rescue which allows to generate the entire efficient frontier from the knowledge of only two efficient points. In the case of $L_\infty$-risk model, we do not have any analogue of two fund theorem, so we look to some other suitable option which does not require simplex algorithm to solve (6.11) for every $\alpha$, but rather gives a solution in close form for every level of return. This is the approach which we shall be following now and describe the details below.

The solution of the bi-criteria optimization problem (6.9) is to be understood in terms of an efficient point which in this context is termed as an *efficient portfolio.*

**Definition 6.4.2 (Efficient Portfolio)** *A feasible portfolio* $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) \in F$ *is said to be an efficient portfolio if there does not exist any other portfolio* $x = (x_1, x_2, \ldots, x_n) \in F$ *such that*

$$(i) \qquad \operatorname*{Max}_{1\le i\le n}\ (q_i \bar{x}_i) \le \operatorname*{Max}_{1\le i\le n} (q_i x_i),$$

*and*

$$(ii) \qquad \sum_{i=1}^{n}\ \mu_i \bar{x}_i\ \ge\ \sum_{i=1}^{n}\ \mu_i x_i.$$

The collection of all *efficient portfolios* is called the *efficient frontier*. As explained in the context of Markowitz's model, an investor is always interested in determining the efficient frontier so as to select that portfolio which gives maximum return for the chosen level of risk. This principle remains valid for $L_\infty$-risk model as well.

The problem $(POL_\infty)$ for the $L_\infty$-risk model can be rewritten as

$$\begin{aligned} \text{Min} \quad & (y\,,\ -\sum_{i=1}^{n} \mu_i x_i) \\ \text{subject to} \quad & \\ & q_i x_i \le y \qquad (i = 1, 2, \ldots, n) \\ & x \in F. \end{aligned} \tag{6.12}$$

Now borrowing the standard technique from the multi-objective literature, we introduce the following parametric optimization problem $PO(\lambda)$ for $0 < \lambda < 1$

$$\begin{aligned} \text{Min} \quad & \lambda y + (1-\lambda)(-\sum_{i=1}^{n} \mu_i x_i) \\ \text{subject to} \quad & \\ & q_i x_i \le y \qquad (i = 1, 2, \ldots, n) \\ & x \in F. \end{aligned} \tag{6.13}$$



We now have the following lemma connecting problems $(POL_\infty)$ and $PO(\lambda)$.

**Lemma 6.4.1.** *The pair* $(\bar{x}, \bar{y})$ *is an efficient solution of* $(POL_\infty)$ *if and only if there exists* $0 < \lambda < 1$ *such that* $(\bar{x}, \bar{y})$ *is an optimal solution of* $PO(\lambda)$.

Here $\lambda$ can be considered as an investor's risk tolerance parameter which helps to do appropriate trade-off between the risk and the return.

In view of Lemma 6.4.1, finding efficient frontier of $(POL_\infty)$ is equivalent to finding solution of $PO(\lambda)$ for all $0 < \lambda < 1$. The discussion given below illustrates that given an arbitrary $\bar{\lambda} \in (0, 1)$, the solution $(\bar{x}, \bar{y})$ of $PO(\bar{\lambda})$ can be obtained explicitly in a close form and there is no need of applying any simplex like numerical optimization algorithm. This greatly reduces the computational burden in determining the efficient frontier.

**Analytical Solution of** $PO(\lambda)$

Without any loss of generality we can assume that (i) $\mu_1 \le \mu_2 \le \ldots \le \mu_n$ and (ii) there do not exist two assets $a_i$ and $a_j$, $i \ne j$, such that $\mu_i = \mu_j$ and $q_i = q_j$. There is obviously no problem with the ordering of asset returns $\mu_i$ and therefore the first assumption can always be met. The second assumption is also valid because if $\mu_i = \mu_j$ and $q_i = q_j$ for two assets $i$, $j$ $(i \neq j)$, then we may treat them as a single aggregated asset. We first consider the case when all assets are risky assets.

**Theorem 6.4.1** *Let all asset $a_1, a_2, \ldots, a_n$ be risky assets. Then for any $0 < \lambda < 1$, an optimal solution $(x^*, y^*)$ of the problem $PO(\lambda)$ is given by*

$$x_i^* = \begin{cases} y^*/q_i & i \in T^*(\lambda) \\ 0 & \text{otherwise,} \end{cases}$$

*where*

$$y^* = M_0 \left( \sum_{l \in T^*(\lambda)} (1/q_l) \right)^{-1},$$

*and the set $T^*(\lambda)$ is identified by the following rule.*

*(a)*

$$\begin{aligned} \frac{(\mu_n - \mu_{n-1})}{q_n} &< \frac{\lambda}{1-\lambda} \\ \frac{(\mu_n - \mu_{n-2})}{q_n} + \frac{(\mu_{n-1} - \mu_{n-2})}{q_{n-1}} &< \frac{\lambda}{1-\lambda} \\ &\vdots \\ \frac{(\mu_n - \mu_{n-k})}{q_n} + \frac{(\mu_{n-1} - \mu_{n-k})}{q_{n-1}} + \ldots + \frac{(\mu_{n-k+1} - \mu_{n-k})}{q_{n-k+1}} &< \frac{\lambda}{1-\lambda}, \end{aligned}$$

*and*

$$\frac{(\mu_n - \mu_{n-k-1})}{q_n} + \frac{(\mu_{n-1} - \mu_{n-k-1})}{q_{n-1}} + \ldots + \frac{(\mu_{n-k} - \mu_{n-k-1})}{q_{n-k}} \geq \frac{\lambda}{1-\lambda}.$$

*Then*

$$T^*(\lambda) = \{n, n-1, \ldots, n-k\}. \tag{6.14}$$

*(b) Otherwise*

$$T^*(\lambda) = \{n, n-1, \ldots, 1\}. \tag{6.15}$$

**Proof.** We apply the Karush Kuhn-Tucker (KKT) conditions to $PO(\lambda)$. For this, we first introduce the Lagrangian of $PO(\lambda)$ as

$$L(x, y, \beta, \delta, \gamma) = \lambda y + (1-\lambda)\Big(-\sum_{i=1}^{n}\mu_i x_i\Big) + \sum_{i=1}^{n}\beta_i(q_i x_i - y) + \delta\Big(\sum_{i=1}^{n}x_i - M_0\Big) - \sum_{i=1}^{n}\gamma_i x_i\,,$$

where $x \in \mathbf{R}^n$, $y \in \mathbf{R}$, $\beta \in \mathbf{R}^n$, $\delta \in \mathbf{R}$ and $\gamma \in \mathbf{R}^n$. We now write the KKT conditions that an optimal solution of $PO(\lambda)$ must satisfy. These are

$$\frac{\partial L}{\partial y} = \lambda - \sum_{i=1}^{n}\beta_i = 0 \tag{6.16}$$

$$\frac{\partial L}{\partial x_i} = -(1-\lambda)\mu_i + \beta_i q_i + \delta - \gamma_i = 0 \qquad (i = 1, \ldots, n) \tag{6.17}$$

$$\sum_{i=1}^{n} x_i = M_0 \tag{6.18}$$

$$(q_i x_i - y)\beta_i = 0 \qquad (i = 1, \ldots, n) \tag{6.19}$$

$$\gamma_i x_i = 0 \qquad (i = 1, \ldots, n) \tag{6.20}$$

$$\beta_i \geq 0 \qquad (i = 1, \ldots, n) \tag{6.21}$$

$$\gamma_i \geq 0 \qquad (i = 1, \ldots, n). \tag{6.22}$$

Our aim now is to obtain a solution of the above system. Define $T^*(\lambda) = \{i : \beta_i > 0\}$. This means that for (6.19) to be satisfied we must have

$$q_i x_i - y = 0 \quad \text{for } i \in T^*(\lambda)\ ,$$

i.e.

$$x_i = \frac{y}{q_i}, \quad \text{for } i \in T^*(\lambda). \tag{6.23}$$

We let

$$x_i = 0, \quad \text{for } i \notin T^*(\lambda). \tag{6.24}$$

But $M_0 = \sum_{i=1}^{n} x_i = \sum_{i \in T^*(\lambda)} x_i + \sum_{i \notin T^*(\lambda)} x_i = \sum_{i \in T^*(\lambda)} x_i$, and hence (6.23) gives

$$M_0 = y \sum_{i \in T^*(\lambda)} (1/q_i) \tag{6.25}$$

i.e.

$$y = M_0 \left( \sum_{l \in T^*(\lambda)} \frac{1}{q_l} \right)^{-1}. \tag{6.26}$$

Therefore, we have

$$x_i^* = \begin{cases} \dfrac{M_0}{q_i} \left( \displaystyle\sum_{l \in T^*(\lambda)} \frac{1}{q_l} \right)^{-1}, & i \in T^*(\lambda) \\ 0, & \text{otherwise.} \end{cases} \tag{6.27}$$

From (6.20), for $x_i > 0$ we have $\gamma_i = 0$. Hence $\gamma_i = 0$ whenever $i \in T^*(\lambda)$. Thus from (6.17), we have for $i \in T^*(\lambda)$

$$\beta_i = \frac{1}{q_i} \left[ (1 - \lambda)\mu_i - \delta \right]. \tag{6.28}$$

From (6.16) and (6.28), we get

$$\lambda = \sum_{l \in T^*(\lambda)} \frac{1}{q_l} \left[ (1 - \lambda)\mu_l - \delta \right] = (1 - \lambda) \sum_{l \in T^*(\lambda)} (\mu_l / q_l) - \delta \sum_{l \in T^*(\lambda)} (1/q_l).$$

Therefore we get

$$\delta = \left( \sum_{l \in T^*(\lambda)} \frac{1}{q_l} \right)^{-1} \left( (1 - \lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_l}{q_l} - \lambda \right). \tag{6.29}$$

Thus, from (6.28) for $i \in T^*(\lambda)$

$$\beta_i = \frac{1}{q_i} \left[ (1 - \lambda)\mu_i - \left( \sum_{l \in T^*(\lambda)} \frac{1}{q_l} \right)^{-1} \left( (1 - \lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_l}{q_l} - \lambda \right) \right]. \tag{6.30}$$

Also from (6.17) and noting that for i $\notin T^*(\lambda)$, $\beta_i = 0$, we have for i $\notin T^*(\lambda)$,

$$\gamma_i = -(1 - \lambda)\mu_i + \delta. \tag{6.31}$$

Clearly, if it is possible to determine the set $T^*(\lambda)$ which ensures that $\beta_i$ and $\gamma_i$ as given by (6.30) and (6.31) are all non negative, then $y$ and $x_i$ as given by (6.26) and (6.27), respectively, will be a solution of KKT conditions (6.16) to (6.22).

Now to prove our theorem, it is required that if there exists an integer $k \in [0, n-2]$ such that inequalities specified in $(a)$ hold, then $T^*(\lambda)$ as given by (6.14) will ensure that $\mu_i \geq 0$ and $\gamma_i \geq 0$. The following analysis proves the above arguments.

By (6.30), it follows that, for any $i \in T^*(\lambda) = \{n, n-1, \ldots, n-k\}$,

$$\beta_i = \left(\sum_{l \in T^*(\lambda)} \frac{q_i}{q_l}\right)^{-1} \left[(1-\lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_i - \mu_l}{q_l} + \lambda\right]$$
$$= \left(\sum_{l \in T^*(\lambda)} \frac{q_i}{q_l}\right)^{-1} (1-\lambda) \left[\left(-\sum_{l=i+1}^{n} \frac{\mu_l - \mu_i}{q_l} + \frac{\lambda}{(1-\lambda)}\right) + \sum_{l=n-k}^{i} \frac{\mu_i - \mu_l}{q_l}\right],$$

which by inequalities in $(a)$, is strictly positive. On the other hand, for $i = 1, 2, \ldots, (n-k-1)$, from (6.31) and (6.29), we have

$$\gamma_i = -(1-\lambda)\mu_i + \left(\sum_{l \in T^*(\lambda)} \frac{1}{q_l}\right)^{-1} \left((1-\lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_l}{q_l} - \lambda\right)$$
$$= \left(\sum_{l \in T^*(\lambda)} \frac{1}{q_l}\right)^{-1} \left((1-\lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_l - \mu_i}{q_l} - \lambda\right)$$
$$\geq \left(\sum_{l \in T^*(\lambda)} \frac{1}{q_l}\right)^{-1} \left((1-\lambda) \sum_{l \in T^*(\lambda)} \frac{\mu_l - \mu_{n-k-1}}{q_l} - \lambda\right),$$

which by (6.14), is nonnegative. The above details show that the KKT conditions (6.21) and (6.22) are satisfied. In the case when there does not exist any integer $k \in [0, n-2]$ such that inequalities in $(a)$ are satisfied, we need to show that the solution given by (6.26) and (6.27) with $T^*(\lambda) = \{n, n-1, \ldots, 2, 1\}$, will satisfy all KKT conditions. To do this, we introduce a dummy asset $a_0$ with $\mu_0 = -L$ and $q_0 = L$, where L is a sufficiently large positive number. Now following in the same manner as above, we can show that all KKT conditions are satisfied.

Thus above analysis shows that the KKT conditions (6.16) to (6.22) are satisfied, if we select $T^*(\lambda)$ by the selection procedure (6.14) to (6.15) and solutions $y$ and $x_i$ are given by (6.26) and (6.27). Now since $PO(\lambda)$ is a convex programming problem the KKT conditions become necessary and sufficient for optimality. So solution given by (6.12) to (6.15) is an optimal solution. This proves the theorem. □

We next consider the case when one risk-free asset is included in the portfolio. To be specific, let the asset $a_1$ be risk-free. It is natural to assume that the risk-free asset has the lowest return. This is because choosing assets having lower returns than that of the risk-free asset in the portfolio can not make it optimal.

Now for the risk-free asset $a_1$, we have $q_1 = 0$. Hence Theorem 6.4.1 can not be directly applied in this scenario. However, we can take $q_1 = \epsilon$ where $\epsilon > 0$ is a small number. Now we can apply Theorem6.4.1 and get the desired result by taking $\epsilon \to 0^+$.

Taking $q_1 = \epsilon(> 0)$, when we apply Theorem 6.4.1 to determine the set $T^*(\lambda)$,there are two possibilities. These are (i) $1 \notin T^*(\lambda)$ and (ii) $1 \in T^*(\lambda)$.

When $1 \notin T^*(\lambda)$, the situation is exactly same as discussed earlier and so the optimal solution for $PO(\lambda)$ as given in Theorem 6.4.1 remains unchanged.

When $1 \in T^*(\lambda)$. Then by Theorem 6.4.1, the optimal solution of $PO(\lambda)$ is given by

$$x_i^* = \begin{cases} \dfrac{y^*}{q_i}, & i \in T^*(\lambda) \\ 0, & i \notin T^*(\lambda), \end{cases} \tag{6.32}$$

where

$$y^* = M_0 \left( \frac{1}{\epsilon} + \sum_{l \neq 1}^{n} \frac{1}{q_l} \right)^{-1}. \tag{6.33}$$



In (6.32) and (6.33), if we now take the limit as $\epsilon \to 0^+$, we obtain $x_i^* = 0$ for all $i > 1$, $x_1^* = M_0$ and $y^* = 0$.

An obvious question now is to state certain condition so that the two cases, namely $1 \notin T^*(\lambda)$ and $1 \in T^*(\lambda)$, can be verified easily. In this context, we state the following condition

$$\frac{(\mu_n - \mu_1)}{q_n} + \frac{(\mu_{n-1} - \mu_1)}{q_{n-1}} + \ldots + \frac{(\mu_2 - \mu_1)}{q_2} < \frac{\lambda}{(1 - \lambda)}. \tag{6.34}$$

Here $\mu_1$ is the risk-free return. For the risk-free asset $a_1$, $q_1 = 0$ but it does not enter in (6.34).

Form the statement of Theorem 6.4.1 and related condition for the determination of $T^*(\lambda)$ for $k = (n - 2)$ it is clear that if the condition (6.34) is not satisfied then $1 \notin T^*(\lambda)$. Otherwise if condition (6.34) is satisfied then $1 \in T^*(\lambda)$. Therefore the above discussion leads to the following theorem

**Theorem 6.4.2** *Let* $\lambda \in (0,1)$ *be given. If the condition (6.34) is not satisfied then* $T^*(\lambda)$ *and associated* $(x^*, y^*)$ *should be determined as per Theorem 6.4.1. Otherwise if (6.34) is satisfied, then all wealth* $M_0$ *should be invested in the risk free asset, i.e.* $x_1^* = M$, $x_i^* = 0$, $i \neq 1$ *and* $y^* = 0$.

**Certain Observations about the Solution of Problem** $PO(\lambda)$

(i) Consider the solution where

$$\frac{(\mu_n - \mu_{n-1})}{q_n} < \frac{\lambda}{(1-\lambda)},$$

and

$$\frac{(\mu_n - \mu_{n-2})}{q_n} + \frac{(\mu_{n-1} - \mu_{n-2})}{q_{n-1}} \geq \frac{\lambda}{(1-\lambda)}.$$

Then Theorem 6.4.1 tells that the optimal portfolio should consist of assets $a_n$ and $a_{n-1}$ only. Further, the actual amounts of investments for these assets should be

$$x_n^* = \frac{M_0}{q_n}\left(\frac{1}{q_{n-1}} + \frac{1}{q_n}\right)^{-1} = M_0\left(1 + \frac{q_n}{q_{n-1}}\right)^{-1},$$

and

$$x_{n-1}^* = \frac{M_0}{q_{n-1}}\left(\frac{1}{q_{n-1}} + \frac{1}{q_n}\right)^{-1} = M_0\left(1 + \frac{q_{n-1}}{q_n}\right)^{-1}.$$

Therefore, if $q_n$ is much larger than $q_{n-1}$, then it is possible that $x_n^*$ is nearly zero while $x_{n-1}^*$ is nearly equal to $M_0$.

The above discussion suggests that the process of constructing an optimal portfolio is a *two phase decision process*. In the first phase, the assets are selected according to their rates of return, thereby giving the set of *investable assets*. In the second phase, the actual amounts allocated to the investable assets (those assets which have been selected in Phase-1) are determined based on their risk levels. Here if the risk of any investable asset is very high, then its wealth allocation is very small and therefore it may be neglected in the optimal portfolio. Thus in Phase-1, an asset may be eliminated if its return is very low, while in Phase-2, it may be eliminated if its risk is very high.

(ii) For the optimal portfolio $x^*$ as given by Theorem 6.4.1, we have

$$x_i^* q_i = \begin{cases} y^*, & i \in T^*(\lambda) \\ 0, & \text{otherwise.} \end{cases} \tag{6.35}$$

Since $q_i x_i^*$ is the risk associated with the $i^{th}$ asset when the amount $x_i^*$ is investable in it, (6.35) tells that for the assets selected for investment, we invest them with the amounts such that they have the same risk $y^*$. Thus the optimal solution of $PO(\lambda)$ gives an investment plan in which we invest a small amount for assets having high risk, and invest a large amount for assets having low risk. This strategy will not increase the maximum risk but certainly increase the overall expected return.

(iii) The amounts $x_i^*$ $(i \in T^*(\lambda))$ as given in Theorem 6.4.1 does not depend on $\mu_i$ as long as the set $T^*(\lambda)$ is selected. Thus the information about the expected returns $\mu_i$ is used to determine the set of investable assets only. Later only risks of the investable assets are used to determine the actual allocation of wealth in the investable assets.

(iv) In Theorem 6.4.1, the inequalities are used to define the ranking rule for selecting the investable assets. Because of the presence of these inequalities, this model has some robustness and is not that much sensitive against errors in the parameters $\mu_i$ $(i = 1, 2, \ldots, n)$. In fact Cai et al. [23] gave an explicit result for allowable perturbation $\delta_i$ in $\mu_i$ so that the optimal portfolio as obtained by Theorem 6.4.1 remains unchanged.

We now present some illustrative examples to verify the above points.

**Example 6.4.1** *Consider a six asset* $(a_1, a_2, a_3, a_4, a_5, a_6)$ *portfolio optimization problem* $PO(\lambda)$ *with the following data*



$$
\begin{aligned}
\mu &= col(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (1, 3, 5, 6, 7, 9)^T \\
q &= col(q_1, q_2, q_3, q_4, q_5, q_6) = (5, 10, 6, 7, 400, 8) \\
M_0 &= Rs\ 60{,}000 \\
\lambda &= 0 \cdot 35.
\end{aligned}
$$

*Use Theorem 6.4.1 to determine the optimal portfolio.*

**Solution** Here the asset returns $\mu_i$ are already in the ascending order. Next we note that

$$
\frac{\lambda}{(1-\lambda)} = 0.5384,
$$

and

$$\frac{(\mu_6 - \mu_5)}{q_6} = \frac{(9-7)}{8} = 0 \cdot 25 < 0 \cdot 5384$$

$$\frac{(\mu_6 - \mu_4)}{q_6} + \frac{(\mu_5 - \mu_4)}{q_5} = \frac{(9-6)}{8} + \frac{(7-6)}{400} = 0 \cdot 3775 < 0 \cdot 5384$$

$$\frac{(\mu_6 - \mu_3)}{q_6} + \frac{(\mu_5 - \mu_3)}{q_5} + \frac{(\mu_4 - \mu_3)}{q_4} = \frac{(9-5)}{8} + \frac{(7-5)}{400} + \frac{(6-5)}{7} = 0 \cdot 647 > 0 \cdot 5384.$$
$$= 0 \cdot 647 > 0 \cdot 5384.$$

Hence $(n-k) = 4$, which gives $k = 6 - 4 = 2$. This gives

$$T^*(0.35) = \{n, n-1, \ldots, n-k\} = \{6, 5, 4\},$$

and therefore the investable assets are $a_4, a_5$ and $a_6$.

Next we determine the allocation in these investable assets using the formulae given in Theorem 6.4.1, we have

$$x_4^* = \frac{60,000}{7}\left(\frac{1}{7} + \frac{1}{400} + \frac{1}{8}\right)^{-1} = 31704 \cdot 09$$
$$x_5^* = \frac{60,000}{400}\left(\frac{1}{7} + \frac{1}{400} + \frac{1}{8}\right)^{-1} = 554 \cdot 82$$
$$x_6^* = \frac{60,000}{8}\left(\frac{1}{7} + \frac{1}{400} + \frac{1}{8}\right)^{-1} = 27741 \cdot 08,$$

and

$$y^* = 60,000\left(\frac{1}{7} + \frac{1}{400} + \frac{1}{8}\right)^{-1} = 221928 \cdot 66.$$

Therefore the optimal portfolio consists of the allocation $x_1^* = 0$, $x_2^* = 0$, $x_3^* = 0$, $x_4^* = 31704 \cdot 09$, $x_5^* = 554 \cdot 82$ and $x_6^* = 27741 \cdot 08$ from the given wealth $M_0 = 60,000$. Further the minimum value of the $L_\infty$-risk measure for the optimal portfolio $x^*$, i.e. $w_{L_\infty}(x^*)$ is $y^* = 221928 \cdot 66$. In fact $y^* = q_4x_4^* = q_5x_5^* = q_6x_6^* = 221928 \cdot 66$ and therefore all investable assets with the chosen optimal allocation have the same risk $y^*$. Further the maximum value of the portfolio expected return is

$$\sum_{i=1}^{n} \mu_i x_i^* = \mu_4 x_4^* + \mu_5 x_5^* + \mu_6 x_6^* = 443778 \cdot 0.$$

Here we see that the returns of the asset $a_1$, $a_2$ and $a_3$ are small in comparison to $a_4$, $a_5$ and $a_6$ and they are not selected for the investment. This is precisely the

Phase-1 of the investable assets selection procedure. Further, as risk of the asset $a_5$ is very large, it is therefore allocated a very small amount of investment. This is the Phase-2 of the investable selection procedure.

□

### Some More Observations on $L_\infty$-Risk Model

(i) Looking at the definition of the $L_\infty$ risk measure $w_{L_\infty}(x)$ or the formula for the optimal allocation $x^*$, one gets the impression that none of these depend on the covariances between the assets. Further it seems that only the risks of the individual assets, rather than the risk of the entire portfolio, are taken care off in this development. But this is not true as being explained below.

Let us recall that the total portfolio risk is given by $E\left(\left|\sum_{i=1}^{n} r_i x_i - \sum_{i=1}^{n} \mu_i x_i\right|\right)$.

Let $\xi$ be a given positive number. We are interested in making $P\left(\left|\sum_{i=1}^{n} r_i x_i - \sum_{i=1}^{n} \mu_i x_i\right| \geq \xi\right)$ as small as possible because this ensures that the deviation of the actual total return from the expected total return is as small as possible. But by Markov inequality we have

$$\begin{aligned} P(\mid \sum_{i=1}^{n} r_i x_i - \sum_{i=1}^{n} \mu_i x_i \mid \geq \xi) &\leq \frac{1}{\xi} E(\mid \sum_{i=1}^{n} r_i x_i - \sum_{i=1}^{n} \mu_i x_i \mid) \\ &\leq \frac{1}{\xi} \sum_{i=1}^{n} E(\mid r_i - \mu_i \mid x_i) \\ &\leq \frac{n}{\xi} w_{L_\infty}(x). \end{aligned} \tag{6.36}$$

The inequality (6.36) illustrates that the total portfolio risk is small if $w_{L_\infty}(x)$ is kept small. But in the expression for the total portfolio risk, the covariances among various assets are certainly involved. This explains that in the $L_\infty$-risk model, the covariances are not ignored and minimizing $w_{L_\infty}(x)$ indirectly attempts minimizing the total portfolio risk.

(ii) In analogy with Markowitz's mean-variance model, it is natural to ask certain questions with regard to the efficient frontier and the status of CAPM type model for $L_\infty$-risk model. Cai et al. [23] developed an explicit procedure for tracing the efficient frontier for $L_\infty$-risk model but did not present any CAPM type result.

(iii) For Markowitz's model we can generate efficient frontier easily provided short selling is allowed (we refer to the *two fund theorem* in this regard). For $L_\infty$-risk model discussed here, generation of efficient frontier becomes easier provided short selling is *not* allowed (we refer to Theorem 6.4.1 in this regard).

(iv) Cai et al. [23] $L_\infty$-risk model can be made more robust if more flexibility is allowed with regard to the parameters $\mu_i$ $(i = 1, 2, \ldots, n)$. Deng et al. [36] presented a minimax type model for portfolio optimization where $\mu_i' s$ are allowed to be in the interval, $a_i \leq \mu_i \leq b_i$ (i=1,2,. . . ,n). They used the celebrated *minimax theorem* along with the results of Cai et al. [23] to solve the resulting optimization problem.

So far we have considered certain variations of Markowitz's mean-variance model where the portfolio risk is taken different from the standard deviation (e.g. $L_1$-risk or $L_\infty$-risk) but it is still essentially based on moments of the portfolio return. In the coming sections, we shall discuss few variations of Markowitz's model where portfolio risk is quantile based. In particular, we discuss VaR and CVaR based portfolio optimization problems.

## 6.5 Value-at-Risk of an Asset

Before defining *value-at-risk* of an asset (denoted by VaR), we consider the below given example for the sake of motivation.

Consider two time instances $t = 0$ and $t = T$. Let $S(0) = 100$ and assume that we buy a share of stock at $t = 0$ to sell it at $t = T$. As S(T) is a random variable, we cannot predict the quantum of profit/loss. Obviously we shall suffer a loss if $S(T) < 100\, e^{rT}$, where $r$ is the risk-free rate under continuous compounding.

A natural question at this stage could be to determine the probability that loss is less than or equal to a specified amount, say Rs 20, i.e. $P[(100\, e^{rT} - S(T)) \leq 20]$. This probability can be easily be computed once the probability distribution of S(T) is known. But it becomes more interesting if this question is reversed. Here we fix the probability, say 95%, and seek the *amount* such that the probability of loss not exceeding *this amount* is more than or equal to 95%. The amount so obtained is essentially called the value-at-risk of the given asset. Thus

$$P[(100\, e^{rT} - S(T)) \leq VaR] \geq 0 \cdot 95\ , \tag{6.37}$$

and VaR represents the predicted maximum loss with specified probability (0.95 in our example) over a certain period of time which is $T$ in our case.

The expression (6.37) suggests that VaR is a measure related to percentiles of loss distributions. In certain sense, it tries to answer the basic question which every investor seems to ask at some point in time, namely, what is the most he/she can lose on his/her investments? Value-at-Risk tries to answer this question within a reasonable bound. VaR and also other downside risk measures are very useful in assessing the risk for securities with asymmetric return distributions, such as call and put options.

VaR has been developed by JP Morgan, and made available through Risk Metrics software in October 1994. We now give the general definition of VaR for a random variable $X$. This random variable may represent the loss distribution of the asset return with $-X$ being represented as gain.

**Definition 6.5.1 (Value at Risk)** *Let $X$ be the given random variable and $\alpha$ be the given probability level. Then the VaR of $X$ with confidence level $(1-\alpha)$, $0 < \alpha < 1$, denoted by $VaR_{(1-\alpha)}(X)$ is defined as*

$$VaR_{(1-\alpha)}(X) = Min\{z : \ F_X(z) \geq (1-\alpha)\} = F_X^{-1}(1-\alpha) \ , \tag{6.38}$$

*where $F_X$ denotes the cumulative distribution function of the random variable $X$.*

In view of the above definition $VaR_{(1-\alpha)}(X)$ is a lower $(1-\alpha)$ percentile of the random variable X. If we substitute the expression for $F_X(z)$ in (6.38), we obtain

$$VaR_{(1-\alpha)}(X) = \text{Min}\{z : \ P(X \leq z) \geq (1-\alpha)\}. \tag{6.39}$$

Thus $VaR_{(1-\alpha)}(X)$ for an asset is the value $z$ such that the probability that the maximum loss $X$ is at most $z$, is at least $(1-\alpha)$.

The use of VaR involves two chosen parameters. These are confidence level $(1-\alpha)$ and the holding period T of the asset. The choice of $\alpha$, and hence$(1-\alpha)$, depends on the purpose to which our risk measure is utilized. In practice $\alpha$ is typically taken as $10\%,\ 5\%$ *and* $1\%$, so that the typical confidence levels are $90\%,\ 95\%$ *and* $99\%$. The usual holding periods are one day or one month, but it can be even one quarter or more. Given the confidence level $(1-\alpha)$ and horizon T, VaR is a bound such that the loss over the horizon is less than this bound with probability equal to the confidence coefficient. For example, if horizon is one week, the confidence level is $99\%$ (so $\alpha = 0 \cdot 01$) and VaR is Rs 50,000, then there is only a 1% chance of loss exceeding Rs 50,000 over the next week.

From (6.39) we note that $VaR_{(1-\alpha)}(X) = F_X^{-1}(1-\alpha)$ and therefore for a continuous loss distribution $VaR_{(1-\alpha)}(X)$ is simply the loss such that

$$P(X \leq VaR_{(1-\alpha)}(X)) = (1-\alpha).$$

Let us recall that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then the $q$-percentile of $X$ is $\mu + \sigma\Phi^{-1}(q)$, where $\Phi$ is the standard normal density function. Therefore

$$VaR_{(1-\alpha)}(X) = \mu + \sigma\Phi^{-1}(1-\alpha) \ ,$$

where $(1-\alpha)$-percentile of $\Phi$, namely $\Phi^{-1}(1-\alpha)$, is that value of the standard normal variate for which the area in the left is $(1-\alpha)$.

The above relation shows that for normally distributed random variables, VaR is proportional to the standard deviation.

**Example 6.5.1** *Suppose that the stock price is lognormal with mean* 12% *and standard deviation* 30%. *Let the interest rate r for the period* $t = 0$ *to* $t = 1$ *be* 8% *and stock price S(0) at time t=0 be Rs 100. Determine the VaR for the given stock at* 95% *confidence level.*

**Solution** We are given that

$$ln\Big(\frac{S(1)}{S(0)}\Big) \ \sim \ \mathcal{N}(0.12, (0.30)^2) \ ,$$

and our aim is to find VaR for $(1-\alpha) = 0.95$, i.e. for $\alpha = 0.05$. By definition, loss equals $(100e^{0.08} - S(1))$ and

$$P[(100e^{0.08} - S(1)) \leq VaR] = 0.95. \tag{6.40}$$

Now from the given information

$$\frac{ln\Big(\frac{S(1)}{S(0)}\Big) - 0.12}{0.30} \ \sim \ \mathcal{N}(0,1). \tag{6.41}$$

Therefore (6.40) gives

$$P\left[\frac{S(1)}{S(0)} \geq \frac{(100e^{0.08} - VaR)}{100}\right] = 0.95$$

i.e.

$$P\left[ln\Big(\frac{S(1)}{S(0)}\Big) \geq ln\frac{(100e^{0.08} - VaR)}{100}\right] = 0.95$$

i.e.

$$P\left[\frac{ln\left(\frac{S(1)}{S(0)}\right)-0.12}{0.30} \geq \frac{ln\left(\frac{(100e^{0.08}-VaR)}{100}\right)-0.12}{0.30}\right] = 0.95. \tag{6.42}$$

But from (6.41) and (6.42), we get

$$\frac{ln\left(\frac{100e^{0\cdot08}-VaR}{100}\right)-0.12}{0.30} = -1.645,$$

which determines

$$VaR = 100(e^{0.08} - e^{-0.3750}) = Rs\ 39.50.$$

□

**Example 6.5.2** *Let an investment A return a gain of* Rs 100 *with probability* 0.96 *and a loss of* Rs 200 *with probability* 0.04. *Obtain VaR of the given investment with* 95% *confidence level.*

**Solution** We note that our definition of VaR is based on the probability distribution of the random variable X, where X represents the loss of the given investment. Therefore from the given information we have

$$X = \begin{cases} -100, & \text{with probability} \quad 0\cdot96 \\ 200, & \text{with probability} \quad 0\cdot04. \end{cases}$$

For determining the value of VaR of the given investment we need to use the relation

$$VaR_{(1-\alpha)}(X) = \text{Min}\{z : P(X \leq z) \geq (1-\alpha)\}$$

with $(1-\alpha) = 0\cdot95$. Since $X$ is a discrete random variable we have

$$P(X \leq -100) = P(X = -100) = 0\cdot96 \geq 0\cdot95,$$

and $P(X \leq 200) = 1 \geq 0\cdot95$. But -100=Min(-100, 200) and therefore the required VaR is $Rs\ -100$.

□

**Example 6.5.3** *Let there be two identical bonds A and B. Each of these defaults with probability 0.04 giving a loss of* Rs 100. *Further there is no loss if default does not occur. Let default occur independently and C be a portfolio consisting of these two bonds A and B. Obtain VaR of bonds A and B; and also of the portfolio C at* 95% *confidence level.*

**Solution** Following on lines similar to Example 6.5.2, it is simple to get VaR(A)=0 and VaR(B)=0. Since C is the portfolio consisting of bonds A and B, its return R becomes the sum of returns of bonds A and B. Writing returns in terms of the loss function, we obtain the distribution of loss $X$ of investment $C$ as

$$X = \begin{cases} 0, & \text{with probability} \quad (0\cdot 96)^2 = 0\cdot 9216 \\ 200, & \text{with probability} \quad (0\cdot 04)^2 = \cdot 0016 \\ 100, & \text{with probability} \quad (1-(0\cdot 96)^2-(0\cdot 04)^2) = 0\cdot 0768. \end{cases}$$

Therefore at 95% confidence level,

$$VaR(C) = Rs\ 100\ ,$$

because P(loss of investment $C \leq 100) = 0.9216 + 0.0768 = .9984 > 0.95$, and P(loss of investment $C \leq 200) = 1 > 0\cdot 95$, but $100 = \text{Min}(100, 200)$. Here we may note that $P(X \leq 0) = P(X = 0) = 0\cdot 9216 < 0\cdot 95$.

□

**Some Theoretical and Computational Difficulties with VaR**

(i) In Example 6.5.3, at 95% confidence level VaR(C)=100 but VaR(A)+VaR(B)=0. Thus $VaR(C) > VaR(A) + VaR(B)$. But this violates the principle that *diversification reduces risk*. We expect a *good* risk measure to respect this principle. Unfortunately VaR does not do so. Mathematically it means that *VaR is not a subadditive risk measure*. We call a risk measure f to be *subadditive* if for two different investments A and B,

$$f(A + B) \leq f(A) + f(B)\ ,$$

i.e. the total risk of two different investment portfolios does not exceed the sum of individual risks.
In Example 6.5.3, diversification has actually increased the risk if VaR is used as a risk measure.

(ii) VaR does not pay any attention to the magnitude of losses beyond the VaR value. For example it is very unlikely that an investor will take a neutral view for two portfolios with identical expected return and VaR, but return distribution of one portfolio having short left tail and other having a long left tail.

(iii) There is additional difficulty with VaR in its computation and optimization. When VaR is computed by generating scenarios, it turns out to be a non smooth and non convex optimization problem is required to be solved.

The above shortcomings of VaR has motivated researchers to look for other quantile based risk measures and CVaR (conditional value at risk) is one such risk measure.

Before we proceed with the discussion of CVaR, we remark that inspite of the difficulties outlined above, VaR is still very popular in the market. Therefore we need to discuss some non-parametric and parametric methods for its estimation.

**Nonparametric Estimation of VaR**

Here we describe a procedure for the nonparametric estimation of VaR. Since it is a nonparametric estimation, the loss distribution is not assumed to be in any parametric family of distributions such as normal distributions.

To motivate the procedure, let us consider a simple example. Suppose that we hold a Rs 20,000 position in an index fund, e.g. NIFTY, so that our returns are those of this index. Suppose we require a 24-hour VaR by using 1000 daily returns on NIFTY for the period ending on December 31,2010. These are approximately the last four years of daily returns in the time series of NIFTY.

Let the confidence level $(1-\alpha)$ be 95%, i.e. $\alpha = 5\%$. Since $5\%$ of 1000 is 50 and VaR is a loss which is minus of the revenue, we can estimate the required VaR by taking the $50^{th}$ smallest daily return. Let this be $r_{(50)}$, the $50^{th}$ *order statistic* of the sample of historic returns. This means that a daily return of $r_{(50)}$ or less occurred only $5\%$ time in the historic data. Equivalently this means that a daily loss of $-r_{(50)}$ or more occurred only $5\%$ time in the historic data. A return of $r_{(50)}$ on a $Rs\ 20,000$ investment yields a revenue of $20,000r_{(50)}$ which can be taken as an estimate of VaR at 95% confidence level. Specifically if $r_{(50)} = -0\cdot 0227$, then $-r_{(50)} = 0\cdot 0227$ and the estimated VaR is $20,000(0\cdot 0227)$, i.e. $Rs\,454$.

The above procedure can be generalized easily. We first note that $(1-\alpha)^{th}$ percentile of loss equals the $\alpha^{th}$ percentile of the returns. Suppose that there are n returns $r_1, r_2, \ldots, r_n$ in the historic sample and let k be equal to $(n\alpha)$ rounded to the nearest integer. Then $\alpha$ percentile of the sample of returns is the $k^{th}$ smallest return, that is, the $k^{th}$ order statistic $r_{(k)}$ of the sample of returns. If S is the size of the initial investment then the required estimate of VaR is $-Sr_{(k)}$. Here minus sign converts revenue (return times initial investment) to a loss.

**Remark 6.5.1** *Estimation using sample percentiles is only feasible if the sample size is large. If T(holding period) is taken a quarter rather than a day, then in a four years duration we will have only 16 observations. In such a situation, increasing the number of years for historical data would also not increase sample size substantially, and also because of volatility there may be bias in our estimate.*

### Parametric Estimation of VaR

For small sample size, VaR can be best estimated by parametric technique. We discuss only the case of normal distribution because it is extremely simple.

Since, in general, the historical data or returns is given, we have

$$VaR_{(1-\alpha)}(loss) = -VaR_{(\alpha)}(return).$$

Therefore

$$VaR_{(1-\alpha)}(X) = -S(\overline{\mu} + \Phi^{-1}(\alpha)\sigma),$$

which can be estimated by $-S(\overline{X} + \Phi^{-1}(\alpha)\hat{\sigma})$. Here $\overline{X}$ and $\hat{\sigma}$ are the mean and standard deviation of sample of returns and S is the size of the initial investment.

Let us assume that from the given historical data, $\overline{X} = -3 \cdot 107 \times 10^{-4}$ and $\hat{\sigma} = 0 \cdot 0151$. Also $\Phi^{-1}(\alpha) = -1 \cdot 645$ for $\alpha = 0 \cdot 05$. Then the estimate of the required VaR is $((-3 \cdot 107 \times 10^{-4}) + (-1 \cdot 645)(0 \cdot 0151)(20,000)) = Rs\ 471$.

In this example, the estimate of the expected return, is negative. This may be because the particular years used include a prolonged bear market. However, we certainly do not expect average future returns to be negative, otherwise we would not have invested Rs 20,000 in the market.

Since in actual practice, normality assumption is not going to hold, we need to apply historical simulation and Monte Carlo strategies to generate a large number of possible scenarios. Also we need to understand VaR for a portfolio of assets which could include options along with certain number of stocks. The estimation of VaR in this situation also requires Monte Carlo simulation. We shall discuss Monte carlo simulation in Chapter 14.

## 6.6 Conditional Value-at-Risk

One major criticism of VaR is that it pays no attention to the magnitude of losses beyond the VaR value. Also VaR is not subadditive, which not only violates the principle of diversification, but also creates computational difficulty in the resulting portfolio optimization problem. These undesirable features of VaR led to the development of *conditional value-at-risk*, denoted by CVaR, which is obtained by computing the expected loss given that the loss exceeds VaR. Some other names for CVaR are *expected shortfall*, *expected tail loss* and *tail VaR*. CVaR has been introduced by Rockafellar and Uryasev [111] who studied many mathematical properties of CVaR in detail.

To define CVaR we first define the occurrence of a tail event. We say that a *tail event occurs* if the loss exceeds the VaR. Then CVaR is the conditional expectation of loss given that the tail event occurs. We now proceed to define CVaR and discuss its minimization mathematically. To be specific, we define CVaR in context of portfolio optimization.

We consider a portfolio of assets with random returns. Let $f(w,r)$ denote the loss function when we choose the investment $w$ from a set of feasible portfolios and $r$ is the realization of random returns. In the context of our portfolio optimization problem, $f(w,r) = -r^T w$ where $r = (r_1, r_2, \ldots, r_n)^T$ is the vector of random returns and $w = (w_1, w_2, \ldots, w_n)^T$ with $e^T w = 1$, e being the vector $(1, 1, \ldots, 1)^T$. Here in the expression of $f(w,r)$, minus sign has been taken to express return $r^T w$ as loss function. We assume that the return vector $r$ has a probability density function $p(r)$, e.g., the random returns may have a multivariate normal distribution.

For a fixed weight vector $w$, we define

$$\psi(w,q) = \int_{f(w,r)\leq q} p(r)dr, \tag{6.43}$$

which represents the cumulative distribution function of the loss associated with the weight vector $w$. In fact $\psi(w,q) = P(loss \leq q)$. Therefore for a given confidence level $(1-\alpha)$, the $VaR_{(1-\alpha)}(w)$ associated with the portfolio is given by

$$VaR_{(1-\alpha)}(w) = \text{Min}\{q \in \mathbf{R} : \psi(w,q) \geq (1-\alpha)\}.$$



**Definition 6.6.1 (Conditional Value-at-Risk)** *Let $w$ be a given portfolio and $(1-\alpha)$ be the given confidence level. Then $CVaR_{(1-\alpha)}(w)$ associated with the portfolio $w$ is defined as*

$$CVaR_{(1-\alpha)}(w) = \frac{1}{\alpha}\int_{f(w,r)\geq VaR_{(1-\alpha)}(w)} f(w,r)p(r)dr. \tag{6.44}$$

**Remark 6.6.1** *In special case of returns having discrete probability distribution, we have*

$$CVaR_{(1-\alpha)}(w) = \frac{1}{\alpha}\left(\sum_{j\in J_1} p_j f(w, r^{(j)})\right), \tag{6.45}$$

*where $J = \{1, 2, \ldots, m\}$, $J_1 = \{j \in J : \ f(w, r^{(j)}) \geq VaR(w)\}$ and the vector $r^{(j)}$ is the $j^{th}$ realization of the return vector $r$ with probability $p_j$.*

**Remark 6.6.2** *In certain sense, (6.44) and (6.45) tell that $CVaR_{(1-\alpha)}(w)$ is the average of the outcomes greater than $VaR_{(1-\alpha)}(w)$. This is certainly true for continuous distribution functions; but for general distributions this is not exactly true. There are certain subtle aspects which need to be explained. We shall refer to Sriboonchitta et el [128] in this regard.*

**Example 6.6.1** *Let the loss function $f(w,r)$ be given by $f(w,r) = -r$ where $r = 75 - j$, $(j = 0,1,2,\ldots,99)$ with probability 1%. Evaluate VaR(w) and CVaR(w) at 95% confidence level.*

**Solution** The loss function $f(w,r) = -r$, where $r = 75 - j$, $(j = 0,1,2,\ldots,99)$, takes 100 values given by $(-75,-74,-73,\ldots,0,1,\ldots,19,20,21,22,23$ and $24)$ with equal probability $p = 0 \cdot 01$. Therefore $VaR(w) = 20$ for $1 - \alpha = 0 \cdot 95$.
We next evaluate $CVaR_{(1-\alpha)}(w)$ for $(1 - \alpha) = 0.95$ by using formula (6.45). This gives

$$CVaR(w) = \frac{1}{0 \cdot 05}(20 + 21 + 22 + 23 + 24)(0 \cdot 01) = 22.$$

Thus at 95% confidence level VaR is 20 and CVaR is 22.

□

In the above example for the same confidence level CVaR is more than VaR. We have the following result in this regard.

**Lemma 6.6.1.** *For the given confidence level $(1-\alpha)$, $CVaR_{(1-\alpha)}(w) \geq VaR_{(1-\alpha)}(w)$.*

**Proof.** We have

$$\begin{aligned} CVaR_{(1-\alpha)}(w) &= \frac{1}{\alpha}\int_{f(w,r)\geq VaR_{(1-\alpha)}(w)} f(w,r)p(r)dr \\ &\geq \frac{1}{\alpha}\int_{f(w,r)\geq VaR_{(1-\alpha)}(w)} VaR_{(1-\alpha)}(w)p(r)dr \\ &= \frac{VaR_{(1-\alpha)}(w)}{\alpha}\int_{f(w,r)\geq VaR_{(1-\alpha)}(w)} p(r)dr \\ &= VaR_{(1-\alpha)}(w). \end{aligned}$$

□

**Remark 6.6.3** *As CVaR of a portfolio is always more than or equal to VaR for same $(1 - \alpha)$, portfolios with small CVaR will also have small VaR. But this does not mean that the minimization of CVaR is equivalent to the minimization of VaR.*

**Minimization of CVaR**

We now discuss Rockafellar and Uryasev's [111] procedure to minimize CVaR. Since the definition of CVaR involves the VaR function explicitly, it is not very convenient to optimize CVaR directly. Therefore we introduce the auxiliary function

$$F_{(1-\alpha)}(w,q) = q + \frac{1}{\alpha}\int_{f(w,r)\geq q}(f(w,r)-q)p(r)dr. \tag{6.46}$$

If we denote $a^+ = Max(a,0),\ a \in \mathbf{R}$, then from (6.46) we get

$$F_{(1-\alpha)}(w,q) = q + \frac{1}{\alpha}\int (f(w,r)-q)^+ p(r)dr. \tag{6.47}$$

We refer to Rockafellar and Uryasev [111] for below given results.

**Lemma 6.6.2.** *The auxiliary function* $F_{(1-\alpha)}(w,q)$ *is a convex function of* $q$.

**Lemma 6.6.3.** $VaR_{(1-\alpha)}(w)$ *is a minimizer of* $F_{(1-\alpha)}(w,q)$ *over* $q$.

**Lemma 6.6.4.** *The minimum value of* $F_{(1-\alpha)}(w,q)$ *over* $q$ *is* $CVaR_{(1-\alpha)}(w)$.

In view of the above Lemmas, we have for a given $w$,

$$CVaR_{(1-\alpha)}(w) = \text{Min}_q(F_{(1-\alpha)}(w,q)) = F_{(1-\alpha)}(w, VaR_{(1-\alpha)}(w)). \tag{6.48}$$

The left equality of (6.48) tells that we can minimize CVaR directly without computing VaR first. Since for portfolios, the loss function $f(w,r) = -r^Tw$ is a linear and hence also convex function of $w$, the auxiliary function $F_{(1-\alpha)}(w,q)$ is a convex function of $w$. Therefore the problem

$$\begin{aligned} &\text{Min} \quad CVaR_{(1-\alpha)}(w) \\ &\text{subject to} \\ &\qquad e^Tw = 1, \end{aligned} \tag{6.49}$$

is equivalent to

$$\begin{aligned} &\underset{w,q}{\text{Min}} \quad F_{(1-\alpha)}(w,q) \\ &\text{subject to} \\ &\qquad e^Tw = 1, \end{aligned} \tag{6.50}$$

which is a smooth convex optimization problem. Though problem (6.50) can be solved by using standard convex optimization techniques, the formulation (6.50) still needs the computation/determination of joint density function p(r) of random return vector r. In practice it is not simple or even possible, so we present another approach which is based on generation of scenarios via computer simulation. An added advantage of this approach is that it results in a linear programming formulation.

Suppose we have scenarios $r^{(s)}$, (s=1,2,...,$n_s$), which may represent historical values of the random vector of returns or obtained via computer simulation. We shall assume that all scenarios have equal probability and define the following empirical distribution of the random returns based on the available scenarios

$$\widetilde{F}_{(1-\alpha)}(w,q) = a + \frac{1}{\alpha\, n_s}\sum_{s=1}^{n_s}(f(w,r^{(s)}) - q)^+. \tag{6.51}$$

Using (6.51) to approximate $F_{(1-\alpha)}(w,q)$, we get an approximation to the problem (6.50) as

$$\begin{aligned} &\underset{w,q}{\text{Min}} \quad \widetilde{F}_{(1-\alpha)}(w,q) \\ &\text{subject to} \\ &\qquad e^T w = 1. \end{aligned} \tag{6.52}$$

Now writing $(f(w,r^{(s)}) - q)^+$ as $z_s$ and using the definition of $a^+$, we obtain



$$\begin{aligned} &\underset{w,q,z_s}{\text{Min}} \quad q + \frac{1}{\alpha.n_s}\sum_{s=1}^{n_s} z_s \\ &\text{subject to} \\ &\qquad z_s \geq f(w,r^{(s)}) - q, \qquad s = 1,2,\ldots,n_s \\ &\qquad e^T w = 1 \\ &\qquad z_s \geq 0, \qquad s = 1,2,\ldots,n_s. \end{aligned} \tag{6.53}$$

In the context of portfolio optimization, $f(w,r^{(s)}) = -w^T r^{(s)}$ is linear and so the problem (13.5) becomes a linear programming problem.

Most often we try to optimize a suitable performance measure (e.g, expected return) while making sure that certain risk measures do not exceed a threshold value. It could be variance or absolute deviation as has been discussed earlier. When the risk measure is CVaR, the resulting optimization problem is

$$\begin{aligned} &\text{Max} \quad \mu^T w \\ &\text{subject to} \\ &\qquad CVaR_{(1-\alpha)}(w) \le u_\alpha \\ &\qquad e^T w = 1, \end{aligned} \tag{6.54}$$

which can be approximated as

$$\begin{aligned} &\underset{w,z,q}{\text{Max}} \quad \mu^T w \\ &\text{subject to} \\ &\qquad u_\alpha \ \ge q + \frac{1}{\alpha\, n_s}\sum_{s=1}^{n_s} z_s \\ &\qquad z_s \ \ge f(w, r^{(s)}) - q \qquad (s = 1, 2, \ldots, n_s) \\ &\qquad e^T w = 1 \\ &\qquad z_s \ \ge 0 \qquad (s = 1, 2, \ldots, n_s). \end{aligned}$$

In (6.54), we can have more than one CVaR constraint for different levels $\alpha$. Also similar to Markowitz's model, we can have a trade off between return and CVaR. We can refer to Mansini et al. [89] for more details in this regard.



## 6.7 Preference Relation, Utility theory and Decision Making

The mean-variance criterion used in the Markowitz portfolio model can also be explained in terms of the expected utility maximization principle in decision making. To understand this aspect we first wish to motivate the readers to the celebrated *Von Neumann- Morgenstern's* model for decision making. This model tries to explain how a *rational person* makes his/her *optimal choice* in a given situation. This requires some discussion on *preference relation* and related results in utility theory.

**Choice under certainty**

We first consider the situation when there is no uncertainty present, i.e. the alternatives are certain. Let this set of alternatives be denoted by $Z$. As the choice is being made under certainty, we shall get the element of our choice for sure.

In order to perform our choice, i.e. the selection procedure, we prescribe our own preferences, and use them to select the desired *optimal* element of $Z$. These preferences are expressed as a *preference relation,* say $\succ$ on $Z$. Thus we say that for $x, y \in Z$, $x \succ y$ states that we *strictly prefer* $x$ to $y$. Here our preference relation should reflect our *valuation* of different alternatives, i.e. different elements of $Z$. These valuations give a quantitative representation of our preference relation.

The above discussion can now be formalized mathematically. A preference relation $\succ$ is a binary relation on $Z$ i.e. a subset $A$ of $Z \times Z$ where $(x, y) \in A$ when $x \succ y$. We demand the following properties with our preference relation $\succ$

(i) The preference relation $\succ$ is *asymmetric* i.e. $x \succ y \Rightarrow y \nsucc x$. Thus if $x$ is strictly preferred to $y$ then $y$ is not strictly preferred to $x$.
(ii) The preference relation $\succ$ is *negative transitive* i.e. $x \nsucc y$ and $y \nsucc z \Rightarrow x \nsucc z$. Thus if $x$ is not strictly preferred to $y$ and $y$ is not strictly preferred to $z$ then $x$ is not strictly preferred to $z$.

We write $x \simeq y$ if $x \nsucc y$ and $y \nsucc x$. The relation $\simeq$ is called an *indifference* or *equivalence* relation because it is reflexive, symmetric and transitive. Thus given $x, y \in Z$, exactly one of $x \succ y$, $y \succ x$, or $x \simeq y$ holds. We further write $x \succeq y$ when $y \nsucc x$, i.e. either $x \succ y$ or $x \simeq y$. The relation $\succeq$ is called the *weak preference relation*. It is simple to verify that $x \succ y$ when $y \nsucceq x$. Also $x \simeq y$ when $x \succeq y$ and $y \succeq x$. For making an *optimal choice* we need to prescribe a numerical representation of our strict preference relation $\succ$ . This is done via the utility function.

**Definition 6.7.1 (Utility Function)** *A function $u : Z \to \mathbf{R}$ is called a utility function if*

$$x \succ y \Leftrightarrow u(x) > u(y).$$

An obvious question at this stage is about the existence of such a function $u$. We have the following theorem in this regard (Sriboonchitta et al. [128]).

**Theorem 6.7.1** *Let $Z$ be finite or countably infinite. Then a utility function $u$ as defined above certainly exists.*

**Choice Under Uncertainty**

We next discuss the more realistic situation of making an *optimal* choice under uncertainty. A typical example of this scenario could be possible losses in investment portfolio which are real-valued random variables. Here the range space $Z$ is the real line $\mathbf{R}$. Taking motivation from the earlier case of choice under certainty, we attempt to give a numerical representation of a preference relation $\succ$ on a class of random variables $\mathfrak{X}$. We say that for $X, Y \in \mathfrak{X}$ and a given preference relation $\succ$, there exists a utility function $u : Z \to \mathbf{R}$ (unique up to a positive linear transformation) such that

$$X \succ Y \Leftrightarrow E(u(X)) > E(u(Y)),$$

provided that the expectations exist. Thus for making their *optimal* choices under uncertainty, decision makers try to maximize their expected utilities. This principle is called *Von Neumann- Mangenstern expected utility maximization principle.* Is this principle always applicable? Does there always exist a utility function as described above for the random variables in the class $\mathfrak{X}$? Apart from various mathematical conditions for its existence it was argued that the existence of utility function is restricted to "rational" people. This latter part has been a topic of debate among economists and psychologists. For this as well as other aspects of Von Neumann - Mangenstern theory we shall refer to Fishburn [47].

**Utilities and Risk Attitudes**

We all know that different people react differently in risky situations. They make decision based upon their own *attitudes* towards risk. Thus, making an investment choice in risky and risk free assets, each person will act differently depending upon his/her risk attitude. Roughly there are three main risk attitudes of an investor. Let the random variable $X$ be a risky prospect, and $u$ be the utility function of the decision maker. We say that the investor is

(i) *risk neutral* if, when facing two risky prospects with same expected value will feel indifferent. In terms of the utility function $u$, this means that $u(E(X)) = E(u(Y))$. This will happen for example if $u$ is linear, say $u(x) = x$.
(ii) *risk averse* if, when facing two risky prospects with the same expected value, will prefer the less risky one. In terms of the utility function $u$, this means that $u(E(X)) > E(u(X))$. This will happen for example if $u$ is concave, say $u(x) = \log x$ or $u(x) = -x^2$.
(iii) *risk seeking* if, when facing two risky prospects with the same expected value, will prefer the more risky one. In terms of the utility function $u$, this means that $u(E(X)) < E(u(X))$. This will happen for example if $u$ is convex, say $u(x) = x^2$.

It is obvious that given a utility function $u(x)$, any function of the form $v(x) = au(x) + b$ with $a > 0$ is a utility function equivalent to $u(x)$. This is because equivalent utility functions give identical rankings in terms of the principle of expected utility maximization.

To explain the meaning of risk averse scenario, let us consider an investor who has two alternatives for future wealth.The first alternative is based on the outcome of tossing of a coin. If the coin turns up 'head' the investor gets $Rs\ x$, otherwise he/she gets $Rs\ y$. The second alternative is a sure event of getting $Rs\ \left(\frac{x+y}{2}\right)$ with certainty. Let the investor's utility function be given by a strictly concave function

$u(x)$. Then the expected utility of first alternative is $\dfrac{u(x)+u(y)}{2}$, whereas the expected utility of the second alternative is $u\left(\dfrac{x+y}{2}\right)$. Since $u$ is strictly concave we have

$$u\left(\frac{x+y}{2}\right) > \frac{u(x)+u(y)}{2}.$$

Hence by the principle of expected utility maximization we shall prefer the sure wealth of $Rs\ \dfrac{(x+y)}{2}$ to a 50-50 chance of $x$ or $y$. Here we must note that both alternatives have the same expected value of $Rs\ \dfrac{(x+y)}{2}$, but the one without risk (certainty of getting) is preferred. This explains the risk averse attitude of the investor.

The below given result is useful in this regard.

**Theorem 6.7.2** *An investor is risk averse if and only if his/her utility function is concave.*

**Remark 6.7.1** *Although it has not been mentioned in the statement of above theorem, $u$ is non-decreasing by the definition of utility function. We refer to Sriboonchitta et al. [128] for the proof of the above theorem. The proof essentially uses Jensen's inequality and the defining condition $u(E(X)) > E(u(X))$ for a risk averse investor.*



Let us take another example to understand the risk attitudes. For an investor let there be three possible scenarios: a $5\%$ chance of loosing $Rs\ 20,000$, a $10\%$ chance of loosing $Rs\ 10,000$, and a sure chance of loosing $Rs\ 1,000$. A risk neutral investor will not find any of these situations worse than the other. For a risk averse investor, the first situation is worse than the second one, and the second one is worse than the last one. This is because risk averse investors dislike uncertainty about the size of losses, so they do not prefer even a small possibility of large amount of loss.

**Certainty Equivalent**

The *certainty equivalent* of a random wealth $X$ is defined to be the amount of a certain (i.e. risk-free) wealth that has a utility level equals to the expected utility of $X$. Thus if $C$ is the certainty equivalent of a random wealth $X$ then by definition

$$u(C) = E(u(X)).$$

The certainty equivalent of a random variable is same for all equivalent utilities and is measured in units of wealth.

## 6.8 Mean-Variance Model Revisited in Utility Frame Work

We now discuss Markowitz's mean-variance model in the frame work of expected utility maximization principle. Here we need to solve the following optimization problem

$$\begin{aligned} \text{Max} \qquad & E\left(u(r^T w)\right) \\ \text{subject to} & \\ & e^T w = 1, \end{aligned} \tag{6.55}$$

where $r^T = (r_1, r_2, \ldots r_i, \ldots, r_n)$ is the vector of asset returns, $u$ is investor's utility function and $w^T = (w_1, w_2, \ldots w_i, \ldots, w_n)$ is the decision vector of asset weights.

Earlier we had agreed to represent a portfolio by $w : (w_1, w_2, \ldots w_i, \ldots, w_n)$ such that $e^T w = 1$, where $e^T = (1, 1, \ldots, 1)$. Therefore $X = r^T w$ denotes the return (random) of the portfolio $w$. The mean-variance rule of Markowitz states that the portfolio $w^{(1)}$ is preferred to the portfolio $w^{(2)}$ if

(i) $E(r^T w^{(1)}) \geq E(r^T w^{(2)})$,

and

(ii) $\sigma^2(r^T w^{(1)}) \leq \sigma^2(r^T w^{(2)})$.

If we denote the random variable $r^T w^{(1)}$ by X and the random variable $r^T w^{(2)}$ by $Y$ then the mean-variance rule says that $X$ is preferred to $Y$ when $E(X) \geq E(Y)$ and $Var(X) \leq Var(Y)$. The essence of Markowitz's theory is that the above mean-variance rule is constrained with the principle of expected utility maximization for the utility function $u(x) = ax - \frac{1}{2}b\,x^2$ with $a > 0$ and $b \geq 0$. This utility function is really meaningful only in the range $x \leq a/b$, where the function is increasing. Also as $b > 0$, the function $u$ is strictly concave every where and therefore corresponds to the risk aversion scenario.

Let the given portfolio have a random wealth value of $y$. Then

$$\begin{aligned} E\left(u(y)\right) &= E\left(ay - \frac{1}{2}b\ y^2\right) \\ &= aE(y) - \frac{1}{2}b(E(y))^2 - \frac{1}{2}bVar(y). \end{aligned} \tag{6.56}$$

Now as per the expected utility maximization principle the optimal portfolio $y$ is the one that maximizes (6.56) with respect to all feasible choices of the random wealth variable $y$. Thus $E(u(\hat{y})) \geq E(u(y))$ for all feasible $y$. Let $E(\hat{y}) = M$ and $S = \{y : y \text{ is feasible and } E(y) = M\}$. We shall now show that $Var(\hat{y}) \leq Var(y)$ for all $y \in S$. If possible let there exists $\overline{y} \in S$ such that $Var(\overline{y}) \leq Var(\hat{y})$. Then as $\overline{y} \in S$ it is feasible and $E(\overline{y}) = M$. Further

$$aM - \frac{1}{2}bM^2 - \frac{1}{2}bVar(\hat{y}) \leq aM - \frac{1}{2}bM^2 - \frac{1}{2}bVar(\overline{y}). \tag{6.57}$$

But then (6.57) contradicts that $\hat{y}$ is optimal. Therefore $Var(\hat{y}) \leq Var(y)$ for all $y \in S$, i.e. $\hat{y}$ has the minimum variance with respect to all feasible $y's$ with $E(y) = M$. Hence $\hat{y}$ corresponds to a mean-variance efficient point. Different mean-variance efficient points are obtained by providing different values for the parameters $a$ and $b$. The readers may identify $y = r^T w,\ E(y) = \mu^T w$ and $Var(y) = w^T C w$ with $C$ as the variance-covariance matrix.

Now we have a very natural question. What happens if we take a general utility function $u$? How different are the mean-variance efficient portfolios and portfolios that maximize the expected utility i.e. those which are the solution of the problem (6.55) for a general $u$? In this context we have the following theorem

**Theorem 6.8.1** *Let $X$ and $Y$ be two random variables, normally distributed with means $\mu_1, \mu_2$ and variances $\sigma_1^2, \sigma_2^2$ respectively. Then following are equivalent*

*(i) $E(u(X)) \geq E(u(Y))$ for any $u : R \to R$ which is non decreasing and concave.*
*(ii) $\mu_1 \geq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$.*

**Proof.** See details in Sriboorchitta et al. [128].

□

**Remark 6.8.1** *In view of the above theorem, if the returns are normally distributed and the utility function is concave, then the portfolio that maximizes the expected utility is on Markowitz's efficient frontier. Exactly which portfolio on the efficient frontier will be obtained, will depend on the choice of utility function.*

## 6.9 Risk Modeling and Financial Risk Measures

Financial risk management is a broad concept involving various perspectives. It is about managing exposure to risk, such as market risk and credit risk. For these purposes we need a *quantitative concept of risk* in its own right.

Mathematically, risk management is a procedure for shaping a loss function or loss distribution. We have already studied several moment based risk measures (e.g. $L_1, L_2$ and $L_\infty$-risk measures) and also some quantile based risk measures (e.g.

VaR and CVaR). Here we attempt to understand risk measures in a broader mathematical framework and introduce the concept of *coherent risk measures*(Artzner [5]). An excellent tutorial on this topic is due to Rockafellor [112].

Basically, we wish to answer the following question: *How can we assess the risk of a financial position mathematically*? For this we note that a financial position is captured by a random variable X, e.g. loss. Information about loss is in its cumulative distribution function F. In the financial context where the *risk* is about the risk of loosing money, we wish to define numerical risk for $X$ as some appropriate number $\rho(X)$ and call it as a risk measure. It seems natural to choose $\rho(\cdot)$ in such a manner that it captures our risk perception and it is also suitable for financial management purposes.

**Domain and Range of Risk Measures**

As stated above, we wish to assign a numerical value $\rho(X)$ to each random variable X to describe its risk. Here X stands for the loss of an investment portfolio, the loss of a financial position, or capital needed to hold for an insurance company to avoid insolvency.

It is very clear that the range of $\rho$ has to be $\mathbf{R}$. But what about its domain? Let $\mathcal{U}$ denote the real vector space of all possible real valued random variables. Mathematically we do not need the entire vector space $\mathcal{U}$ as the domain of $\rho$. But rather we need a particular subset of $\mathcal{U}$ namely a convex cone $\mathfrak{X}$.

**Definition 6.9.1 (Cone)** *A subset $\mathfrak{X}$ in $\mathcal{U}$ is called a cone if for $X \in \mathfrak{X}$ and $\lambda > 0$, we have $\lambda X \in \mathfrak{X}$.*

**Definition 6.9.2 (Convex Cone)** *A subset $\mathfrak{X}$ in $\mathcal{U}$ is called a convex cone if it is a cone and also a convex set. Thus $\mathfrak{X}$ is a convex cone if for $X, Y \in \mathfrak{X}$, $\lambda > 0$, we have $X + Y \in \mathfrak{X}$ and $\lambda X \in \mathfrak{X}$.*

It is economically meaningful to take the domain of risk measures as a convex cone of the vector space £. This is because we know that diversification should reduce risk. But for this we need to evaluate our risk measure on an investment portfolio of type $X = \sum_{i=1}^{k} \lambda_i Y_i$, where $\lambda_i > 0$, $\sum_{i=1}^{k} \lambda_i = 1$ and $Y_i$ denotes the rate of return of $i^{th}$ asset in the portfolio. Thus if $Y_i$ are in a domain $\mathfrak{X}$ of £, we also require that $\sum_{i=1}^{n} \lambda_i Y_i$, ($\lambda_i > 0$, $\sum_{i=1}^{k} \lambda_i = 1$), should also be in $\mathfrak{X}$; i.e. $\mathfrak{X}$ is a convex cone.

**Desirable properties of Risk Measures**

Let $\mathfrak{X}$ denote the convex cone of *loss* random variable X. Then, as per Artzner [5], we consider the following desirable properties for a risk measure $\rho : \mathfrak{X} \mapsto \mathbf{R}$.

(A1) (Monotonicity) If $X \leq Y$ almost surely, i.e. $P(X \leq Y) = 1$, then $\rho(X) \leq \rho(Y)$.
(A2) (Positive Homogeneity) $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda \geq 0$.
(A3) (Translation Invariance) $\rho(X + a) = \rho(X) + a$, $a \in \mathbf{R}$.
(A4) (Subadditivity) $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

**Definition 6.9.3 (Coherent Risk Measure)** *A functional* $\rho : \mathfrak{X} \mapsto \mathbf{R}$ *is called a coherent risk measure if it satisfies axioms (A1)-(A4) listed above.*

**Remark 6.9.1** *Though we are calling above properties as axioms, our approach is not axiomatic as these do not characterize a risk measure. These are simply the desirable properties and therefore even if a functional satisfies these axioms, it may not have any meaning in terms of risk. This can be verified for* $\rho(X) = E(X)$.

The above axioms are based on our perception about risk. In case of *loss* random variables, it is natural to believe that smaller losses have smaller risks. This is precisely the axiom (A1). Also the risk of a loss of a financial position should be proportional to the size of the loss of the position. This leads to the axioms (A2) and (A3). The principle that *diversification should reduce risk* leads to the axiom (A4). Also axioms (A2) and (A4) imply that the risk measure is convex, which is certainly a desirable property.

Before we analyze some of already studied risk measures, we need to introduce the notion of *stochastic dominance*. Stochastic dominance rules refer to criteria to order distributions of random variables modelling uncertain quantities such as future returns or investments. Thus these are various partial orders on the space of distribution functions of random variables and they are based on the principle of maximization of expected utility (Sriboonchitta et al. [128]). While stochastic dominance rules shed light on which future prospects are better than others, they do not provide us with any numerical value which we need to manage further our financial positions. This numerical value is provided by risk measures. Therefore for further analysis of risk measures, apart from axioms (A1)-(A4), we also need to check consistency of risk assessment with respect to stochastic dominance.

**Definition 6.9.4 (First Order Stochastic Dominance)** *Let X, Y be two random variables with distribution functions F, G respectively. Then X is said to dominate Y in the first order stochastic dominance (FSD), denoted by* $X \succeq_1 Y$, *if* $F(.) \leq G(.)$.

**Remark 6.9.2** *If* $X \succeq_1 Y$ *then* $E(X) \geq E(Y)$ *but the converse is not true. This is because*

$$E(X) = \int_0^\infty (1 - F(t))\, dt - \int_{-\infty}^0 F(t) dt.$$

As preference relations on the space of distribution functions, stochastic dominance rules are essentially *stochastic orders* which have been introduced much earlier in statistics. These were rediscovered and made use in Economics/Finance. We may refer to the survey article by Levy [83] in this regard. An obvious question at this stage is: How can we rank distributions when they intersect? In this situation FSD cannot be applied and therefore we need to weaken the criteria which is consistent with FSD.

**Definition 6.9.5 (Second Order Stochastic Dominance)** *Let X, Y be two random variables with distribution functions F, G respectively. Then X is said to dominate Y in the second order stochastic dominance (SSD), denoted by* $X \succeq_2 Y$, *if*

$$\int_{-\infty}^{t} (G(u) - F(u))du \geq 0,$$

*for all* $t \in \mathbf{R}$.

To analyze any proposed risk measure we have to take care of two main criteria. These are being coherent and being consistent with the stochastic dominance rules. In general we say that X *precedes* Y when $X \preceq Y$, and a risk measure $\rho$ is said to be *consistent* with respect to the partial order $\preceq$ if

$$(X \preceq Y) \Rightarrow (\rho(X) \leq \rho(Y)).$$

Here $\preceq$ could be $\preceq_1$ or if need be $\preceq_2$ or some other higher order stochastic dominance rule. Since risks are related to preferences, such a consistency with respect to a suitable partial order $\preceq$ is important.

**Remark 6.9.3** *First order stochastic dominance (FSD) implies the Second order stochastic dominance (SSD). Thus* $(X \succeq_1 Y) \Rightarrow (X \succeq_2 Y)$. *Also if* $(X \succeq_2 Y)$, *then* $E(X) \geq E(Y)$. *Further if* $(X \succeq_2 Y)$ *and* $E(X) = E(Y)$, *then* $Var(X) \leq Var(Y)$.

Though higher order stochastic dominance rules have also been introduced in the literature, in context of portfolio optimization only FSD and SSD have been popular.

### Some Already Studied Popular Risk Measures

(i) **Variance** In the early theory of portfolio selection, Markowitz proposed the use of variance as a risk measure. Thus the risk measure $\rho$ is given by

$$\rho(X) = Var(X) = E(X - E(X))^2 ,$$

and portfolio with small variance is considered less risky.
Here we note that variance as a risk measure is not a coherent risk measure as it is not subadditive. Further it is not consistent with the first order stochastic dominance because $(X \geq_1 Y)$ does not imply $Var(X) \leq Var(Y)$. Due to these reasons, and also to take care of non symmetry in the loss distributions, risk measures based on downside tail distribution of returns have been introduced.

(ii) **Value-at-Risk** For a loss random variable X with the cumulative distribution function F, the value-at-risk of X with confidence level $(1-\alpha),\ 0<\alpha<1$ is defined as

$$VaR_{(1-\alpha)}(X) = F^{-1}(1-\alpha) = \text{Min}\{x \in \mathbf{R} : F(x) \geq (1-\alpha)\}.$$

Here $(1-\alpha)$-quantile is the position such that $P(X \leq F^{-1}(1-\alpha)) \geq (1-\alpha))$. Thus, $P(X > F^{-1}(1-\alpha)) \leq \alpha$.
It can now be shown that VaR is again not a coherent risk measure because it is not subadditive. We have already given such an example but below given another example which is probably more general.

**Example 6.9.1** *Let $X, Y$ be two independent and identically distributed Bernoulli random variables with parameter $p$, i.e. $P(X=0)=(1-p)$ and $P(X=1)=p$. Let $(1-p)^2 < \alpha < 1$. Show that VaR is not sub additive in this situation.*

**Solution** As $(1-p)^2 < \alpha < 1$, we have $0 < (1-\alpha) < 2p(1-p)$. Therefore $F_X^{-1}(1-\alpha) = 0 = F_Y^{-1}(1-\alpha)$. But $F_{X+Y}^{-1}(1-\alpha) > 0$. Therefore for this choice of $\alpha$, the subadditive law does not hold.

□

Let us recall the definition of SSD. We say that $X \geq_2 Y$ if

$$\int_{-\infty}^{t} (F_Y - F_X)(u)du \geq 0, \quad \forall\ t \in \mathbf{R}.$$

With respect to SSD, $\geq_2$ we define $\geq_{2'}$ as follows: $X \geq_{2'} Y$ if

$$\int_{t}^{\infty} (F_Y - F_X)(u)du \geq 0, \quad \forall\ t \in \mathbf{R}.$$

The order $\geq_{2'}$ is called *risk seeking stochastic dominance* (RSSD), also called *stop-loss-order* which is very popular in acturial science.

**Definition 6.9.6 (Distortion Function)** *A function $g : [0,1] \to [0,1]$ is called a distortion function if (i) it is nondecreasing and (ii) g(0)=0, g(1)=1.*

**Definition 6.9.7 (Risk Measure $\rho_g$)** *Let g be the given distortion function. A risk measure $\rho_g$ (risk measure with respect to the given distortion function g) is defined as*

$$\rho_g(X) = \int_0^{\infty} g\,(1 - F(t))\,dt + \int_{-\infty}^{0} [g\,(1 - F(t)) - 1]dt,$$

*where F is the cumulative distribution function of the random variable X (say loss random variable).*

**Remark 6.9.4** *VaR is a risk measure of type $\rho_g$. Some readers may like to obtain the distortion function g for VaR.*

**Lemma 6.9.1.** *If the distortion function g is concave then*

$$X \succeq_{2'} Y \Rightarrow \rho_g(X) \leq \rho_g(Y).$$

*Thus any risk measure $\rho_g$ defined with respect to a concave distortion function g is $\preceq_{2'}$ always consistent.*

Using the above lemma it can be shown that VaR is consistent with respect to the stochastic dominance rule induced by the partial order $\preceq_{2'}$. Thus $(X \preceq_{2'} Y) \Rightarrow VaR_{(1-\alpha)}(X) \leq VaR_{(1-\alpha)}(Y)$ for all $\alpha \in (0, 1)$.

(iii) **Conditional Value-at-Risk** Let us recall that $CVaR_{(1-\alpha)}$ is the conditional expectation of loss given that the loss exceeds $VaR_{(1-\alpha)}$. It can be shown that unlike other risk measures CVaR is a coherent risk measure. Further using Lemma 6.9.1, it is coherent with respect to the stochastic dominance rules induced by the partial order $\preceq_{2'}$. Thus $(X \preceq_{2'} Y)$ implies that $CVaR_{(1-\alpha)(X)} \leq CVaR_{(1-\alpha)(Y)}$ for all $\alpha \in (0, 1)$.

**Remark 6.9.5** *The proofs of most of results in this section have been omitted. This has been done deliberately because the proofs are involved and need a good knowledge of measure theory, in particular the theory of Choquet integrals. Interested readers are encouraged to refer to Sriboonchitta et al. [128] and references cited therein.*

## 6.10 Stochastic Dominance and Portfolio Optimization

Let $r^T = (r_1, r_2, \ldots, r_n)$ be the vector of random returns for n assets and $w : (w_1, w_2, \ldots, w_n)$ be the given portfolio. Then the portfolio return $R(w)$ is given by the random variable $r^T w$ where $w^T = (w_1, w_2, \dots, w_n)$. In the context of portfolio optimization we consider the stochastic dominance relations between the random returns $r^T w^{(1)}$ and $r^T w^{(2)}$ corresponding to the two portfolios $w^{(1)}$ and $w^{(2)}$. Thus we say that portfolio $w^{(1)}$ dominates portfolio $w^{(2)}$ under the first order stochastic dominance, written as $w^{(1)} \succeq_1 w^{(1)}$ if $R(w^{(1)}) \succeq_1 R(w^{(2)})$. Thus $w^{(1)} \succeq_1 w^{(2)}$ if

$$F\left(R(w^{(1)});\eta\right) \leq F\left(R(w^{(2)});\eta\right) \qquad \forall\ \eta \in \mathbf{R}. \tag{6.58}$$

Here $F\left(R(w^{(1)}),\cdot\right)$ is the cumulative distribution function of the random variable $R(w)$.

We say that $w^{(1)} \succ_1 w^{(2)}$ if $R(w^{(1)}) \succeq_1 R(w^{(2)})$ and $R(w^{(2)}) \not\succeq_1 R(w^{(1)})$. This amounts to saying that at least one inequality is strict at (6.58).

Similarly $w^{(1)} \succeq_2 w^{(2)}$ if $R(w^{(1)}) \succeq_2 R(w^{(2)})$, i.e.

$$F_2\left(R(w^{(1)});\eta\right) \leq F_2\left(R(w^{(2)});\eta\right) \quad \forall\ \eta \in \mathbf{R}, \tag{6.59}$$

where

$$F_2(V;\eta) = \int_{-\infty}^{\eta} F(V,\xi)d\xi, \quad \forall\ \eta \in \mathbf{R}.$$

We say that $w^{(1)} \succ_2 w^{(2)}$ if $R(w^{(1)}) \succeq_2 R(w^{(2)})$ and $R(w^{(2)}) \not\succeq R(w^{(1)})$. This amounts to saying that at least one inequality is strict at (6.59).

Since stochastic dominance rules also try to capture preference between two random alternatives, it is natural that there should be some connection between these rules and associated utility functions. We have following results in this regard (Sriboonchitta et al. [128])

**Theorem 6.10.1** *Let $w^{(1)}$ and $w^{(2)}$ be two portfolios. Then $R(w^{(1)}) \succeq_1 R(w^{(2)})$ if and only if*

$$E\left(u(R(w^{(1)}))\right) \geq E\left(u(R(w^{(2)}))\right),$$

*for any nondecreasing function $u : \mathbf{R} \to \mathbf{R}$ for which these expectations are finite.*

**Theorem 6.10.2** *Let $w^{(1)}$ and $w^{(1)}$ be two portfolios. Then $R(w^{(1)}) \succeq_2 R(w^{(2)})$ if and only if*

$$E\left(u(R(w^{(1)}))\right) \geq E\left(u(R(w^{(2)}))\right),$$

*for any nondecreasing function $u : \mathbf{R} \to \mathbf{R}$ for which these expectations are finite.*

**Definition 6.10.1 (FSD-efficient portfolio)** *A feasible portfolio $\overline{w}$ is called FSD- efficient if there is no other feasible portfolio $w$ such that $R(w) \succ_1 R(\overline{w})$.*

**Definition 6.10.2 (SSD-efficient portfolio)** *A feasible portfolio $\overline{w}$ is called SSD- efficient if there is no other feasible portfolio $w$ such that $R(w) \succ_2 R(\overline{w})$.*

## 6.11 Portfolio Optimization with Stochastic Dominance Constraints

In view of the above discussion it makes sense to study a typical portfolio optimization problem having stochastic dominance constraints. Let us assume that a reference random return $Y$ having a finite expected value is available. For example this could be the (random) return of a reference portfolio like an index portfolio or the current portfolio itself. Then we introduce the following problem

$$\begin{aligned} &\text{Max} \quad E\left(R(w)\right) \\ &\text{subject to,} \\ &\qquad R(w) \succeq_2 Y \\ &\qquad e^T w \ = \ 1, \end{aligned} \tag{6.60}$$

where $R(w) = r^T w$.

The problem (6.60) aims to determine those portfolios whose return dominate the reference return under second order stochastic dominance rule. This problem has continum of constraints because of the presence of the constraints $R(w) \succeq_2 Y$, and as such is difficult to handle. Never the less, the application of stochastic dominance in portfolio selection problem has been an important recent contribution in the literature. Some notable contributions in this direction are due to Dentcheva and Ruszczynski [37], Ruszczynski and Vanderbei [116], and Fáබián, Mitra and Roman [44].



## 6.12 Summary and Additional Notes

- This chapter presents certain developments post Markowitz's mean-variance theory with regard to other risk measures.
- Section 6.2 discusses certain theoritical and computational issues pertaining to Markowitz's model and need to discuss other risk measures.
- Section 6.3 presents mean absolute deviation based portfolio optimization whereas Section 6.4 discusses minimax rule based portfolio optimization.
- Sections 6.5 and 6.6 are devoted to the study of value at risk and conditional value at risk respectively. These are quantile based risk measures, in contrast to moment based risk measures ($L_1, L_2$ and $L_\infty$) which are discussed in earlier sections.
- The expected utility maximization principle of Von Neumann - Mongestern's is important for rational decision making. This requires some discussion on

preference relation. After presenting a very brief introduction of these topics in Section 6.7, Markowitz's model is revisited in Section 6.8 in utility theory frame work.

- Sections 6.9 and 6.10 respectively introduce coherent risk measures and stochastic dominance. The application of stochastic dominance in portfolio optimization is relatively new. These two topics provide a very basic theoretical frame work of portfolio optimization. Much of the current research on portfolio optimization is devoted to the application of this theoretical frame work to study the real life portfolio selection problems.
- We have not discussed certain other interesting portfolio optimization problems. For example we have not included transaction costs in our study. Also we have not looked into multi period and continuous portfolio optimization problems. We may refer to Yu et al. [148], Zhou and Li [150] and references cited therein in this regard. An interesting application of dynamic programming to study multiperiod portfolio optimization problem is presented in Sun et al. [131].
- Is there any tool which can help us to describe the possible dependency of several random variables? Let us first try to understand the question itself. We explain it through a simple example. Suppose $X_1$ and $X_2$ are two real-valued random variables that can take values from $\{1, 2, \ldots, 6\}$. Assume that we know the value of $X_1$ and we wish to infer the value of $X_2$ from it. An important thing to know here is that how much information we gained from knowing the value of $X_1$! In other words, it is same if we ask what is the interrelation or dependence of these two random variables? In certain simple situations we can indeed find the answer. For instance, match the above situation with the procedure of throwing a fair die twice and the outcome on the first throw is $X_1$ and the one of the second is $X_2$. The variables are independent; the knowledge of $X_1$ gives us no information about $X_2$. The situation will be quite different if it is given that $X_1$ is always the smaller number of the two throws and $X_2$ is the larger one. We then have a strict monotonic relation $X_1 \leq X_2$. Now, if $X_1 = 6$, then we also know $X_2 = 6$. If $X_1 = 5$, we would guess that $X_2$ is either 5 or 6, both with a probability $1/2$, and so on. We need some mathematical tool to depict the latter dependency. It is precisely what we ask for in the beginning. The answer lies in the notion called *Copulas*. Copulas are the tools for modeling and describing dependence of several random variables. The name first appeared, in context of finance, in the work of Sklar (1959) and is derived from the Latin word 'copulare', meaning 'to connect or to join'.

  The concept of copula has received growing attention in finance and economics in recent years. In fact one of the main issues of risk management is the aggregation of individual risks. Copulas have been used to aggregate risk in [42] and [18] and many more references cited therein. They are often used to identify market risk, credit risk and operational risk. Though in this book we have used correlation coefficient to compute interdependence of returns of two or more assets in a portfolio, yet it is equally important to inform readers about a relatively complex but highly useful notion of copula. Useful because, correlation parameter works well with normal distributions, while distributions in financial markets are mostly skewed and copulas are very handy to deal with the skewness. For this reason, they have been applied in option pricing, portfolio value-at-risk, and interest rates derivatives, to name a few. There are various types of copulas, like, Frank copula, normal (Gaussian) copula, student copula, singular copula, and many more. For more details on how copulas are used in risk management, we can refer to [17, 18, 29], and numerous excellent articles on web. We encourage you to hunt for this concept and its applications in finance.

## 6.13 Exercises

**Exercise 6.1** *Consider the data of Exercise 1.1 of Chapter 1. Let there be Rs 50,000 to be invested on these three assets. Formulate and solve the resulting portfolio optimization problem if*



*(i)* $L_1$ *risk measure is used.*
*(ii)* $L_2$ *risk measure is used.*
*(iii)* $L_\infty$ *risk measure is used.*

**Exercise 6.2** *Suppose that the portfolios are constructed using three securities* $a_1, a_2, a_3$ *with expected returns* $\mu_1 = 20\%, \mu_2 = 15\%$ *and* $\mu_3 = 4\%$. *Further, let expected absolute deviations be* $q_1 = 50, q_2 = 400$ *and* $q_3 = 10$ *respectively. Let there be* $M_0 = Rs\ 50{,}000$ *to be invested. Then for the risk tolerance parameter* $\lambda = 0.5$, *obtain the minimax rule based optimal portfolio.*

**Exercise 6.3** *Solve the above problem if* $q_3 = 0$ *and there is no change in the other data.*

**Exercise 6.4** *An investor has utility function* $u(x) = x^{1/4}$ *for his/her salary. He/she has a new job offer which pays Rs 80,000 with a bonus. The bonus will be Rs 0, Rs 10,000, Rs 20,000, Rs 30,000, Rs 40,000, Rs 50,000, or Rs 60,000 each with equal probability. Find the certainty equivalent of this job offer.*

**Exercise 6.5** *Let* $X \sim \mathcal{N}(\mu, \sigma^2)$. *Show that*

*(i) for each* $\alpha \in (0,1)$, $F_X^{-1}(\alpha)$ *is proportional to* $\alpha$.
*(ii) if* $X \leq Y$, *a.s. then* $F_X^{-1}(\alpha) \leq F_Y^{-1}(\alpha)$.
*(iii)* $F_{\lambda X}^{-1}(\alpha) = \lambda F_X^{-1}(\alpha)$ *for* $\lambda > 0$.
*(iv)* $F_{X+a}^{-1}(\alpha) = F_X^{-1}(\alpha) + a$ *for any* $a \in \mathbf{R}$.

**Exercise 6.6** *Find* $Var_{(1-\alpha)}(X)$ *when*

*(i)* $X \sim$ *uniform[a,b].*
*(ii)* $X \sim \mathcal{N}(\mu, \sigma^2)$.
*(iii)* $\log X \sim \mathcal{N}(\mu, \sigma^2)$.

**Exercise 6.7** *Compute* $CVar_{(1-\alpha)}(X)$ *when*

*(i)* $X \sim \mathcal{N}(\mu, \sigma^2)$.
*(ii)* $\log X \sim \mathcal{N}(\mu, \sigma^2)$.

**Exercise 6.8** *Let* $\mathfrak{X}$ *be the class of profit or loss random variables with finite variances. Verify that*

*(i)* $\mathfrak{X}$ *is a convex cone.*
*(ii) If* $\rho : \mathfrak{X} \to \mathbf{R}^+$ *given by* $\rho(X) = Var(X)$, *then* $\rho$ *is not a coherent risk measure.*

**Exercise 6.9** *Let X and Y be two random variables such that*

$$P(X = 2) = 1 - P(X = 12) = 1/5,$$

*and*

$$P(Y = 8) = 1 - P(Y = 18) = 4/5.$$

*(i) Verify that* $E(X) = E(Y)$ *and* $V(X) = V(Y)$.
*(ii) Consider the utility function* $u(x) = \log x$ *for* $x > 0$. *Does any risk averse investor feel the indifference between the two risky prospects X and Y?*
*(Hint: Is* $E\left(u(X)\right) = E\left(u(Y)\right)$*?*

**Exercise 6.10** *Let $\succ$ be a preference relation on a given set* $\mathbf{Z}$. *Show that*

*(i)* $z \succ x$ *and* $x \approx y \Rightarrow z \succ y$.
*(ii)* $z \succ x$ *and* $z \approx y \Rightarrow y \succ x$.

**Exercise 6.11** *Prove that*

*(i)* $X \succeq_1 Y \Rightarrow X \succeq_2 Y$.
*(ii)* $X \succeq_2 Y, E(X) = E(Y) \Rightarrow V(X) \leq V(Y)$.
*(iii)* $X \preceq_1 Y \Leftrightarrow F_X^{-1}(\cdot) \leq F_Y^{-1}(\cdot)$.

**Exercise 6.12** *Let* $(t - X)^+ = Max(t - X, 0)$. *Show that* $X \succeq_2 Y$ *if and only if* $E\left((t - X)^+\right) \leq E\left((t - Y)^+\right)$ *for all* $t \in \mathbf{R}$.

# 7
# Stochastic Processes

## 7.1 Introduction

In our earlier chapters, we have been making frequent use of terms like *random variables* and *random vectors*. In fact, any first course on probability theory discusses these concepts in detail. The aim of this chapter is to go beyond the notion of a random vector, and introduce a collection of random variables parameterized by a parameter, say time. Such a collection of random variables is essentially a *stochastic process*.

Stochastic processes occur naturally in real life applications. For instance consider the (opening) exchange rate: Indian rupees (INR)/US dollar (USD) at every day between August 18, 2011 and February 2, 2012 except weekends. The actual exchange rate is shown in Fig. 7.1. As on any given day, the exchange rate is random, we can interpret this figure as a realization $x_n$ of the random variable $X_n$, where $n$ is the day $1, 2, \ldots$. In order to make a guess of the interest rate on a future date, it is reasonable to look at the whole evolution of $X_n$ between August 18, 2011 and February 2, 2012. Therefore, there is a need into develop theory which provides almost continuous information about the process considered. A mathematical representation for describing such a phenomenon leads to the notion of stochastic process.

Stochastic processes play a vital role in financial mathematics, because the proper understanding of asset dynamics is crucial for making any meaningful financial decision. The derivation of Black-Scholes formula for option pricing illustrates this point very well. The derivative pricing and interest rate modeling can be studied in greater depth using the concept of stochastic process. In this chapter, we aim to present a brief introduction of stochastic processes and some of the related concepts keeping finance in view.

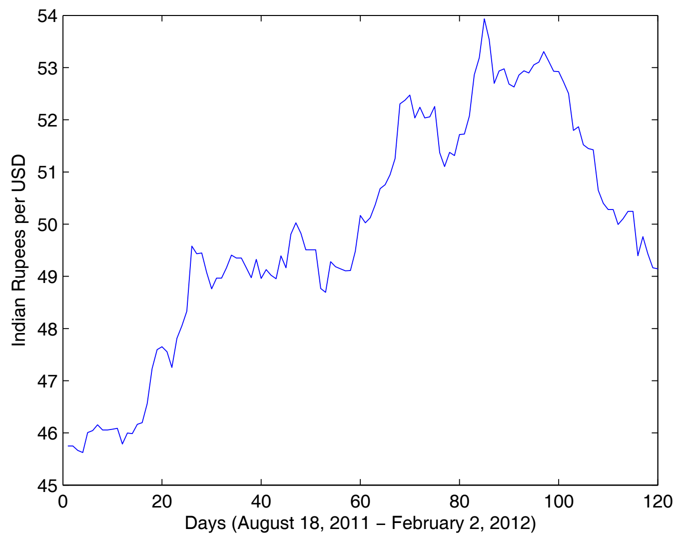



**Fig. 7.1.** Exchange Rate: Indian Rupees (INR) / US Dollar (USD) from August 18, 2011 to February 2, 2012



## 7.2 Definitions and Simple Stochastic Processes

In this section, we first review certain definitions for random variables and then proceed formally to define a stochastic process. We shall denote the random experiment by $\mathcal{E}$ and the corresponding sample space by $\Omega$.

**Definition 7.2.1** **($\sigma$-field)** *A $\sigma$-field $\mathcal{F}$ (or $\sigma$-algebra) is a family of subsets of $\Omega$ which satisfies the following properties*

*(i)* $\emptyset \in \mathcal{F}$,
*(ii) if* $A \in \mathcal{F}$ *then its complement* $A^c \in \mathcal{F}$,
*(iii) if a countable sequence of sets* $A_1, A_2, \dots$ *is in* $\mathcal{F}$, *then* $\cup_i A_i \in \mathcal{F}$.

From the above it is simple to conclude that $\Omega \in \mathcal{F}$, and if a countable sequence of sets $A_1, A_2, \dots$ is in $\mathcal{F}$ then $\cap_i A_i \in \mathcal{F}$.

Obviously, the smallest possible $\sigma$-field is $\{\emptyset, \Omega\}$. It is called the *trivial $\sigma$-field.* When the set $\Omega$ is finite, the collection of all subsets of $\Omega$ is a $\sigma$-field, which is same as the power set of $\Omega$. It is called the *total $\sigma$-field* and is the largest $\sigma$-field on $\Omega$. The set $\{\emptyset, A, A^c, \Omega\}$ where $A$ is a proper subset of $\Omega$ ($\emptyset \subset A \subset \Omega$), is another $\sigma$-field on $\Omega$; it is called the *$\sigma$-field generated by the set* $A$. The *$\sigma$-field $\mathcal{F}$ generated by two proper subsets* $A$ and $B$ is

$$\mathcal{F} = \{\emptyset, A, B, A^c, B^c, A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c, (A \vartriangle B)^c, (A \vartriangle B),$$
$$A \cup B, A \cup B^c, A^c \cup B, (A \cap B)^c, \Omega\}.$$

Here $A \vartriangle B$ is defined as $(A \cap B^c) \cup (B \cap A^c)$ and is called the symmetric difference of sets $A$ and $B$. Further it may be noted that given two $\sigma$-fields $\mathcal{F}_1$ and $\mathcal{F}_2$, $\mathcal{F}_1 \cap \mathcal{F}_2$ is always a $\sigma$-field but $\mathcal{F}_1 \cup \mathcal{F}_2$ may not be a $\sigma$-field.

**Example 7.2.1** *Let $\Omega = \{1, 2, 3\}$ and $\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \Omega\}$. Verify that $\mathcal{F}$ is a $\sigma$-field on $\Omega$.*

**Solution** We must show that properties (i), (ii) and (iii) of Definition 7.2.1 of $\sigma$-field are satisfied by given $\mathcal{F}$. We have $\emptyset \in \mathcal{F}$. Also $\emptyset^c = \Omega, \{1\}^c = \{2, 3\}, \{1, 2\}^c = \{3\}$ and so on. Hence (ii) is satisfied. Also, $\{1\} \cup \{2, 3\} = \{1, 2, 3\}, \{1\} \cup \{3\} = \{1, 3\}$. Further this property can be checked for other sets as well. Hence $\mathcal{F}$ satisfies (i), (ii) and (iii). Therefore, $\mathcal{F}$ is a $\sigma$-field on $\Omega$.

□

**Example 7.2.2** *Let $\Omega = \{a, b, c, d\}$. Let $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$ and $\mathcal{F}_2 = \{\emptyset, \{a, b\}, \{c, d\}, \Omega\}$. Show that $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a $\sigma$-field. What is the largest $\sigma$-field in $\Omega$?*

**Solution** By the definition of $\sigma$-field, $\mathcal{F}_1$ and $\mathcal{F}_2$ are $\sigma$-fields. Now, $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \{a\}, \{a, b\}, \{c, d\}, \{b, c, d\}, \Omega\}$. Clearly, $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a $\sigma$-field because $\{a\} \cup \{c, d\} = \{a, c, d\} \notin \mathcal{F}_1 \cup \mathcal{F}_2$. Further, $\mathcal{F} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{c, d\}, \{a, c\}, \{b, d\}, \{a, d\}, \{b, c\}, \{a, b, c\}, \{b, c, d\}, \{a, c, d\}, \{a, b, d\}, \Omega\}$ is the largest $\sigma$-field.

□

Let $\mathcal{F}$ be the given $\sigma$-field. Then $\mathcal{F}$ consists of those subsets of $\Omega$ which satisfy conditions (i), (ii) and (iii) of Definition 7.2.1. These subsets of $\Omega$ are called *$\mathcal{F}$-events* or *$\mathcal{F}$-measurable sets*. The pair $(\Omega, \mathcal{F})$ is called a *measurable space*. Measurable spaces are the building blocks for defining probability spaces.

**Definition 7.2.2 (Probability Measure)** *Let $\mathcal{F}$ be a $\sigma$-field over $\Omega$. Let $P$ be a real-valued function defined on $\mathcal{F}$ such that*

*(i) $P(A) \geq 0$ for all $A \in \mathcal{F}$,*
*(ii) $P(\Omega) = 1$,*
*(iii) if $A_1, A_2, A_3, \ldots$ are mutually exclusive events, then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

*Then $P(.)$ is called a probability measure. The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space.*

In most of the random experiments, either it is not feasible or it is difficult to obtain the probability of an event using definition. However, it is possible to calculate the probability of an event and related probability distributions by mapping the set of all possible outcomes into real line. This leads to the definition of a random variable.

**Definition 7.2.3 (Random Variable)** *Let $(\Omega, \mathcal{F}, P)$ be given probability space. A function $X : \Omega \to \mathbf{R}$ is called a random variable if for any $x$ in $\mathbf{R}$, $X^{-1}\{(-\infty, x]\}$ belongs to $\mathcal{F}$. Here*

$$X^{-1}\{(-\infty, x]\} = \{w \in \Omega : X(w) \in (-\infty, x]\} \ .$$

**Remark 7.2.1** *For a given probability space $(\Omega, \mathcal{F}, P)$, when $\mathcal{F}$ is the largest $\sigma$-field on $\Omega$, any real-valued function defined on $\Omega$ will be a random variable.*

**Remark 7.2.2** *Let $\Omega = \mathbf{R}$ and $\mathcal{U}^{(1)} = \{(a, b] : -\infty < a < b < \infty\}$. Then the $\sigma$-field generated by $\mathcal{U}^{(1)}$ is called the Borel $\sigma$-field, and its elements are called Borel sets. Thus Borel $\sigma$-field is the $\sigma$-field generated by the family of semi-open intervals of type $(a, b]$, $-\infty < a < b < \infty$.*

**Remark 7.2.3** *A function $f$ from a measurable space $(\Omega, \mathcal{F})$ into $(\mathbf{R}, \mathcal{B})$ is a said to be measurable function if for each Borel set $B \in \mathcal{B}$, the set $\{w \in \Omega : f(w) \in B\} \in \mathcal{F}$. Here $\mathcal{B}$ is the Borel $\sigma$-field of subsets of $\mathbf{R}$. From the above definition, a random variable $X$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbf{R}, \mathcal{B})$.*

**Remark 7.2.4** *As $X : \Omega \to \mathbf{R}$, a statement of type $a \le X \le b$ is to be understood as $\{w \in \Omega : a \le X(w) \le b\}$. In that case $P(a \le X \le b) = P\{w \in \Omega : a \le X(w) \le b\}$. In general for a Borel set $\mathcal{B}$, $P(X \in \mathcal{B}) = P\{w \in \Omega : X(w) \in \mathcal{B}\}$. Here it is assumed that $(\Omega, \mathcal{F}, P)$ is the underlying probability space.*

The function which records the probabilities associated with the random variable is defined as below.

**Definition 7.2.4 (Cumulative Distribution Function of a Random Variable)** *The cumulative distribution function (CDF) of a random variable $X$ is given as*

$$F_X(x) = P(X \le x) = P(w \in \Omega : X(w) \le x), \quad -\infty < x < \infty \ .$$

*For $a < b$, we have*

$$P(a < X \le b) = P(w \in \Omega : a < X(w) \le b) = F_X(b) - F_X(a) \ .$$

**Example 7.2.3** *Let* $\Omega = \{a, b, c, d\}$ *and* $\mathcal{F} = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$. *Let* $(\Omega, \mathcal{F}, P)$ *be the given probability space where* $P(\{a\}) = 1/4$ *and* $P(\{b, c, d\}) = 3/4$. *Define*

$$X(w) = \begin{cases} 0, & w = a \\ 1, & w = b \\ 2, & w = c \\ 3, & w = d, \end{cases}$$

*and*

$$Y(w) = \begin{cases} 0, & w = a \\ 1, & w \in \{b, c, d\}. \end{cases}$$

*Check whether X and Y are random variables? If so, find the corresponding CDFs.*

**Solution** As $X^{-1}\{(-\infty, 1]\} = \{a, b\} \notin \mathcal{F}$, the function $X$ is not a random variable. Hence CDF of $X$ is not defined.

Next, for $Y$ we have

$$Y^{-1}\{(-\infty, y]\} = \begin{cases} \emptyset, & -\infty < y < 0 \\ \{a\}, & 0 \le y < 1 \\ \Omega, & 1 \le y < \infty\ . \end{cases}$$



Since, for every $y \in \mathbf{R}$, $Y^{-1}\{(-\infty, y]\} \in \mathcal{F}$, $Y$ is a random variable. The CDF of $Y$ is given by

$$F_Y(y) = \begin{cases} 0, & -\infty < y < 0 \\ \frac{1}{4}, & 0 \le y < 1 \\ 1, & 1 \le y < \infty\ . \end{cases}$$

□

To understand uncertainty in a static system, we employ standard probability theory. But to handle uncertainty in a dynamic system, e.g. the price of a share of stock, we need to go beyond to capture evolution over time. This can be studied via a collection of random variables namely a stochastic process with time as index set. Now, we define the stochastic process formally.

**Definition 7.2.5 (Stochastic Process)** *Let* $(\Omega, \mathcal{F}, P)$ *be a given probability space. A collection of random variables* $\{X(t), t \in T\}$ *defined on the probability space* $(\Omega, \mathcal{F}, P)$ *is called a stochastic process.*

A stochastic process is also called a random process or a chance process.

**Remark 7.2.5** *Given a probability space $(\Omega, \mathcal{F}, P)$, the stochastic process $\{X(t), t \in T\}$ can be identified as a real-valued function $X : T \times \Omega \to \mathbf{R}$ of two independent variables $t \in T, w \in \Omega$ such that $X^{-1}(t)\{(-\infty, x]\}$ belongs to $\mathcal{F}$ for every $x \in \mathbf{R}$ and for every $t \in T$. Here*

$$X^{-1}(t)\{(-\infty, x]\} = \{w \in \Omega : X(t, w) \in (-\infty, x]\} \ .$$

*In this sense, the stochastic process $\{X(t), t \in T\}$ can be thought of as $\{X(t, w) : t \in T, w \in \Omega\}$.*

**Remark 7.2.6** *The mapping $X$ gives rise to two mappings, $X(\cdot, w)$ and $X(t, \cdot)$ as follows. For fixed $w \in \Omega$, the mapping $X(., w) : T \to \mathbf{R}$ takes $t$ to $X(t, w)$. The sequence of real numbers $\{X(t, w), t \in T\}$ for fixed $w \in \Omega$ is called the realization or trajectory or sample path of the stochastic process. This is not a random variable. Next, for each fixed $t \in T$, the mapping $X(t, .) : \Omega \to \mathbf{R}$ takes $w$ to $X(t, w)$ which is a random variable.*

**Definition 7.2.6 (Parameter Space and State Space)** *Let $\{X(t), t \in T\}$ be a given stochastic process. The set $\{t \in T\}$ is called the parameter space or index set. The collection of all possible values of $X(t)$ for $t \in T$ is called the state space. The state space is denoted by $S$.*

A given stochastic process is classified based on the values of state space and parameter space. For example, if both the state space and parameter space are finite or countably infinite, then it is called discrete time discrete space stochastic process. We shall be denoting a *discrete time stochastic process* by $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ while the *continuous time stochastic process* will be denoted by $\{X(t), t \in T\}$. Note that, a time series $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ is a discrete time stochastic process with $T = \mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$.

Based on parameter space $T$ and state space $S$, we now present some of the examples of different stochastic processes.

(i) Continuous time discrete space stochastic process:
$T = \mathbf{R}_+ = [0, \infty)$, $S = \mathbf{N}$, e.g. total number of shares $\{X(t), t \in [0, \infty)\}$ held by an investor at any time $t$,

(ii) Discrete time continuous space stochastic process:
$T = \mathbf{N}$, $S = \mathbf{R}_+ = [0, \infty)$, e.g. $\{X_n, n = 1, 2, 3, \ldots\}$ value of one US dollar (USD) in rupees at the end of $n$-th day.

(iii) Continuous time continuous space stochastic process:
$T = \mathbf{R}^+ = [0, \infty)$, $S = \mathbf{R}_+ = [0, \infty)$, e.g. $\{X(t), t \in [0, \infty)\}$ stock price of a particular item at any time $t$.

(iv) Discrete time discrete space stochastic process:
$T = \mathbf{N}$, $S = \mathbf{N}$, e.g. total number of shares $\{X_n, n = 1, 2, 3 \ldots\}$ held by a particular person at the end of $n$-th day.

We already know that a random variable $X$ can be characterized by its CDF $F_X(.)$. The main question now is to characterize a stochastic process. For this we need to define *finite dimensional distributions of the given process.*

**Definition 7.2.7 (Finite Dimensional Distributions of a Stochastic Process)** *Let $\{X(t), t \in T\}$ be a given stochastic process. Then finite dimensional distributions of the given process are the distributions of the finite dimensional random vectors $(X(t_1), X(t_2), \ldots, X(t_n))$, $t_1, t_2, \ldots, t_n \in T$, for all possible choices of times $t_1, t_2, \ldots, t_n \in T$ and for all $n \geq 1$.*

It can be shown that the stochastic process $\{X(t), t \in T\}$, is known completely if all its finite dimensional distributions are known.

Now we introduce certain properties of stochastic process $\{X(t), t \in T\}$.

**Definition 7.2.8 (Strict Sense Stationary Stochastic Process)** *The stochastic process $\{X(t), t \geq 0\}$ is called strict sense stationary stochastic process if, for arbitrarily chosen $n$, and for $0 < t_1 < t_2 < \ldots < t_n$, the finite dimensional random vectors $(X(t_1), X(t_2), \ldots, X(t_n))$ and $(X(t_1 + h), X(t_2 + h), \ldots, X(t_n + h))$ have the same joint distributions for all $h > 0$ and all $t_1, t_2, \ldots, t_n$.*

Note that a strict sense stationary stochastic process is invariant with respect to time translation. It is also called a *strong stationary stochastic process.*

**Definition 7.2.9 (Wide Sense Stationary Stochastic Process)** *The stochastic process $\{X(t), t \geq 0\}$ is wide sense stationary if it satisfies the following*

*(i)* $E(X(t))$ *is independent of* $t$,
*(ii)* $Cov(X(t), X(s))$ *depends only on the time difference* $|t - s|$ *for all* $t, s$,
*(iii)* $E((X(t))^2) < \infty$ *(finite second order moment).*

A wide sense stationary stochastic process is also called a *covariance stationary* or *weak stationary* or *second order stationary* stochastic process.

In other words, a stochastic process with finite second order moments is said to be weak stationary if its mean (or expectation) function is independent of $t$ and covariance function is invariant under time shift.

**Example 7.2.4** *Consider the process $X(t) = A\cos(\theta t) + B\sin(\theta t)$ where $A$ and $B$ are uncorrelated random variables with mean 0 and variance 1 and $\theta$ is a positive constant. Is $\{X(t), t \geq 0\}$ covariance or wide sense stationary stochastic process?*

**Solution** Since $E(A) = 0$ and $E(B) = 0$, we have, for a fixed $t$,

$$E(X(t)) = \cos(\theta t)E(A) + \sin(\theta t)E(B) = 0.$$

Hence $E(X(t))$ is independent of $t$.

Now for $s < t$,

$$\begin{aligned} Cov(X(t), X(s)) &= E((X(t) - 0)(X(s) - 0)) \\ &= E(X(t)X(s)) \\ &= E\left((A\cos(\theta t) + B\sin(\theta t))\left(A\cos(st) + B\sin(st)\right)\right). \end{aligned}$$

Since $E(A^2) = 1$ and $E(AB) = E(A)E(B) = 0$, we get

$$\begin{aligned} Cov(X(t), X(s)) &= \cos(\theta t)\cos(\theta s) + \sin(\theta t)\sin(\theta s) \\ &= \cos(t - s)\theta. \end{aligned}$$

Hence $Cov(X(t), X(s))$ depends only on $| t - s |$.

Next for fixed $t$,

$$\begin{aligned} E(X^2(t)) &= Var(X(t)) - (E(X(t)))^2 \\ &= \cos^2(\theta t)Var(A) + \sin^2(\theta t)Var(B) + 2\cos(\theta t)\sin(\theta t)Cov(A, B) \\ &= \cos^2(\theta t) + \sin^2(\theta t) \\ &= 1. \end{aligned}$$



Hence, by Definition 7.2.9, $\{X(t), t \geq 0\}$ is a wide sense stationary process.

□

**Definition 7.2.10 (Stationary Increments)** *The process* $\{X(t), t \in T\}$ *is said to have stationary increments if* $X(t+\tau) - X(t)$ *is a strict sense stationary process for all* $\tau > 0$, *i.e. whenever* $s < t$, $X(t) - X(s)$ *has the same distribution of* $X(t-s) - X(0)$.

**Definition 7.2.11 (Independent Increments)** *If for all* $n$ *and for* $t_1 < t_2 < \ldots < t_n$, $X(t_2) - X(t_1)$, $X(t_3) - X(t_2), \ldots,$ $X(t_n) - X(t_{n-1})$ *are independent random variables, then the process is said to have independent increments.*

**Definition 7.2.12 (Independent Process)** *If for all* $n$ *and for* $t_1 < t_2 < \ldots < t_n$ *the CDF satisfies* $F(X, T) = \prod_{i=1}^{n} F(X(t_i), t_i)$ *where* $T = (t_1, t_2, \ldots, t_n)$ *and* $X = (X(t_1), X(t_2), \ldots X(t_n))$, *then the process is called an independent process. Here* $F(X, T)$ *is the joint distribution function of the random vector* $X$.

Note that, a stochastic process having independent increments is different from a stochastic process which is an independent process.

**Definition 7.2.13 (Markov Property)** *A given process* $\{X(t), t \in T\}$ *is said to have Markov property if for all* $n$ *and for all* $0 < t_1 < t_2 < \ldots < t_n < t$, *the conditional CDF satisfies*

$$P\left(X(t) \leq x / X(0) = x_0, X(t_1) = x_1, \ldots, X(t_n) = x_n\right) = P\left(X(t) \leq x / X(t_n) = x_n\right) \ .$$

*In other words, if the future prediction depends only on the current state of the stochastic process and does not depend on the past information, then it has the Markov property.*

**Definition 7.2.14 (Markov Process)** *A given stochastic process* $\{X(t), t \in T\}$ *is said to be a Markov process if it satisfies Markov property.*

A Markov process is a stochastic process with property that, given the value of $X(s)$, the values of $X(t)$, $t > s$, do not depend on the values of $X(u)$, $u < s$, i.e. the probability of any particular future behavior of the process, when it's present state is known exactly, is not altered by additional knowledge concerning it's past behavior. For instance, binomial process, Poisson process and Brownian motion are examples of Markov process whereas Gaussian process is an example of non Markov process. All these processes will be discussed later in this chapter.

Stochastic processes can be classified according to different criteria. One of them is the distribution of finite dimensional random variables. Based on this, we list few stochastic processes which are related to stochastic study of financial mathematics.

**Definition 7.2.15 (Bernoulli Process)** *A discrete time discrete space stochastic process* $\{X_n, n = 1, 2, \ldots\}$ *is called a Bernoulli process if for each* $n$, $X_n$ *is a Bernoulli random variable with parameter* $p$, $0 < p < 1$.

Here we may note that the Bernoulli process is strict sense as well as wide sense stationary process.

**Definition 7.2.16 (Binomial Process)** *Let for* $n = 1, 2, \ldots$, $S_n = X_1 + X_2 + \ldots + X_n$ *where* $X_i$*'s are mutually independent Bernoulli distributed random variables with parameter* $p$, $0 < p < 1$. *Then* $\{S_n, n = 1, 2, \ldots\}$ *is called a binomial process.*

Here we may note that $\{S_n, n = 1, 2, \ldots\}$ is a discrete time discrete space stochastic process. Also each $S_n$ is a binomial distributed random variable.

**Lemma 7.2.1** *The binomial process* $\{S_n, n = 1, 2, \ldots\}$ *is a Markov process.*

**Proof.** For the binomial process $\{S_n, n = 1, 2, \ldots\}$, we have, for $n = 2, 3, \ldots$,

$$S_n = S_{n-1} + X_n \ .$$

Hence,

$$P(S_n = x/S_{n-1} = x_{n-1}, \ldots, S_1 = x_1) = \frac{P(S_n = x, S_{n-1} = x_{n-1}, \ldots, S_1 = x_1)}{P(S_{n-1} = x_{n-1}, \ldots, S_1 = x_1)}$$

$$= \frac{P(S_n = x/S_{n-1} = x_{n-1})P(S_{n-1} = x_{n-1}/S_{n-2} = x_{n-2})\ldots P(S_2 = x_2/S_1 = x_1)P(S_1 = x_1)}{P(S_{n-1} = x_{n-1}/S_{n-2} = x_{n-2})\ldots P(S_2 = x_2/S_1 = x_1)P(S_1 = x_1)}$$
$$= \frac{P(X_n = x - x_{n-1})P(X_{n-1} = x_{n-1} - x_{n-2})\ldots P(X_2 = x_2 - x_1)P(S_1 = x_1)}{P(X_{n-1} = x_{n-1} - x_{n-2})\ldots P(X_2 = x_2 - x_1)P(S_1 = x_1)}$$
$$= P(X_n = x - x_{n-1})$$
$$= P(S_n = x/S_{n-1} = x_{n-1}) \ .$$

Thus $\{S_n, n = 1, 2, \ldots\}$ satisfies the Markov property. Hence $\{S_n, n = 1, 2, \ldots\}$ is a Markov process.

□



**Definition 7.2.17 (Gaussian Process)** *A given stochastic process* $\{X(t), t \in T\}$ *is said to be a Gaussian process if, for* $t_1, t_2, \ldots, t_n$ *and all* $n$, $(X(t_1), X(t_2), \ldots, X(t_n))$ *is a multivariate normal distributed random vector.*

We may observe the following for a Gaussian process.

(i) The joint pdf of $(X(t_1), X(t_2), \ldots, X(t_n))$ is given by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(det(\Sigma))^{1/2}} \exp\left[-\frac{1}{2}(X - \mu)^T \Sigma^{-1}(X - \mu)\right] ,$$

where

$$\mu = E(X) = (E(X(t_1)), E(X(t_2)), \ldots, E(X(t_n))) ,$$
$$\Sigma = (Cov(X(t_i), X(t_j)); i, j = 1, 2, \ldots, n),$$

and

$$Cov(X(t_i), X(t_j)) = E[(X(t_i) - E(X(t_i))(X(t_j) - E(X(t_j)))] \ .$$

(ii) A Gaussian process which is weak stationary is also strict sense stationary. Further, in any Gaussian process, its finite dimensional random variables are independent if and only if those random variables are uncorrelated, i.e. the corresponding covariance matrix $\Sigma$ is a diagonal matrix.

**Definition 7.2.18 (Symmetric Random Walk)** *Consider a random experiment of tossing a fair coin infinitely many times. Let the successive outcomes be denoted as* $w = (w_1, w_2, w_3, \ldots)$ *e.g.* $w = (w_1, w_2, w_3, \ldots) = (H, T, T, \ldots)$ *or* $(T, H, T, \ldots)$ *etc. We now define, for* $j = 1, 2, \ldots$

$$X_j = \begin{cases} 1, & \text{if} \quad w_j = H \\ -1, & \text{if} \quad w_j = T, \end{cases}$$

*and*

$$P(X_j = 1) = P(w_j = H) = 0.5 \; ; \qquad P(X_j = -1) = P(w_j = T) = 0.5 \; .$$

*Set* $S_0 = 0$*. Let*

$$S_k = \sum_{j=1}^{k} X_j, \quad (k = 1, 2, \ldots) \; .$$

*Then,* $\{S_k, k = 0, 1, \ldots\}$ *is known as a symmetric random walk.*

Now, we present the sample path of a random walk $\{S(k), k = 0, 1, \ldots\}$ for fixed $w \in \Omega$. Suppose $w = (w_1, w_2, w_3, \ldots) = (H, H, H, T, H, H, T, H, H, H, H, \ldots)$. This realization is shown in Fig. 7.2. This symmetric random walk $\{S_k, k = 0, 1, \ldots\}$ is a discrete time, discrete space stochastic process. By considering continuous time and continuous space, we get Brownian motion which will be discussed in the next section.

Next, we present the following results for a symmetric random walk.

**Theorem 7.2.1** *Let* $\{S_k \, , \, k = 0, 1, 2, \ldots\}$ *be a symmetric random walk. Then*

*(i) for each* $k$, $E\,(S_k) = 0$ *and* $Var(S_k) = k$,
*(ii) it has independent increments,*
*(iii) it has stationary increments,*
*(iv) it is a Markov process.*

**Proof.**

(i) For $j = 1, 2, \ldots$, we have $E(X_j) = 0$ and $Var(X_j) = 1$. Therefore, for $k = 1, 2, \ldots$,

$$E(S_k) = 0 \quad \text{and} \quad Var(S_k) = k \; .$$

Here we are using the fact that $X_i's$ are mutually independent.

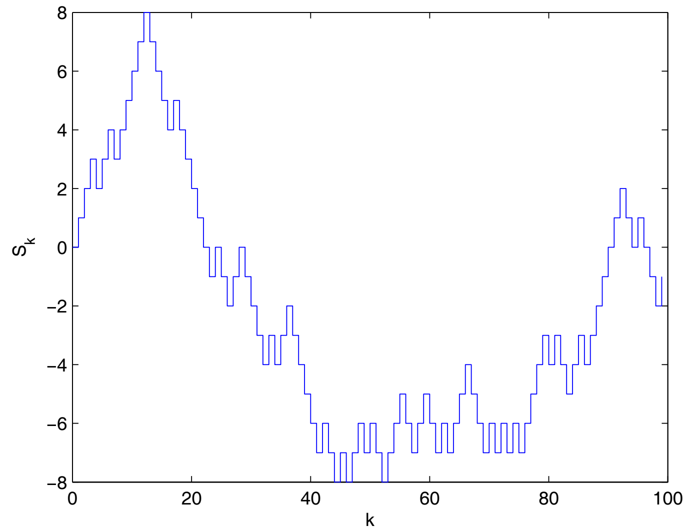



**Fig. 7.2.** Sample path of symmetric random walk



(ii) We choose an arbitrary positive integer $n$ and then choose non-negative integers $0 = k_0 < k_1 < \ldots < k_n$. Then

$$S_{k_{i+1}} - S_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j \ .$$

Since $X_j$ are independent and identically distributed (i.i.d) random variables having Bernoulli distribution, $S_{k_1} - S_{k_0}, S_{k_2} - S_{k_1}, \ldots, S_{k_n} - S_{k_{n-1}}$ are mutually independent variables. Hence the stochastic process $\{S_k, k = 0, 1, \ldots\}$ has the independent increment property.

(iii) Choose non-negative integers $k_1 < k_2$. Then

$$S_{k_2} - S_{k_1} = \sum_{j=k_1}^{k_2} X(j) \ .$$

Since $X_j$ are i.i.d random variables having Bernoulli distribution, $S_{k_2} - S_{k_1}$ has the same distribution of $S_{k_2-k_1} - S_0$. Hence the stochastic process $\{S_k, k = 0, 1, \ldots\}$ has the stationary increment property.

(iv) We have for $k = 1, 2, \ldots$

$$S_k = S_{k-1} + X_k \ .$$

Now,

$$P(S_k \leq x/S_{k-1} = x_{k-1}, \ldots, S_1 = x_1 = \frac{P(S_k \leq x, S_{k-1} = x_{k-1}, \ldots, S_1 = x_1)}{P(S_{k-1} = x_{k-1}, \ldots, S_1 = x_1)}$$

$$= \frac{P(S_k \leq x/S_{k-1} = x_{k-1}) \ldots P(S_2 = x_2/S_1 = x_1)P(S_1 = x_1)}{P(S_{k-1} = x_{k-1}/S_{k-2} = x_{k-2}) \ldots P(S_2 = x_2/S_1 = x_1)P(S_1 = x_1)}$$
$$= P(S_k \leq x/S_{k-1} = x_{k-1}) \ .$$

From the above discussion we can easily conclude that $\{S_k, k = 1, 2, \ldots\}$ is a Markov process.

□

**Definition 7.2.19 (Poisson Process)** *A stochastic process* $\{N(t), t \geq 0\}$ *is said to be a Poisson process with parameter* $\lambda > 0$*, if it satisfies the following properties*

*(i)* $N(0) = 0$*,*
*(ii) For all* $n$ *and for all* $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$*, increments* $N(t_i) - N(t_{i-1})$*,* $i = 1, 2, \ldots, n$ *are independent and stationary,*
*(iii) For* $0 \leq s < t$*,* $N(t) - N(s)$ *is a Poisson distributed random variable with parameter* $\lambda(t-s)$*, i.e.*

$$P(N(t) - N(s) = n) = \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!}, \quad (n = 0, 1, 2, \ldots) \ .$$

The sample path of a Poisson process with parameter $\lambda = 2$ is shown in Fig. 7.3. It is observed that in any finite interval, the sample path has finite jumps of size one and in the interval $[0, \infty)$ it has countably infinite jumps.

Let $\{N(t), t \geq 0\}$ be a Poisson process. Then, for each $t$, $E(N(t)) = \lambda t$ and $Var(N(t)) = \lambda t$.

**Remark 7.2.7** *By property (ii) of Definition 7.2.19, for every* $s < t$*, the increment* $N(t) - N(s)$ *is independent from the history of the process up to time* $s$ *and has the same law as* $N(t-s)$*. Therefore* $\{N(t), t \geq 0\}$ *satisfies Markov property. Hence, it is a Markov process.*

**Remark 7.2.8** *Suppose* $N(t)$ *denotes the number of events that occur in the interval* $[0, t]$*. Let* $T_i, i = 1, 2, \ldots$ *denote the occurrence of events. The inter arrival of successive events* $T_1, T_2 - T_1, \ldots, T_{i+1} - T_i$ *are independent and are exponentially distributed with parameter* $\lambda$*.*

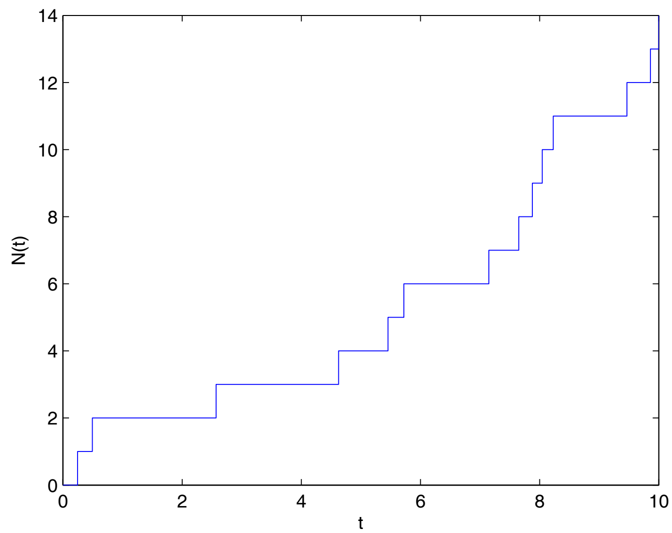



**Fig. 7.3.** Sample path of Poisson process with $\lambda = 2$



**Example 7.2.5** *Consider a financial risk model that may occur in an insurance company. Let* $N(t)$ *denote the number of claims received by time* $t$, $t \geq 0$. *Thus,* $N(t)$ *counts the number of claims received in* $(0, t]$. *Suppose the process* $\{N(t), t \geq 0\}$ *is a Poisson process with claim arrival rate 9 per month of 30 days. In a randomly chosen month of 30 days,*

*(i) what is the probability that there are exactly 4 claims were received in the first 15 days?*

*(ii) given that exactly 4 claims were received in the first 15 days, what is the probability that all the four claims were received in the last 7 days out of these 15 days?*

**Solution** We have

$$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \quad (k = 0, 1, \ldots),$$

where $\lambda = \dfrac{9}{30}$ per day.

(i) The required probability is

$$P(N(15) = 4) = \frac{e^{-\frac{9}{2}} 9^4}{384}\ .$$

(ii) The required probability is

$$\frac{P(N(8)=0)P(N(7)=4)}{P(N(15)=4)} = \frac{e^{-36/15}e^{-63/30}(63/30)^4}{e^{-9/2}(9/2)^4} = \left(\frac{7}{15}\right)^4 .$$

□

## 7.3 Brownian Motion and its Properties

The long studied model known as Brownian motion, is named after the English botanist Robert Brown. In 1827, Brown described the unusual motion exhibited by a small particle totally immersed in a liquid or a gas. Early nineteenth century, Brownian motion is introduced to model the price movements of stocks and commodities. A formal mathematical description of Brownian motion and its properties was first given by the great mathematician Norbort Wiener beginning 1918.

**Definition 7.3.1 (Brownian Motion)** *A stochastic process* $\{W(t), t \geq 0\}$ *is said to be a Brownian motion (BM), if it satisfies the following properties*

*(i)* $W(0) = 0$,
*(ii) for* $t > 0$, *the sample path of* $W(t)$ *is continuous,*
*(iii) for all* $n$ *and for all* $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$, *increments* $W(t_i) - W(t_{i-1})$, $i = 1, 2, \ldots, n$ *are independent and stationary,*
*(iv) for* $0 \leq s < t < \infty$, $W(t) - W(s)$ *is normally distributed random variable, with mean* $0$ *and variance* $(t - s)$.

A Brownian motion is also called a Wiener process. The sample path of Wiener process is shown in Fig. 7.4. Wiener process can also be visualized as a scaling limit of a symmetric random walk.

**Remark 7.3.1** *We observe that, the sample paths are always continuous functions. In other words, the sample paths do not have any jumps. On the other hand, these paths are also essentially nowhere differentiable, that is, it is not possible to define a unique tangent line at any point on a curve. We can show this by using the convergence in the second order moment. For this we shall like to show that*

$$\lim_{\Delta t \to 0} Var\left(\frac{W(t_0 + \Delta t) - W(t_0)}{\Delta t}\right)$$

*does not exist for any arbitrary point* $t_0$.

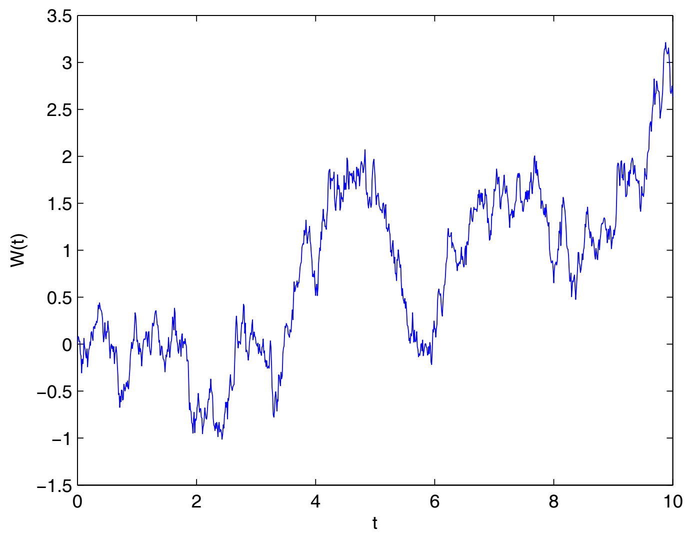



**Fig. 7.4.** Sample path of Wiener process

*We know that,* $W(t_0 + \Delta t) - W(t_0)$ *has normal distribution with mean zero and variance* $\Delta t$*, i.e.* $\mathcal{N}(0, \Delta t)$*. Hence,*

$$\lim_{\Delta t \to 0} Var\left(\frac{W(t_0 + \Delta t) - W(t_0)}{\Delta t}\right) = \lim_{\Delta t \to 0} \frac{1}{(\Delta t)^2} \times \Delta t = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \ .$$

*Thus the above limit does not exist. Hence, we conclude that Brownian motion has non-differentiability at a fixed point. We observe that nowhere differentiability of Brownian motion requires a more careful argument than non-differentiability at a fixed point. The precise mathematical proof is not presented in this book but rather the above argument may convince to conclude nowhere differentiability of Brownian motion (refer Karatzas and Shreve [75] for proof).*

**Remark 7.3.2** *The Wiener process is not wide sense stationary. This is because for* $s < t$*, the covariance function* $Cov\,(W(t), W(s))$ *is not a function of* $(t - s)$*. In fact*

$$\begin{aligned} Cov\,(W(t), W(s)) &= E[(W(t) - E(W(t)))(W(s) - E(W(s)))] \\ &= E[W(t)W(s)] \\ &= E[(W(t) - W(s) + W(s))W(s)] \\ &= E[W(t) - W(s)]E[W(s)] + E[(W(s))^2] \\ &= (0 \times 0) + s = s \ . \end{aligned}$$

*Hence,* $Cov\,(W(t), W(s)) = Min(s, t)$.

**Remark 7.3.3** *Given* $W(t)$*, the future* $W(t+h)$ *for any* $h > 0$ *only depends on the increment* $W(t+h) - W(t)$ *and this is independent of the past. Hence by Definition 7.3.1, the Markov property is satisfied. Thus,* $\{W(t), t \geq 0\}$ *is a Markov process.*

**Lemma 7.3.1** *The joint probability density function of* $(W(t_1), W(t_2))$ *is given by*

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{t_1(t_2 - t_1)}} \exp\left[-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{(t_2 - t_1)}\right\}\right] .$$

**Proof.** From Definition 7.2.17, a Wiener process is also a Gaussian process. Since $\{W(t), t \geq 0\}$ is a Markov process as well as a Gaussian process, we have

$$P\,(W(t) \leq x / W(t_n) = x_n) = P\,(W(t) - W(t_n) \leq x - x_n)$$

$$= \int_{-\infty}^{x - x_n} \frac{1}{\sqrt{2\pi(t - t_n)}} \exp\left[-\frac{s^2}{2(t - t_n)}\right] ds\ .$$

Let us now consider the joint distribution of $(W(t_1), W(t_2))$. We know that, $W(t_1)$ and $(W(t_2) - W(t_1))$ are independent. Also, $W(t_1)$ is $\mathcal{N}(0, t_1)$ and $W(t_2) - W(t_1)$ is $\mathcal{N}(0, t_2 - t_1)$. Therefore, the joint probability density function of $(W(t_1), W(t_2))$ is



$$f(x_1, x_2) = p(x_1, t_1)p(x_2 - x_1, t_2 - t_1),$$

where

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{x^2}{2t}\right], t > 0; -\infty < x < \infty\ .$$

Thus,

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{t_1(t_2 - t_1)}} \exp\left[-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{(t_2 - t_1)}\right\}\right].$$

□

**Remark 7.3.4** *The above Lemma can be generalized in the sense that for* $0 < t_1 < t_2 < \cdots < t_n$, $X = (W(t_1), W(t_2), \ldots, W(t_n))$ *is jointly normally distributed with the joint probability density function given by*

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(det(\Sigma))^{1/2}} \exp\left[-\frac{1}{2}(X - \mu)^T \Sigma^{-1}(X - \mu)\right],$$

*where*

$$\mu = E(X) = (E(W(t_1)), E(W(t_2)), \ldots, E(W(t_n))) \in \mathbf{R}^n$$

*and*

$$\Sigma = [Cov(W(t_i), W(t_j)); i, j = 1, 2, \ldots, n] \in \mathbf{R}^{n \times n} \ .$$

We now present some important properties of Brownian motion and refer to Shreve [122] for their proofs.

**Result 7.3.1** *Let* $\{W(t), t \geq 0\}$ *be a Wiener process. Then,*

*(i)* $\{-W(t), t \geq 0\}$ *is a Wiener process.* *(symmetric property)*

*(ii)* $\{\frac{1}{\sqrt{c}} W(ct), t \geq 0\}$ *is a Wiener process for each fixed* $c > 0$. *(scaling property)*

*(iii) Taking* $\widetilde{W}(0) = 0, \widetilde{W}(t) = tW(\frac{1}{t}), t > 0$, *the process* $\{\widetilde{W}(t), t \geq 0\}$ *is a Wiener process.* *(time inversion property)*

**Example 7.3.1** *Let* $\{W(t), t \geq 0\}$ *be a Wiener process. Find the conditional distribution of* $W(s)$ *for* $0 < s < t$, *given* $W(t) = x$ .

**Solution** There are two possible approaches to solve this problem. In the first approach we find the joint distribution of $(W(t), W(s))$ and then find the conditional distribution. In the second approach we use the time inversion property in Result 7.3.1. The required conditional distribution is $sW\left(\frac{1}{s}\right)$ given that $tW\left(\frac{1}{t}\right) = x$. This is same as the distribution of $sW\left(\frac{1}{s}\right) - sW\left(\frac{1}{t}\right) + sW\left(\frac{1}{t}\right)$ given $W\left(\frac{1}{t}\right) = \frac{x}{t}$. Hence, we need to find the distribution of $s\left(W\left(\frac{1}{s}\right) - W\left(\frac{1}{t}\right)\right) + \frac{sx}{t}$. This distribution is a normal distribution with mean $\frac{sx}{t}$ and variance $s^2\left(\frac{1}{s} - \frac{1}{t}\right)$.

□

## 7.4 Processes Derived from Brownian Motion

A general Brownian motion need not have $W(0) = 0$ and $\sigma^2 = 1$. Therefore we define a general Brownian motion with drift $\mu$ and variance $\sigma^2$ as follows.

**Definition 7.4.1 (Brownian motion with drift $\mu$ and volatility $\sigma$)** *A stochastic process $\{X(t), t \geq 0\}$ is said to be a Brownian motion with drift $\mu$ and volatility $\sigma$ if $X(t) = \mu t + \sigma W(t)$ where (i) $W(t)$ is a standard Brownian motion, (ii) $-\infty < \mu < \infty$ is a constant and (iii) $\sigma > 0$ is a constant.*

This is a generalization of standard Brownian motion. In this process, the mean function $E(X(t)) = \mu t$ and covariance function $Cov\,(W(s), W(t)) = \sigma^2 \text{Min}(s,t),\ s,t \geq 0$. Fig. 7.5 shows a sample path of Brownian motion with drift $\mu = 0.1$ and volatility $\sigma^2 = 1.5$.

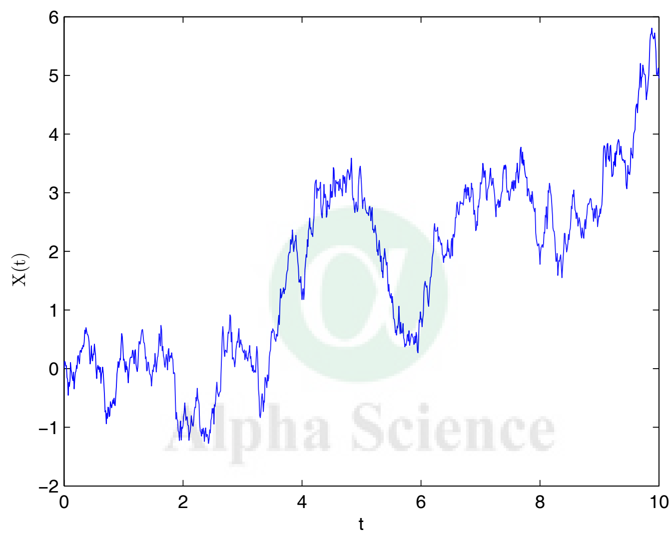



**Fig. 7.5.** Sample path of Brownian motion with $\mu = 0.1$ $\sigma^2 = 1.5$

**Definition 7.4.2 (Brownian Bridge)** *A standard Brownian bridge $\{X(t), 0 \leq t \leq 1\}$ is defined as $X(t) = W(t) - tW(1)$, where $W(t)$ is a standard Brownian motion. Clearly, $X(0) = 0 = X(1)$.*

Since for $0 < t < 1$, $W(t) \sim \mathcal{N}(0,t)$, $X(t) \sim \mathcal{N}(0, t(1-t))$. The covariance function is $s(1-t)$ for $0 \leq s \leq t \leq 1$. Therefore, the Brownian bridge is a Gaussian process but is not a Brownian motion. For fixed $T > 0$, the general Brownian bridge $\{X(t), 0 \leq t \leq T\}$ can be defined as

$$X(t) = W(t) - \frac{t}{T} W(T), \quad 0 \leq t \leq T \,.$$

The covariance function is given by

$$\begin{aligned} Cov(X(s), X(t)) &= E[X(s)X(t)] - E(X(s))E(X(t)) \\ &= E\left[W(t)W(s) + \frac{st}{T^2}W^2(T) - \frac{t}{T}W(s)W(T) - \frac{s}{T}W(t)W(T)\right] \\ &= \text{Min}\{s,t\} + \frac{st}{T} - \frac{t}{T}\text{Min}\{s,T\} - \frac{s}{T}\text{Min}\{t,T\} \\ &= \text{Min}\{s,t\} - \frac{st}{T}\ . \end{aligned}$$

Brownian motion model for stock prices is open to the objection that prices are by definition positive quantities while $W(t)$, being normally distributed, is negative with strictly positive probability. In 1965, Samuelson [117] introduced what has now become the standard model, namely geometric Brownian motion.

**Definition 7.4.3 (Geometric Brownian Motion)** *A stochastic process* $\{X(t), t \geq 0\}$ *is said to be a geometric Brownian motion (GBM) if* $X(t) = X(0)\ e^{W(t)}$ *where* $W(t)$ *is a standard Brownian motion.*

**Result 7.4.1** *For any* $h > 0$*, we have*

$$\begin{aligned} X(t+h) &= X(0)\ e^{W(t+h)} \\ &= X(0)\ e^{W(t)+W(t+h)-W(t)} \\ &= X(t)\ e^{W(t+h)-W(t)}\ . \end{aligned}$$

*We note that, BM has independent increments. Hence given* $X(t)$*, the future* $X(t+h)$ *only depends on the future increment of the BM. Thus future is independent of the past and therefore the Markov property is satisfied. Hence,* $\{X(t), t \geq 0\}$ *is a Markov process.*

Because a geometric Brownian motion is nonnegative, it provides for a more realistic model of stock prices. Also, the GBM model considers the ratio of stock prices to have the same normal distribution. Therefore, the percentage change in price as opposed to the absolute change in price is modeled by a GBM. Fig. 7.6 shows a sample path of geometric Brownian motion.

How does geometric Brownian motion relate to stock prices? One possibility is to think of modeling the rate of return of the stock price as a Brownian motion. Suppose that the stock price $S(t)$ at time $t$ is given by

$$S(t) = S(0)\ e^{H(t)},$$

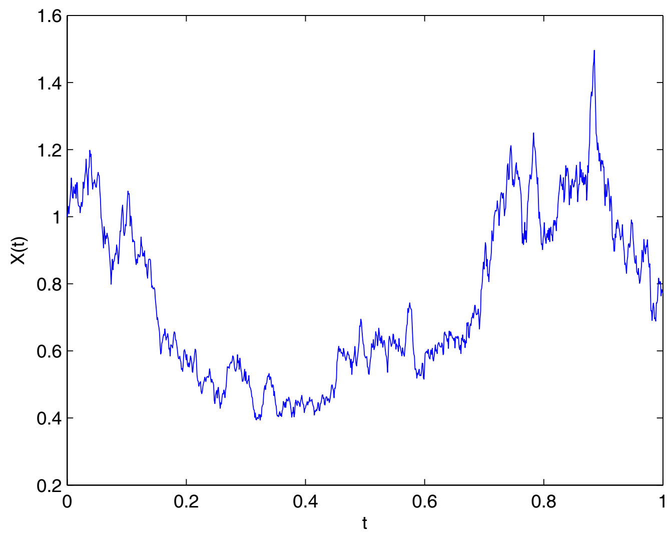



**Fig. 7.6.** Sample path of geometric Brownian motion

where $S(0)$ is the initial price and $H(t) = \mu t + \sigma W(t)$ is a Brownian motion with drift. In this case, $H(t)$ represents a continuously compounded rate of return of the stock price over the period of time $[0, t]$. Here, $H(t)$ refer to the logarithmic growth of the stock price, satisfies



$$H(t) = ln\left(\frac{S(t)}{S(0)}\right) .$$

This gives

$$ln(S(t)) = ln(S(0)) + H(t) .$$

Therefore, $ln(S(t))$ has a normal distribution with mean $\mu t + ln(S(0))$ and variance $\sigma^2 t$. As we have seen, if a random variable $X$ has the property that $lnX$ has normal distribution, then the random variable $X$ is said to have a lognormal distribution. Accordingly, $S(t)/S(0)$ is lognormal distributed random variable.

**Example 7.4.1** *Suppose that the stock price* $S(t)$ *at time* $t$ *is given by* $S(t) = S(0)\ e^{H(t)}$ *where* $S(0)$ *is the initial price and* $H(t) = \mu t + \sigma W(t)$ *is a Brownian motion with drift* $\mu$ *and volatility* $\sigma$. *Prove that*

*(i)* $E(S(t)) = S(0)\ \exp\left(\left(\mu + \frac{\sigma^2}{2}\right) t\right)$ .

*(ii)* $Var(S(t)) = \left(S(0)\ \exp\left(\left(\mu + \frac{\sigma^2}{2}\right) t\right)\right)^2 (\exp(\sigma^2 t) - 1)$ .

**Solution** We know that, if a random variable $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$, then the moment generating function of $X$ is given by

$$M_X(\theta) = E\left(e^{\theta X}\right) = \exp\left(\mu\theta + \frac{1}{2}\sigma^2\theta^2\right) . \tag{7.1}$$

Since for every $t$, $W(t)$ is normally distributed with mean zero and variance $t$, $H(t) = \mu t + \sigma W(t)$ is normally distributed with mean $\mu t$ and variance $\sigma^2 t$. Hence,

$$M_{H(t)}(\theta) = E\left(e^{\theta H(t)}\right) = \exp\left(\mu t\theta + \frac{1}{2}\sigma^2 t\theta^2\right) . \tag{7.2}$$

(i) Now

$$\begin{aligned} E(S(t)) &= E\left(S(0)e^{H(t)}\right) \\ &= S(0)E\left(e^{H(t)}\right) \end{aligned}$$

Using (7.2) and substituting $\theta = 1$, we obtain

$$E(S(t)) = S(0)\exp\left(\left(\mu + \frac{\sigma^2}{2}\right)t\right) .$$

(ii)

$$\begin{aligned} Var(S(t)) &= E(S^2(t)) - (E(S(t)))^2 \\ &= E\left(S^2(0)e^{2H(t)}\right) - S^2(0)\exp\left(2\left(\mu + \frac{\sigma^2}{2}\right)t\right) \end{aligned}$$

Using (7.2) and substituting $\theta = 2$, we obtain

$$\begin{aligned} Var(S(t)) &= S^2(0)\exp\left(2\left(\mu + \sigma^2\right)t\right) - S^2(0)\exp\left(2\left(\mu + \frac{\sigma^2}{2}\right)t\right) \\ &= \left(S(0)\exp\left(\left(\mu + \frac{\sigma^2}{2}\right)t\right)\right)^2 \left(\exp\left(\sigma^2 t\right) - 1\right) . \end{aligned}$$

□

**Remark 7.4.1** *Letting* $\bar{r} = \mu + \frac{1}{2}\sigma^2$, *we get*

$$E(S(t)) = S(0)\ e^{\bar{r}t}$$

$$Var(S(t)) = \left(S(0)\ e^{\ (\bar{r}+\frac{\sigma^2}{2})t}\right)^2 \left(e^{\sigma^2 t} - 1\right)\ .$$

*Here we observe that, the expected stock price depends not only on the drift* $\mu$ *of* $H(t)$ *but also on the volatility* $\sigma$*. Further, it shows that, the expected price grows like a fixed-income security with continuously compounded interest rate* $\bar{r}$*. In real scenario,* $r$ *is much lower than* $\bar{r}$*, the real fixed-income interest rate, that is why one invests in stocks. But the stock has variability due to the randomness of the underlying Brownian motion and hence a risk is involved here.*

**Example 7.4.2** *Suppose that stock price* $\{S(t), t \geq 0\}$ *follows geometric Brownian motion with drift* $\mu = 0.12$ *per year and volatility* $\sigma = 0.24$ *per annum. Assume that, the current price of the stock is* $S(0) = Rs\ 40$*. What is the probability that a European call option having four years to exercise time and with a strike price* $K = Rs\ 42$*, will be exercised?*

**Solution** We have

$$\begin{aligned} P\,(S(4) > 42) &= P\left(\frac{S(4)}{40} > \frac{42}{40}\right) \\ &= P\left(ln\left(\frac{S(4)}{40}\right) > ln\left(\frac{42}{40}\right)\right)\ . \end{aligned}$$

Since $ln\left(\frac{S(4)}{40}\right)$ follows normal distribution with mean 0.48 and variance $(0.48)^2$, we get

$$\begin{aligned} P\left(ln\left(\frac{S(4)}{40}\right) > ln\left(\frac{42}{40}\right)\right) &= P\left(\frac{ln\left(\frac{S(4)}{40}\right) - 0.48}{0.48} > \frac{ln\left(\frac{42}{40}\right) - 0.48}{0.48}\right) \\ &= 1 - \Phi\left(\frac{ln\left(\frac{42}{40}\right) - 0.48}{0.48}\right) \\ &= 1 - \Phi(-0.8983) \\ &= \Phi(0.8983) = 0.3133, \end{aligned}$$

where $\Phi$ is the cumulative standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

□

**Definition 7.4.4 (Ornstein-Uhlenbeck Process)** *Let* $\{W(t), t \geq 0\}$ *be a Wiener process. Define*

$$X(t) = X(0)e^{-at} + be^{-at}W(e^{2at} - 1) \ ,$$

*where* $a$ *and* $b$ *are strictly positive real numbers and* $X(0)$ *is independent of* $W(t)$*. Then, we say* $\{X(t), t \geq 0\}$ *is an Ornstein-Uhlenbeck process.*

We note that, $\{X(t), t \geq 0\}$ is a Markov process. It is also a Gaussian process if $X(0) = x(0)$ is fixed or $X(0)$ is Gaussian.

In the "real" world, we observe that asset price processes have jumps or spikes and risk-managers have to take them into account. We need processes that can describe the observed reality of financial markets in a more accurate way than models based on Brownian motion. Levy processes provide us with the appropriate framework to model both in the "real" and in the "risk-neutral" world.

**Definition 7.4.5 (Levy Process)** *A stochastic process* $\{X(t), t \geq 0\}$ *is said to be a Levy process if it satisfies the following properties*

*(i)* $X(0) = 0$,
*(ii) for all* $n$ *and for* $0 \leq t_0 < t_1 < t_2 < \ldots < t_n$*, increments* $X(t_i) - X(t_{i-1}),\ i = 1, 2, \ldots, n$*, are independent and stationary,*
*(iii) for* $a > 0$*,* $P(\mid X(t) - X(s) \mid > a) \rightarrow 0$ *when* $t \rightarrow s$*.*

**Remark 7.4.2** *Let* $b$ *be a constant and* $X(t) = bt$*. Then* $\{X(t), t \geq 0\}$ *is a Levy process.*

**Remark 7.4.3** *A Wiener process* $\{W(t), t \geq 0\}$ *defined in Definition 7.3.1, is a Levy process in* **R** *that has continuous paths and has the Gaussian distribution with mean zero and variance* $\Delta t$ *for its increments* $W(t + \Delta t) - W(t)$*. The most general continuous Levy process in* **R** *has the form* $X(t) = bt + cW(t)$, $t \geq 0$*, where* $b$ *and* $c$ *are real constants.*

**Remark 7.4.4** *A Poisson process* $\{N(t), t \geq 0\}$ *with parameter* $\lambda$ *defined in Definition 7.2.19 is a Levy process that is a counting process having the Poisson distribution with mean* $\lambda\Delta t$ *for its increments* $N(t + \Delta t) - N(t)$*.*

The above three processes namely deterministic process $X(t) = bt$, Wiener process and Poisson process are Levy processes. It turns out that all Levy processes can be built up out of these building blocks. We may refer to Applebaum [4] for further reading in this regard.

## 7.5 Summary and Additional Notes

- As early as 1900, Louis Bachelier proposed Brownian motion (also, named as Wiener process) as a model of the fluctuations of stock prices. Since the 1970s, the Wiener process has been widely applied in financial mathematics and economics to model the evolution in time of stock prices and bond interest rates.
- The earlist attempt to model Brownian motion mathematically can be traced to three different sources. The first was that of T.N. Thiele in 1880 who effectively created a model of Brownian motion while studying time series; the second was that of L. Bachelier (mentioned above) and the third was that of Einstien, who in 1905 proposed a model of the motion of small particles suspended in a liquid so as to convince other physicists of the molecular nature of matter. The readers may refer to Vassiliou [144] for further historical details.
- Section 7.2 presents the basics of stochastic process while the Brownian motion is introduced in Section 7.3.
- In Section 7.4, processes derived from Brownian motion are presented. It includes geometric Brownian motion which has been of tremendous use in modern finance theory.

## 7.6 Exercises

**Exercise 7.1** *Consider the binomial model for trading in stock, $t = 1, 2$, where at each time the stock can go up by the factor $u$ or down by the factor $d$. The sample space $\Omega = \{(u,u),(u,d),(d,u),(d,d)\}$. Create one non-trivial $\sigma$-field and the largest $\sigma$-field on $\Omega$.*

**Exercise 7.2** *Construct an example of $\sigma$-fields $\mathcal{F}(\infty)$ and $\mathcal{F}(\in)$ such that $\mathcal{F}(\infty) \cup \mathcal{F}(\in)$ is not a $\sigma$-field.*

**Exercise 7.3** *Prove that there does not exist a $\sigma$-field which contains exactly 6 elements.*

**Exercise 7.4** *Consider* $\Omega = \{1, 2, 3, 4\}$. *Let* $\sigma$*-algebra* $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3, 4\}, \Omega\}$. *Construct a random variable on the measurable space* $(\Omega, \mathcal{F})$.

**Exercise 7.5** *Consider a random experiment of tossing an unbiased coin three times. Let* $\Omega$ *denote the set of all possible outcomes. Let the random variable* $X$ *be the number of heads observed. Find the cumulative distribution function* $F(X, x)$.

**Exercise 7.6** *Let* $X_n$, *for* $n$ *even take values* $+1$ *and* $-1$ *each with probability 0.5, and for* $n$ *odd, take values* $\sqrt{a}$, $\frac{-1}{\sqrt{a}}$ *with probability* $\frac{1}{a+1}$, $\frac{a}{a+1}$ *respectively* ($a > 0$, $a \neq 1$)*. Further, let* $X_n$*'s be independent, show that the stochastic process* $\{X_n, n \geq 1\}$ *is wide sense stationary but not strict stationary.*

x

**Exercise 7.7** *Let* $\{X(t), t \geq 0\}$ *be a stochastic process with independent increments and* $X(0) = 0$. *Show that* $Cov\,(X(s), X(t)) = Var\,(X(Min(s, t)))$, *for any* $s, t$ $(t, s > 0)$ .

**Exercise 7.8** *Let* $X$ *and* $Y$ *be i.i.d. random variables each having uniform distribution on the interval* $(-\pi, \pi)$. *Let* $Z(t) = \cos(tX + Y), t \geq 0$. *Is* $\{Z(t), t \geq 0\}$ *wide sense stationary process?*

**Exercise 7.9** *Consider an urn containing 100 red balls and 100 black balls. Balls are drawn one by one without replacement. Let* $X_n$ *be the number of red balls remaining in the urn after the* $n^{th}$ *ball is drawn. Is* $\{X_n, n = 0, 1, \ldots\}$ *a Markov process?*

**Exercise 7.10** *Let* $Y_n = a_0 X_n + a_1 X_{n-1}; n = 1, 2, \ldots$ *where* $a_0, a_1$ *are constants and* $X_n$, $(n = 0, 1, \ldots)$ *are i.i.d random variables with mean 0 and variance* $\sigma^2$. *Is* $\{Y_n, n = 1, 2, \ldots\}$ *a Markov process?*

**Exercise 7.11** *Let* $\{N(t), t \geq 0\}$ *be a Poisson process with parameter* $\lambda$. *Suppose* $N(t)$ *denotes the number of events that occur in the interval* $[0, t]$, *prove that the inter arrival of successive events are independent and are exponentially distributed with parameter* $\lambda$.

**Exercise 7.12** *Let* $\{W(t), t \geq 0\}$ *be the Brownian motion. Prove that,* $E((W(t) - W(s))^4) = 3(t - s)^2$.

**Exercise 7.13** *Let* $\{W(t), t \geq 0\}$ *be the Brownian motion. Prove that* $(W(t_1), W(t_2), \ldots, W(t_n))$ *is jointly normal distributed with CDF for* $0 < t_1 < t_2 < \cdots < t_n$ *is given by*

$$P(W(t_1) \le a_1, W(t_2) \le a_2, \ldots, W(t_n) \le a_n) = \frac{1}{\sqrt{(2\pi)^n t_1(t_2 - t_1)\cdots(t_n - t_{n-1})}} \times$$
$$\int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \exp\left[-\frac{1}{2}\left(\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right)\right] dx_1 \cdots dx_n \ .$$

**Exercise 7.14** *Let $X$ be a normally distributed random variable with mean $\mu$ and variance $\sigma^2$. Let $u$ be a fixed number in $\mathbf{R}$ and define the convex function $\phi(x) = e^{ux}$ for all $x \in \mathbf{R}$. Prove that*

*(i)* $E(\phi(X)) = e^{u\mu + \frac{1}{2}u^2\sigma^2}$.
*(ii) Verify the Jensen's inequality holds* $E(\phi(X)) \ge \phi(E(X))$.

**Exercise 7.15** *Let $\{W(t), t \ge 0\}$ be a Wiener process. Find the conditional distribution of $W(t)$ given that $W(s) = c$ (where $c$ is a constant) when $s < t$.*

**Exercise 7.16** *Let $\{W(t), t \ge 0\}$ be a Wiener process. Find the conditional distribution of $W(s/2)$ given that $W(s) = x$.*

**Exercise 7.17** *Show that for any $T > 0$, $V(t) = W(t + T) - W(T)$ is a Wiener process if $W(t)$ is a Wiener process.*

**Exercise 7.18** *Let $\{W(t), t \ge 0\}$ be a Brownian motion. Prove that $\{tW(1/t), t \ge 0\}$ where $tW(1/t)$ is taken to be zero when $t = 0$, is a Brownian motion.*

**Exercise 7.19** *Let $\{W(t), t \ge 0\}$ be a Wiener process. Prove that distribution of $\{W(t), 0 \le t \le 1\}$ and $\{W(1) - W(1 - t), 0 \le t \le 1\}$ are the same.*

**Exercise 7.20** *Consider the process*

$$Y(t) = \begin{cases} W(t), & t < 1 \\ W(t) + Z, & t \ge 1 \end{cases}$$

where $W(t)$ is a Wiener process and $Z \sim \mathcal{N}(0, 1)$ and is independent of $W(t)$. Show that $\{Y(t), t \ge 0\}$ not a Levy process.

# 8
# Filtration and Martingale

## 8.1 Introduction

Conditional expectation is an extremely important concept in probability theory. Traditionally in a typical probability course, the well known *gambler's ruin problem* is presented as a motivational example. But now the concept of conditional expectation has found much favor in financial mathematics because it provides the basis for two of the most important concepts namely, the *filtration* and the *martingale*.

In gambler's ruin problem, a gambler starts with an amount of Rs $N$. Then an unbiased coin is flipped, landing head with probability $0.5$ and landing tail with probability $0.5$. If the coin lands head, he/she gains $Rs\ 1$, otherwise he/she loses $Rs\ 1$. The game continues until he/she loses all his/her amount of Rs $N$. Let $X_n$ be the fortune at the $n$th game, and $S_n$ be his/her capital after $n$ games. An interesting question here is to know his/her fortune, on an average, on the next game given his/her current fortune. To answer this question, we need the conditional expectation of the random variables $\{X_n, n = 0, 1, \ldots\}$ given the information up to $r$, $r < n$.

We next present another example which is more relevant to us. Let $X(t)$ denote the share price of a particular stock at time $t$. Assume that, we have the information about the share price up to time $s > 0$. We may like to know the expected or average share price of the stock at a future time, say, $s + 5$ given that $X(s) = 10$ (say). To make such statements precise and also to answer them, we need the conditional expectation of random variables $\{X(t), t \geq 0\}$ given the information up to time $s$, $s < t$.

The above two examples certainly make a case for the study of conditional expectation. But for a better understanding and deeper study of various other aspects we need to develop the concepts of *filtration* and *martingale* property of the underlying stochastic process.

A filtration is an increasing sequence of $\sigma$-fields on a measurable space $(\Omega, \mathcal{F})$. Since a $\sigma$-field defines the set of events that can be measured, a filtration is often used to represent the change in the set of events that can be measured through gain or loss of information. In financial mathematics, a filtration represents the information available at each time $t$, and it becomes more and more precise as information from the present becomes available.

The concept of martingale in probability theory is referred to a class of betting strategies. It says that, the conditional expected value of the next observation, given all the past observations is equal to the last observation. It is important to note that the property of being a martingale involves both the filtration and the probability measure. It is possible that one stochastic process could be a martingale with respect to one probability measure but not with respect to another one.

In this chapter, we first discuss the conditional expectation and then proceed to study filtrations and martingales in detail.

## 8.2 Conditional Expectation $E(X/Y)$ and Calculation Rules



Let $(\Omega, \mathcal{F}, P)$ be the given probability space and $X$ be the given random variable. Then its expectation $E(X)$ is defined as

$$E(X) = \int_{\Omega} X(w) dP(w) \ , \tag{8.1}$$

where the integration is to be understood in the Lebesgue-Stieltjes sense. In case $X$ is a discrete random variable taking values $\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots$ we have

$$E(X) = \sum_i \alpha_i P(X = \alpha_i) \ ,$$

where $P(X = \alpha_i) = P(w \in \Omega : X(w) = \alpha_i)$. If $X$ is a continuous random variable having probability density function (p.d.f.) $f(x)$ then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \ . \tag{8.2}$$

We say that the expectation $E(X)$ exists provided $E(X) < \infty$. Obviously not all random variables have expectation. For example, if the discrete random variable $X$ takes value $2^n,\ n = 1, 2, \ldots$ with probability mass function (p.m.f.)

$$P(X = 2^n) = 2^{-n},\ (n = 1, 2, \ldots)\ ,$$

then $E(X)$ does not exist. In a similar manner, if the continuous random variable $X$ has the p.d.f.

$$f(x) = \frac{1}{\pi(1 + x^2)}\ , \quad -\infty < x < \infty\ ,$$

then $E(X)$ does not exist.

Now, we define the conditional expectation and present some of its properties.

**Definition 8.2.1 (Conditional Expectation of a Random Variable)** *For a discrete random variable $X$ and a discrete random variable $Y$, the conditional expectation of $X$ given the event $Y = y$ is defined as*

$$E(X/Y = y) = \sum_x x\ P(X = x/Y = y)\ .$$

**Remark 8.2.1** *Looking at the expression of $E(X/Y = y)$, it is simple to note that it is a function of $y$. If we write this function of $y$ as $f(y) = E(X/Y = y)$, then we write $E(X/Y)$ for $f(Y)$. This is a discrete random variable. It is called the conditional expectation of $X$ given $Y$ and is denoted by $E(X/Y)$.*

**Remark 8.2.2** *Definition 8.2.1 of conditional expectation can be extended in an obvious manner even when $X$ is continuous random variable. We note that $Y$ is still a discrete random variable. Therefore if we define*

$$E(X/Y)(w) = E(X/A_i) \quad \text{for } w \in A_i \ \ (i = 1, 2, \ldots)\ ,$$

*and $A_i = \{w \in \Omega : Y = y_i\}$, then it takes care of both discrete as well as continuous cases of the random variable $X$. Here again $E(X/Y)$ is a discrete random variable.*

**Remark 8.2.3** *Definition 8.2.1 of conditional expectation can further be extended when the event $Y = y$ is replaced by a general event $B$ for which $P(B) > 0$. Thus when $P(B) > 0$, the conditional expectation of the random variable $X$ given $B$ is defined as*

$$E(X/B) = \frac{E(X1_B)}{P(B)}\ ,$$

*where*

$$1_B(w) = \begin{cases} 1, & w \in B \\ 0, & w \notin B \end{cases},$$

*denotes the indicator function of the set $B$.*

**Example 8.2.1** *Let $X$ denote the outcome of tossing of an unbiased die. Let $Y$ be a random variable which takes the value $+1$ if the outcome is an even number and it takes the value $-1$ if the outcome is an odd number.*

*(i) Find $E(X/Y = 1)$ and $E(X/Y = -1)$.*
*(ii) Let $B = \{1, 2, 3\}$. Find $E(X/B)$.*

**Solution** Here $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $P(X = i) = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$.

(i) We have $A_1 = \{w \in \Omega : Y(w) = 1\} = \{2, 4, 6\}$ and $A_2 = \{w \in \Omega : Y(w) = -1\} = \{1, 3, 5\}$. Hence

$$E(X/Y = 1) = E(X/A_1) = \frac{E(X1_{A_1})}{P(A_1)} .$$

But

$$E(X1_{A_1}) = 2 \times \frac{1}{6} + 4 \times \frac{1}{6} + 6 \times \frac{1}{6} ,$$

and

$$P(A_1) = P(\{2, 4, 6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} .$$

Hence,

$$E(X/Y = 1) = \frac{E(X1_{A_1})}{P(A_1)} = 4 .$$

Similarly,

$$E(X/Y = -1) = \frac{E(X1_{A_2})}{P(A_2)} = 3 .$$

(ii) We have

$$E(X/B) = \frac{E(X1_B)}{P(B)} = 2 .$$

□

**Properties of Conditional Expectation**

Using the above definition of conditional expectation we present the following properties for $E(X/Y)$. We do not prove these results here and shall refer to Mikosch [97] for the same.

(i) For random variables $X_1, X_2$ and reals $a_1$, $a_2$, we have

$$E\left((a_1X_1 + a_2X_2)/Y\right) = a_1E(X_1/Y) + a_2E(X_2/Y) \ .$$

This property is called the *linear property* of $E(X/Y)$.

(ii) $E(E(X/Y)) = E(X)$. This property is called the *average property* of $E(X/Y)$.

(iii) For Borel measurable function $g$, we have $E(Xg(Y)/Y) = g(Y)E(X/Y)$. This property is called the *out property*.

(iv) $E(E(X/Y, Z)/Z) = E(X/Z)$. This property is called the *tower property*.

(v) If $X$ and $Y$ are independent random variables then $E(X/Y) = E(X)$.

**Remark 8.2.4** *Conditional expectation tells us how to use the observation of the random variable $Y$ to estimate another random variable $X$. This conditional expectation is the best guess about $X$ given $Y$ if we want to minimize the mean square error.*

**Example 8.2.2** *Consider two i.i.d random variables $X$ and $Y$ each having uniform distribution between the intervals $0$ and $1$. Define $Z = X + Y$. Find*

*(i)* $E(Z/X)$.
*(ii)* $E(XZ/X)$.



**Solution** Since $X$ and $Y$ are uniformly distributed random variables, we have $E(X) = E(Y) = \frac{1}{2}$.

(i) We have

$$E(Z/X) = E[(X + Y)/X] = E(X/X) + E(Y/X) = X + E(Y) = X + \frac{1}{2} \ .$$

(ii) Using property (iii) of conditional expectation, we have

$$E(XZ/X) = XE(Z/X) = X\left(X + \frac{1}{2}\right) \ .$$

□

**Example 8.2.3** *Consider a 2-period binomial model. Let $S(n)$ be the stock price at period $n$. Assume that, $P(up) = p$ and $P(down) = 1 - p$ such that $0 < p < 1$. Find $E\left(S(2)/S(1)\right)$.*

**Solution** The stock price $S(1)$ is a discrete random variable taking two values $uS(0)$ and $dS(0)$ with probabilities $p$ and $1 - p$ respectively. This can be formed as partition $\mathcal{B} = \{B_1, B_2\}$ with $B_1 = \{S(1) = uS(0)\}$ and $B_2 = \{S(1) = dS(0)\}$. Then the forecast of $S(2)$ given the information $B_1$ is given by

$$E(S(2)/S(1) = uS(0)) = pS(0)u^2 + (1-p)S(0)ud\ .$$

In a similar manner the forecast of $S(2)$ given the information $B_2$ is given by

$$E\left(S(2)/S(1) = dS(0)\right) = pS(0)ud + (1-p)S(0)d^2\ .$$

Note that, this forecast differs depending on whether $B_1$ or $B_2$ actually occurs in period $n = 1$. □

**Example 8.2.4** *Consider Example 7.3.1. Find the conditional expectation of* $W(s)$ *given* $W(t) = x$ *for* $0 < s < t$.

**Solution** We know that the conditional distribution of $W(s)$ given $W(t) = x$ for $0 < s < t$ is a normal distribution with mean $\frac{sx}{t}$ and variance $s^2\left(\frac{1}{s} - \frac{1}{t}\right)$. Hence,

$$E\left(W(s)/W(t) = x\right) = \frac{sx}{t}\ .$$

□



## 8.3 More on $\sigma$-Fields

In the previous section we have defined the conditional expectation $E(X/Y)$ where $Y$ is a discrete random variable. We have observed that $E(X/Y)$ is a function of $Y$ which is again a discrete random variable. The definition required the construction of sets $A_i = \{w \in \Omega : Y(w) = y_i\}$ and then define

$$E(X/Y)(w) = E(X/A_i) \quad \text{for } w \in A_i \ \ (i = 1, 2, \ldots)\,.$$

Thus $E(X/Y)$ can be understood as a discrete random variable constructed from a collection of subsets $A_i$ of $\Omega$.

In Chapter 7, we have studied $\sigma$-fields and also learnt how to generate a $\sigma$-field from a given collection of subsets of $\Omega$. Let us recall the following definitions.

**Definition 8.3.1 ($\sigma$-Field)** *A $\sigma$-field on $\Omega$ is a collection $\mathcal{F}$ of subsets of $\Omega$ satisfying*

*(i)* $\emptyset \in \mathcal{F}$,
*(ii) if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$,
*(iii) if a countable collection of sets* $A_1, A_2, \ldots$ *is in* $\mathcal{F}$, *then* $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

**Definition 8.3.2** (**$\sigma$-Field Generated by a Collection of Subsets of $\Omega$**) *Let $\mathcal{U}$ be a collection of subsets of $\Omega$. Then the smallest $\sigma$-field containing $\mathcal{U}$ is called the $\sigma$-field generated by the collection $\mathcal{U}$ of subsets of $\Omega$ and it is denoted by $\sigma(\mathcal{U})$.*

Based on Definition 8.3.2, in this section we wish to introduce *$\sigma$-field generated by a random variable*, *$\sigma$-field generated by a random vector* and *$\sigma$-field generated by a stochastic process*. This is important because we need the $\sigma$-fields generated by stochastic processes to define filtration and martingale.

**Definition 8.3.3** (**$\sigma$-Field Generated by a Random Variable**) *Let $Y$ be a discrete random variable taking values $y_i$, $(i = 1, 2, \ldots)$. Let $A_i = \{w \in \Omega : Y(w) = y_i\}$ $(i = 1, 2, \ldots)$, and $\mathcal{U} = \{A_1, A_2, \ldots\}$ be the corresponding family of subsets of $\Omega$. Then the $\sigma$-field generated by the random variable $Y$ is defined as the $\sigma$-field generated by the family $\mathcal{U}$. In that case we denote $\sigma(Y) = \sigma(\mathcal{U})$.*

**Remark 8.3.1** *The $\sigma$-field $\sigma(Y)$ certainly consists of all sets of form $A = \cup_{i\in I} A_i$, where $I$ is any subsets of $\mathbf{N} = \{1, 2, \ldots\}$ including $I = \emptyset$. For $I = \emptyset$, we get $A = \emptyset$ and for $I = \mathbf{N}$ we get $A = \Omega$. Also $\sigma(Y)$ consists of all subsets of the form*

$$A_{a,b} = \{Y \in (a, b]\} = \{w \in \Omega : a < Y(w) \le b\}, -\infty < a < b < \infty \ ,$$

*because*

$$A_{a,b} = \cup_{i\in I}\{w : Y(w) = y_i\} \in \sigma(Y) \ ,$$

*where $I = \{i : a < y_i \le b\}$.*



**Remark 8.3.2** *Recall, if $\Omega = \mathbf{R}$ and $\mathcal{U}^{(1)} = \{(a, b] : -\infty < a < b < \infty\}$, then the $\sigma$-field generated by $\mathcal{U}^{(1)}$ is called the Borel $\sigma$-field, and its elements are called Borel sets. Thus the Borel $\sigma$-field is the $\sigma$-field generated by the family of semi-open intervals of type $(a, b]$, $-\infty < a < b < \infty$.*

*In case $\Omega = \mathbf{R}^n$, we can consider the family of rectangles*

$$\mathcal{U}^{(n)} = \{((a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_n, b_n]), -\infty < a_i < b_i < \infty, (i = 1, 2, \ldots, n)\} \ ,$$

*and the $\sigma$-field generated by $\mathcal{U}^{(n)}$, i.e. $\sigma(\mathcal{U}^{(n)})$. This $\sigma$-field is called the $n$-dimensional Borel field. Thus $n$-dimensional Borel field is the $\sigma$-field generated by the collection of rectangles in $\mathbf{R}^n$.*

**Example 8.3.1** *Consider a binomial model with $t = 1, 2$. Let $S_t$ be the stock price at time $t$. Let $\Omega = \{(u, u), (u, d), (d, u), (d, d)\}$. Define a discrete random variable $X$ as*

$$X(w) = \begin{cases} 0.5, & w \in \{(u, u), (u, d)\} \\ 1.5, & w \in \{(d, u), (d, d)\} \ . \end{cases}$$

*Find $\sigma(X)$.*

**Solution** The random variable $X$ is a discrete random variable on $\mathcal{F}_1$. We have

$$A_1 = \{w \in \Omega : X(w) = 0.5\} = \{(u,u),(u,d)\} ,$$

and

$$A_2 = \{w \in \Omega : X(w) = 1.5\} = \{(d,u),(d,d)\} .$$

Hence to determine the required $\sigma$-field we need to consider the family $\mathcal{U} = \{A_1, A_2\}$ of subsets of $\Omega$. Thus

$$\mathcal{U} = \{\{(u,u),(u,d)\},\{(d,u),(d,d)\}\} .$$

Then the $\sigma$-field generated by the random variable $X$ is defined as the $\sigma$-field generated by the family $\mathcal{U}$. In this case,

$$\begin{aligned}\sigma(X) &= \sigma(\mathcal{U}) \\ &= \{\emptyset,\{(u,u),(u,d)\},\{(d,u),(d,d)\},\Omega\} .\end{aligned}$$

□

**Definition 8.3.4 ($\sigma$-Field Generated by a Random Vector)**
*Let $Y = (Y_1,\ldots,Y_n)$ be an $n$-dimensional random vector. Then the $\sigma$-field generated by the random vector $Y = (Y_1,Y_2,\ldots,Y_n)$ is the smallest $\sigma$-field containing all $n$-dimensional Borel sets, i.e. sets of the form*

$$\{w \in \Omega : a_i < Y_i(w) \le b_i \ (i = 1,2,\ldots,n), -\infty < a_i < b_i < \infty \ (i = 1,2,\ldots,n)\} .$$

*This $\sigma$-field is denoted by $\sigma(Y) = \sigma(Y_1,Y_2,\ldots,Y_n)$.*

**Example 8.3.2** *Consider a 2-period binomial model. Let $S_n$ be the stock price at period $n$. Let $\Omega = \{(u,u),(u,d),(d,u),(d,d)\}$. Define a discrete random vector $X = (X_1,X_2)$ where*

$$X_1(w) = \begin{cases} 0.5, & w \in \{(u,u),(u,d)\} \\ 1.5, & w \in \{(d,u),(d,d)\} ,\end{cases}$$

*and*

$$X_2(w) = \begin{cases} 0.25, & w = (u,u) \\ 0.5, & w = (u,d) \\ 0.75, & w = (d,u) \\ 1.00, & w = (d,d) .\end{cases}$$

*Find the $\sigma$-field generated by the random vector $X = (X_1,X_2)$.*

**Solution** We have

$$A_1 = \{w \in \Omega : X_1(w) = 0.5\} = \{(u, u), (u, d)\} ,$$

and

$$A_2 = \{w \in \Omega : X_1(w) = 1.5\} = \{(d, u), (d, d)\} .$$

Similarly, we have

$$\begin{aligned} B_1 &= \{w \in \Omega : X_2(w) = 0.25\} = \{(u, u)\} \\ B_2 &= \{w \in \Omega : X_2(w) = 0.5\} = \{(u, d)\} \\ B_3 &= \{w \in \Omega : X(w) = 0.75\} = \{(d, u)\} , \end{aligned}$$

and

$$B_4 = \{w \in \Omega : X_2(w) = 1.00\} = \{(d, d)\} .$$

Hence $\sigma$-field generated by $X = (X_1, X_2)$ is the $\sigma$-field generated by the family $\mathcal{U}$ where

$$\mathcal{U} = \{A_1, A_2, B_1, B_2, B_3, B_4\} .$$

Therefore the desired $\sigma$-field is

$$\begin{aligned} \sigma(X) = \sigma(X_1, X_2) = \{&\emptyset, A_1, A_2, B_1, B_2, B_3, B_4, B_1 \cup B_3, B_1 \cup B_4, B_2 \cup B_3, B_2 \cup B_4, \\ &B_1 \cup B_2 \cup B_3, B_1 \cup B_2 \cup B_4, B_2 \cup B_3 \cup B_4, B_1 \cup B_3 \cup B_4, \Omega\} . \end{aligned}$$

□

**Definition 8.3.5** **($\sigma$-Field Generated by a Stochastic Process)** *Let $\{Y(t), t \in T\}$ be the given stochastic process. Then the $\sigma$-field generated by the stochastic process $\{Y(t), t \in T\}$ is the smallest $\sigma$-field containing all sets of the form*

$$\{w : \text{ the sample path } (Y(t, w), t \in T) \text{ belongs to } C\}$$

*for all suitable sets $C$ of functions on $T$.*

**Remark 8.3.3** *The set $C$ in Definition 8.3.5 probably needs some more specifications. But at this stage and with our limited background, it is not possible to answer this question for a general stochastic process. However we shall present the $\sigma$-field generated by a Brownian motion, as this is the one which we shall be using most often.*

**Definition 8.3.6 ($\sigma$-Field Generated by a Brownian Motion)** *Let* $W = \{W(s), 0 < s \leq t\}$ *be the given Brownian motion on* $[0, t]$*. Then the* $\sigma$*-field generated by all sets of the form*

$$A_{t_1,t_2,\ldots,t_n} = \{w \in \Omega : (W(t_1, w), W(t_2, w), \ldots, W(t_n, w)) \in C\} \ ,$$

*for any* $n$*-dimensional Borel set* $C$*, and for any choice of* $t_i \in [0, t]$*,* $i \geq 1$ *is called the* $\sigma$*-field generated by the Brownian motion* $W$*.*

**Remark 8.3.4** *For a random variable, a random vector or a stochastic process* $Y$ *on* $\Omega$*, the* $\sigma$*-field* $\sigma(Y)$ *generated by* $Y$ *contains all the essential information about the structure of* $Y$ *as a function of* $w \in \Omega$*. It consists of all subsets* $\{w : Y(w) \in C\}$ *for all suitable sets* $C$*. In general,* $C$ *has to be any* $n$*-dimensional Borel set* $(n \geq 1)$*,* $n$ *being equal to 1 for the case of a random variable. In this situation we agree to the terminology that* $Y$ *contains information represented by* $\sigma(Y)$ *or* $Y$ *carries the information* $\sigma(Y)$*.*

## 8.4 Filtration

Filtration is important in financial mathematics, because it allows to model the flow of information. Of course, the information increases as time goes by. Before we give the formal definition of a filtration, we consider below the given situation.

Let us consider the random experiment of tossing a coin three times. Here the set of all possible outcomes is

$$\Omega = \{(HHH, (HHT), (HTH), (HTT), (THH), (THT), (TTH), (TTT)\} \ .$$

We choose $\Omega$ as the universal or global set as it contains all information on the process. Let $E_H$ be consists of all possible outcomes starting with head in 1st toss and $E_T$ consists of all possible outcomes starting with tail in 1st toss. Now, we consider the family $\mathcal{F}_1 = \{\emptyset, \{E_H\}, \{E_T\}, \Omega\}$ of subsets of $\Omega$. Then it can be verified that $\mathcal{F}_1$ is a $\sigma$-field. Note that, $\mathcal{F}_1$ is the set of events that have been decided by the end of 1st toss. Similarly, let $E_{HH}$ consists of all possible outcomes starting with head in 1st toss and head in 2nd toss. We define $E_{HT}$ , $E_{TH}$ and $E_{TT}$ analogously. Then, we consider the following family $\mathcal{F}_2 = \{\emptyset, \{E_H\}, \{E_T\}, \{E_{HH}\}, \{E_{HT}\}, \{E_{TH}\}, \{E_{TT}\}, \{E^c_{HH}\}, \{E^c_{HT}\}, \{E^c_{TH}\}, \{E^c_{TT}\}, \{E_{HH} \cup E_{TH}\}, \{E_{HH} \cup E_{TT}\}, \{E_{HT} \cup E_{TH}\}, \{E_{HT} \cup E_{TT}\}, \Omega\}$ of subsets of $\Omega$. We can again verify that $\mathcal{F}_2$ is a $\sigma$-field. Observe that, this $\sigma$-field $\mathcal{F}_2$ contains the information learned by observing the first two tosses of a coin.

Now, after the coin is tossed three times, every subset of $\Omega$ is resolved. Hence the family $\mathcal{F}_3$ of 256 subsets of $\Omega$ which is the set of all subsets of $\Omega$. That is, $\mathcal{F}_3$ is the total $\sigma$-field. Since $\mathcal{F}_i$ is the set of events that have been decided by the end of $i$th toss, we have more information at the end of $(i+1)^{th}$ toss than at the end of $i^{th}$ toss.

Taking motivation from the above example we give the definition of filtration in discrete time.

**Definition 8.4.1 (Filtration in a Discrete Time)** *Let $\Omega$ be the set of all possible outcomes of a random experiment and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then a filtration in discrete time is an increasing sequence $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots$ of $\sigma$-fields, one per time instant.*

The $\sigma$-field $\mathcal{F}_n$ may be thought of as the events of which the occurrence is determined at or before time $n$, the "known events" at time $n$. The interpretation of $\mathcal{F}_0 = \{\emptyset, \Omega\}$ is that, at the beginning, one has no information. Further the interpretation of $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ is that the state of information increases over time.

**Example 8.4.1** *Consider $\Omega = \{a, b, c, d\}$. Construct $\sigma$-fields $\mathcal{F}_i$, $(i = 0, 1, 2)$, such that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$.*

**Solution** Obviously $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Let $\mathcal{F}_1 = \{\emptyset, \{a, b\}, \{c, d\}, \Omega\}$ and $\mathcal{F}_2 = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{c, d\}, \{a, d\}, \{a, c\}, \{b, c\}, \{b, d\}, \{a, b, c\}, \{c, d, a\}, \{d, a, b\}, \{b, c, d\}, \Omega\}$. Then, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$.

□

**Definition 8.4.2 (Filtration in a Continuous Time)** *Let $\Omega$ be the set of all possible outcomes of a random experiment. Let $T$ be a fixed positive number and assume that for each $t \in [0, T]$, there is a $\sigma$-field $\mathcal{F}_t$. Assume further that, if $s \leq t$, then every set in $\mathcal{F}_s$ is also in $\mathcal{F}_t$. Then, the collection of $\sigma$-fields $\{\mathcal{F}_t,\ 0 \leq t \leq T\}$ is called a filtration in continuous time.*

Thus a collection of $\sigma$-fields $\{\mathcal{F}_t,\ t \geq 0\}$ is called a filtration in continuous time if $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \leq s \leq t$.

**Remark 8.4.1** *Filtration is used to model the flow of information over time. As an example, we think of $X_t$ as the price of some asset at time $t$ and $\mathcal{F}_t$ as the information obtained by watching all the prices in the market up to time $t$.*

**Definition 8.4.3 (Natural Filtration)** *The natural filtration of a discrete time stochastic process $\{X_0, X_1, \ldots\}$ is defined by the filtration $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ where $\mathcal{F}_n$ is the $\sigma$-field generated by the random vector $(X_0, X_1, \ldots, X_n)$.*

Thus $\mathcal{F}_n$ contains all events that depend on the first $(n+1)$ elements of the stochastic process. It gives the "history" of the process up till time $n$. A convenient notation to describe a $\sigma$-field corresponding to observing a random vector $X$ is $\sigma(X)$. Thus $\sigma(X)$, the $\sigma$-field generated by X, consists of all events that can be expressed in $X$: events of the type $\{X \in C\}$, $C$ being $(n+1)$ dimensional Borel sets. In this notation, the natural filtration of a discrete time stochastic process $\{X_0, X_1, \ldots\}$ can be written as $\mathcal{F}_n = \{(X_0, X_1, \ldots, X_n)\}$.

The natural filtration of a continuous time stochastic process $\{X(t), t \geq 0\}$ is defined by the filtration $\{\mathcal{F}_t, t \geq 0\}$ where $\mathcal{F}_t = \sigma(X(s), s \leq t)$.

**Definition 8.4.4 (Adapted Process)** *We say that a discrete time stochastic process* $\{X_0, X_1, \ldots\}$ *is adapted to a given filtration* $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ *if the* $\sigma$*-field generated by* $X_n$ *is a subset of* $\mathcal{F}_n$ *means* $\sigma(X_n) \subset \mathcal{F}_n$*, for every* $n$*. In a similar manner, a continuous time stochastic process* $\{X(t), t \geq 0\}$ *is said to be adapted to a given filtration* $\mathcal{F}_t, t \geq 0\}$ *if* $\sigma(X(t)) \subset \mathcal{F}_t$ *for all* $t \geq 0$.

Thus the events connected to an adapted process up to time $n$ are known at time $n$. For instance, suppose $S_n$ is the price of a stock at the end of $n$th day then the price process $\{S_n, n = 0, 1, 2, \ldots\}$ is adapted to natural filtration $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\}$ where $\mathcal{F}_n$ is the history up to the end of $n$th day, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\Omega = [0, \infty)$. In a similar manner for continuous time stochastic process $\{S(t), t \geq 0\}$ is adapted to $\{\mathcal{F}_t, t \geq 0\}$ where $\mathcal{F}_t$ is the history up to time $t$.

**Remark 8.4.2** *The natural filtration corresponding to a process is the smallest filtration to which it is adapted. If the process* $\{Y_0, Y_1, \ldots\}$ *is adapted to the natural filtration of a stochastic process* $\{X_0, X_1, \ldots\}$ *then for each* $n$ *the variable* $Y_n$ *is a function* $\sigma(X_0, X_1, \ldots, X_n)$ *of the sample path of the process* $X$ *up till time* $n$.

**Example 8.4.2** *Consider Example 8.4.1. Define*

$$X_0(w) = \{1, \ w \in \{a, b, c, d\}\},$$
$$X_1(w) = \begin{cases} 1, & w \in \{a, b\} \\ -1, & w \in \{c, d\}, \end{cases}$$

*and*

$$X_2(w) = \begin{cases} 1, & w = a \\ 2, & w = b \\ 3, & w = c \\ 4, & w = d. \end{cases}$$

*Verify that* $\{X_i, i = 0, 1, 2\}$ *is an adapted process to* $\mathcal{F}_i$, $(i = 0, 1, 2)$.

**Solution** From Example 8.4.1, we have $\mathcal{F}_i$, $(i = 0, 1, 2)$ are the natural filtration of the stochastic process $\{X_i, i = 0, 1, 2\}$. The $\sigma$-field generated by the random variable $X_0$ is defined as the $\sigma$-field generated by the family $\{a, b, c, d\}$.

$$\sigma(X_0) = \{\emptyset, \Omega\} \subseteq \mathcal{F}_0 \ .$$

Similarly, the $\sigma$-field generated by the random variable $X_1$ is defined as the $\sigma$-field generated by the family $\{\{a, b\}, \{c, d\}\}$.

$$\sigma(X_1) = \{\emptyset, \{a, b\}, \{c, d\}, \Omega\} \subseteq \mathcal{F}_1 \ .$$

Finally, the $\sigma$-field generated by the random variable $X_2$ is defined as the $\sigma$-field generated by the family $\{\{a\}, \{b\}, \{c\}, \{d\}\}$.

$$\begin{aligned}\sigma(X_2) = \{&\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{c, d\}, \{a, d\}, \{a, c\}, \{b, c\}, \{b, d\},\\ &\{a, b, c\}, \{c, d, a\}, \{d, a, b\}, \{b, c, d\}, \Omega\}\\ \subseteq\ &\mathcal{F}_2 \ .\end{aligned}$$

Hence, $\{X_i, (i = 0, 1, 2)\}$ is an adapted process to $\mathcal{F}_i$ $(i = 0, 1, 2)$.

□

**Remark 8.4.3** *If a stochastic process $\{X(t), t \geq 0\}$ is adapted to the natural filtration $\{\mathcal{F}_t, t \geq 0\}$ of the Brownian motion $\{W(t), t \geq 0\}$, then it means that $X(t)$ is a function of $W(s)$, $s \leq t$. For example, $X(t) = W^2(t) - t$ and $Y(t) = \sup_{s \leq t} W(s)$ are adapted processes whereas $Z(t) = \sup_{s \leq t+1} W(s)$ and $U(t) = W(1) - W(t)$ are not adapted processes to the natural filtration $\{\mathcal{F}_t, t \geq 0\}$ of the Brownian motion $\{W(t), t \geq 0\}$.*

**Remark 8.4.4** *Many stochastic calculus texts specify a probability space as $(\Omega, \{\mathcal{F}_t, t \geq 0\}, P)$ thereby, specifying the filtration explicitly from the very beginning.*

**Remark 8.4.5** *The natural filtration is the name given to the filtration for which $\mathcal{F}_t$ consists of those sets that can be decided by observing trajectories of the specified process up to time $t$.*

## 8.5 Conditional Expectation $E(X/\mathcal{F})$ and Calculation Rules

In Section 8.2, we discussed the conditional expectation $E(X/Y)$ while in Section 8.3, we defined the $\sigma$-field generated by the random variable $Y$. In this section, we study the conditional expectation $E(X/\sigma(Y))$.

**Definition 8.5.1 (Conditional Expectation of a Random Variable given $\sigma$-field)** *Let $X$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a $\sigma$-field of subsets of $\Omega$. A random variable $Z$ is called the conditional expectation of $X$ given the $\sigma$-field $\mathcal{G}$ if*

*(i) $Z$ does not contain more information than that contained in $\mathcal{G}$, i.e. $\sigma(Z) \subset \mathcal{G}$.*
*(ii) For every $A \in \mathcal{G}$, $Z$ satisfies $E(X1_A) = E(Z1_A)$ where $1_A$ is the indicator function of the event $A$.*

*We denote $Z$ by $E(X/\mathcal{G})$ and call the same as the conditional expectation of $X$ given $\mathcal{G}$.*

**Remark 8.5.1** *Thus we may have $Z$ and $Z'$ such that $Z = E(X/\mathcal{G})$ and also $Z' = E(X/\mathcal{G})$, but then we have $Z = Z'$ almost surely. In that sense there is a unique $\mathcal{G}$-measurable random variable $Z$ which satisfies $E(X1_A) = E(Z1_A)$ for every $A \in \mathcal{G}$.*

**Remark 8.5.2** *The interpretation of (i) in the above definition is that, the information in $\mathcal{G}$ is sufficient to determine $Z$. Then, we say that $Z$ is $\mathcal{G}$-measurable.*

**Remark 8.5.3** *Obviously, every random variable is $\mathcal{F}$-measurable where $\mathcal{F}$ is the largest or total $\sigma$-field on $\Omega$. Also, if $X$ is $\mathcal{G}$-measurable, then $h(X)$ is also $\mathcal{G}$-measurable for any Borel-measurable function $h$. The other extreme is when a random variable is independent of a $\sigma$-field. In this case, the information contained in the $\sigma$-field gives no clue about the random variable.*

**Remark 8.5.4** *When $\mathcal{G} = \sigma(Y)$, the conditional expectation $E(X/Y)$ is a special case of the above definition of $E(X/\mathcal{G})$. We know that, every element $A$ of $\sigma(Y)$ is of the form*

$$A = \cup_{i\in I} A_i = \cup_{i\in I}\{w \in \Omega : Y(w) = y_i\},\ I \subset \mathbf{N}\ .$$

*Hence,*

$$E(X1_A) = E\left(X \sum_{i\in I} 1_{A_i}\right) = \sum_{i\in I} E(X1_{A_i})\ .$$

*For the case of discrete random variable $Z1_A$, its expectation is given by*

$$E(Z1_A) = \sum_{i\in I} E(X/A_i)P(A_i) = \sum_{i\in I} E(X1_{A_i})\ .$$

*From Definition 8.5.1, we conclude that when $\mathcal{G}$ is the $\sigma$-field generated by $Y$, then we write $E(X/\mathcal{G})$ for the random variable $E(X/Y)$. Thus $E(X/\mathcal{G})$ is the expected value of $X$ given the information $\mathcal{G}$. This result is valid for the continuous random variable also.*

**Example 8.5.1** *Consider a binomial model with* $t = 1, 2$*. Let* $S_t$ *be the stock price at time* $t$*. Let* $\Omega = \{(u,u),(u,d),(d,u),(d,d)\}$*. Let* $\sigma$*-field* $\mathcal{F}$ *be the power set of* $\Omega$ *and* $P(w) = \frac{1}{4}$ *for all* $w \in \Omega$*. Let* $\mathcal{G} = \{\emptyset, \{(u,u),(u,d)\}, \{(d,u),(d,d)\}, \Omega\}$*. Define a discrete random variable* $X$ *as*

$$X(w) = \begin{cases} 0.25, & w = (u,u) \\ 0.5, & w = (u,d) \\ 0.75, & w = (d,u) \\ 1.00, & w = (d,d) \ . \end{cases}$$

*Find* $E(X/\mathcal{G})$*.*

**Solution** Define

$$Z(w) = \begin{cases} \frac{3}{8}, & w \in \{(u,u),(u,d)\} \\ \frac{7}{8}, & w \in \{(d,u),(d,d)\} \ . \end{cases}$$

Then we have

$$Z^{-1}\{(-\infty, z]\} = \begin{cases} \emptyset, & -\infty < z < \dfrac{3}{8} \\ \{(u,u),(u,d)\}, & \dfrac{3}{8} \leq z < \dfrac{7}{8} \\ \Omega, & \dfrac{7}{8} \leq z < \infty \ . \end{cases}$$

Hence, $Z$ is $\mathcal{G}$-measurable.



For $A_1 = \{(u,u),(u,d)\}$, we have

$$E(X1_{A_1}) = 0.25 \times \frac{1}{4} + 0.50 \times \frac{1}{4} = \frac{3}{16} \ ,$$

and for $A_2 = \{(d,u),(d,d)\}$, we have

$$E(X1_{A_2}) = 0.75 \times \frac{1}{4} + 1.00 \times \frac{1}{4} = \frac{7}{16} \ .$$

Now, for $A_1 = \{(u,u),(u,d)\}$, we have

$$E(Z1_{A_1}) = \frac{3}{8} \times \left(\frac{1}{4} + \frac{1}{4}\right) = \frac{3}{16} \ ,$$

and for $A_2 = \{(d,u),(d,d)\}$, we have

$$E(Z1_{A_2}) = \frac{7}{8} \times \left(\frac{1}{4} + \frac{1}{4}\right) = \frac{7}{16} \ .$$

By Definition 8.5.1, the random variable $Z$ is the conditional expectation of $X$ given the $\sigma$-field $\mathcal{G}$.

□

**Result 8.5.1** *Let $X$ and $Y$ be two integrable random variables and $a$ and $b$ be two real numbers, and $\mathcal{G}$ be the sub-$\sigma$-field of $\mathcal{F}$. Then,*

*(i)* $E((aX + bY)/\mathcal{G}) = aE(X/\mathcal{G}) + bE(Y/\mathcal{G})$,

*(ii) If $X$ and $\mathcal{G}$ are independent, then $E(X/\mathcal{G}) = E(X)$. It means that we do not gain any information about $X$, if we know $\mathcal{G}$ and vice versa.*

*(iii) If $\sigma$-field $\sigma(X) \subset \mathcal{G}$, then $E(X/\mathcal{G}) = X$. This means that the information contained in $\mathcal{G}$ provides us with the whole information about the random variable $X$. Hence, $X$ can be treated as a constant.*

*(iv) If $X$ is $\mathcal{G}$-measurable, then $E(XY/\mathcal{G}) = XE(Y/\mathcal{G})$. Given $\mathcal{G}$, we can deal with $X$ as if it is a constant, hence we can pull $X(w)$ out of the updated expectation and write it in front of $E(Y/\mathcal{G})$.*

*(v) If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two sub $\sigma$-fields of $\mathcal{F}$ with $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1) = E(X/\mathcal{G}_1)$.*

For the proofs of above results we shall refer to Mikosch [97].

**Example 8.5.2** *Suppose $\mathcal{G} = \{\emptyset, \Omega\}$. Find $E(X/\mathcal{G})$.*

**Solution** The given $\sigma$-field $\{\emptyset, \Omega\}$ is the trivial $\sigma$-field containing no information. The only random variables which are measurable with respect to the trivial $\sigma$-field are constants. Hence, $E(X/\mathcal{G}) = E(X) = c$, where $c$ is a constant.

□

**Example 8.5.3** *Consider a discrete random variable $X$ which takes value $-\sqrt{b}$ and $\sqrt{b}$ with probabilities $\frac{1}{3}$ and $\frac{2}{3}$ respectively. Find $E(X/\sigma(X^2))$. Here $\sigma(X^2)$ is the $\sigma$-field generated by the random variable $X^2$.*

**Solution** We have

$$P(X = \sqrt{b}/X^2 = b) = \frac{2}{3}\ ; \quad \text{and} \quad P(X = -\sqrt{b}/X^2 = b) = \frac{1}{3}\ .$$

Hence,

$$E(X/\sigma(X^2)) = \frac{2\sqrt{b}}{3} - \frac{\sqrt{b}}{3} = \frac{\sqrt{b}}{3}\ .$$

□

## 8.6 Martingales

The theory of martingale plays a very important and useful role in the study of financial mathematics. A formal definition is given below.

**Definition 8.6.1 (Discrete Time Martingale)** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{X_n, n = 0, 1, \ldots\}$ be a stochastic process and $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ be the filtration. The stochastic process $\{X_n, n = 0, 1, \ldots\}$ is said to be a martingale corresponding to the filtration $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ if it satisfies the following conditions*

*(i) For every $n$, $E(X_n)$ exists.*
*(ii) Each $X_n$ is $\mathcal{F}_n$-measurable.*
*(iii) For every $n$, $E(X_{n+1}/\mathcal{F}_n) = X_n$.*

**Remark 8.6.1** *The definition of martingale depends on the collection of $\sigma$-fields $\mathcal{F}_n$. For clarity, one can say that $(X_n, \mathcal{F}_n)$ is a martingale.*

**Remark 8.6.2** *From the definition of martingale and using the properties of conditional expectation, we observe that if $\{X_n\}$ is a martingale then $E(X_{n+1}) = E(X_n)$ for every $n$. This implies that $E(X_n) = c$, a constant. Therefore if, for some $n > 0$, $E(X_n) < \infty$ and the increments $X_{n+1} - X_n$ of the martingale $\{X_n\}$ are bounded, then $E(X_n) = E(X_0)$.*

**Result 8.6.1** *We can generate martingale sequences by the following procedure. Given any increasing family of $\sigma$-fields $\{\mathcal{F}_n\}$, and any integrable random variable $X$ on $(\Omega, \mathcal{F}, P)$, we take $X_n = E(X/\mathcal{F}_n)$ and it is easy to check that $\{(X_n, \mathcal{F}_n)\}$ is a martingale sequence. Of course, every finite martingale sequence is generated this way for we can always take $X$ to be $X_n$, the last one.*

**Result 8.6.2** *In equation (8.6.1), if '$=$' is replaced by '$\geq$' then $X_n : n = 0, 1, 2, \ldots$ is called a* submartingale *while if it is replaced by '$\leq$' then $X_n : n = 0, 1, 2, \ldots$ is called a* supermartingale. *Obviously, $\{X_n\}$ is a* supermartingale *if and only if $\{-X_n\}$ is a submartingale.*

**Result 8.6.3** *Let $\{X_n, n = 0, 1, \ldots\}$, $\{Y_n, n = 0, 1, \ldots\}$ be discrete time stochastic processes. We say $\{X_n\}$ is a martingale with respect to $\{Y_n\}$ if $E(|X_n|) < \infty$ and $E(X_{n+1}/Y_0, Y_1, \ldots, Y_n) = X_n$. We may think of $Y_0, Y_1, \ldots, Y_n$ as the information or history upto stage $(n+1)$. It may include more information than just $X_0, X_1, \ldots, X_n$.*

*By calculating the conditional expectation value of $X_{n+1}$ given the information about $X_n$ (or $Y_n$) upto time $n$, we are making a forecast for the random variable. The martingale relation implies that the "best" forecast for the next value of the random variable is its current value.*

**Result 8.6.4** *Martingale theory provides a classification scheme for the time series. If a time series exhibits no discernible trend then it has a martingale like behavior. On the other hand, if the trend is an increasing (decreasing) one, then the time series behaves like submartingale (supermartingale).*

**Example 8.6.1** *Consider the gambler's ruin problem. A gambler starts with Rs* $N$*. Then a coin is flipped, landing head with probability* $p$ *and landing tail with probability* $1-p$*. If the coin lands heads, he/she gains one rupee, otherwise he /she loses a rupee. The game continues until he/she loses all his/her* $N$ *rupees. Let* $X_n$ *be fortune at the* $n$*th game, and* $S_n$ *be his/her capital after* $n$ *games. What will be his/her fortune, on an average, on the next game given that his/her current fortune?*

**Solution** For this we suppose that $S_n$ is adapted to the filtration $\mathcal{F}_n$. Then we can interpret a martingale as a fair game. Thinking $(S_n - S_r)$ as the net winnings of the game per unit stake in time frame $(r, n]$, the best prediction of the net winnings given the information at the time $r < n$ has value

$$E\left((S_n - S_r)/\mathcal{F}_r\right) = E(S_n/\mathcal{F}_r) - S_r \ .$$

If $S_n$ is a martingale then $E(S_n/\mathcal{F}_r) - S_r = 0$. This means that the best prediction of the future net winnings per unit stake in the interval $(r, n]$ is zero. This is exactly what we expect to be a fair game. It says that gambler expected capital after one more game played with the knowledge of the entire past and present is exactly equal to his/her current capital.

□

**Example 8.6.2** *Let* $X_1, X_2, \ldots$ *be a sequence of i.i.d random variables each taking two values +1 and -1 with equal probabilities. Let us define* $S_0 = 0$ *and* $S_n = \sum_{j=1}^{n} X_j$*,* $(n = 1, 2, \ldots)$*. This discrete time stochastic process* $\{S_n, n = 0, 1, \ldots\}$ *is a symmetric random walk. Prove that,* $\{S_n, n = 0, 1, \ldots\}$ *is a martingale with respect to* $\{X_n, n = 1, 2, \ldots\}$*.*

**Solution** We have $E(|S_n|) \leq E(|X_1|) + E(|X_2|) + \cdots + E(|X_n|) < \infty$ . Also

$$\begin{aligned} E(S_{n+1}/X_1, X_2, \ldots, X_n) &= E\left((S_n + X_{n+1})/X_1, X_2, \ldots, X_n\right) \\ &= E(S_n/X_1, X_2, \ldots, X_n) + E(X_{n+1}/X_1, X_2, \ldots, X_n) \\ &= S_n + E(X_{n+1}) \ \ \text{(using independent of } X_1, X_2, \ldots, X_n, X_{n+1}) \\ &= S_n + 0 \\ &= S_n \ . \end{aligned}$$

Hence $\{S_n, n = 0, 1, \ldots\}$ is a martingale with respect to $\{X_n, n = 1, 2, \ldots\}$. Suppose $\mathcal{F}_k$ is the $\sigma$-field of information corresponding to the first $k$ random variables $X_k$, we have for non-negative integers $k < n$, $E(S_n/\mathcal{F}_k) = S_k$ .

□

**Example 8.6.3** *Consider a symmetric random walk* $\{S_n, n = 0, 1, \ldots\}$ *which is a martingale with respect to the filtration* $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ *where* $\mathcal{F}_0 = \{\emptyset, \Omega\}$ *and* $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n), (n \geq 1)$, *is the* $\sigma$*-field of information corresponding to the first* $n$ *random variables* $X_n$. *Verify if* $\{S_n^2, n = 0, 1, \ldots\}$ *is a martingale with respect to the filtration* $\{\mathcal{F}_n, n = 0, 1, \ldots\}$.

**Solution** For each $n = 1, 2, \ldots$, $S_n^2$ is $\mathcal{F}_n$-measurable. Also

$$E(S_n^2) = \sum_{i=1}^{n} E(X_i^2) < \infty \ .$$

Now,

$$\begin{aligned} E(S_{n+1}^2/\mathcal{F}_n) &= E\left[(S_{n+1} - S_n + S_n)^2/\mathcal{F}_n\right] \\ &= E\left[(S_{n+1} - S_n)^2/\mathcal{F}_n\right] - 2E\left[S_n(S_{n+1} - S_n)/\mathcal{F}_n\right] + E(S_n^2/\mathcal{F}_n) \\ &= E(X_{n+1}^2/\mathcal{F}_n) - 2E(X_{n+1}\dot{S}_n/\mathcal{F}_n) + E(S_n^2/\mathcal{F}_n). \end{aligned}$$

Since $X_{n+1}$ is independent of $\mathcal{F}_n$ and since $S_n^2$ is $\mathcal{F}_n$-measurable, we have

$$\begin{aligned} E(S_{n+1}^2/\mathcal{F}_n) &= E(X_{n+1}^2) - 2S_n E(X_{n+1}) + S_n^2 \\ &= 1 - 0 + S_n^2 \\ &= 1 + S_n^2 \ . \end{aligned}$$

Hence, $E(S_{n+1}^2/\mathcal{F}_n) \neq S_n^2$ . Therefore, $\{S_n^2, n = 0, 1, \ldots\}$ is not a martingale. Since $E(S_{n+1}^2/\mathcal{F}_n) \geq S_n^2$, $(n = 1, 2, \ldots)$, hence $\{S_n^2, n = 0, 1, \ldots\}$ is a submartingale.

□

**Example 8.6.4** *Let* $Y_0 = 0$ *and let* $Y_1, Y_2, \ldots$ *be a sequence of i.i.d random variables with* $E(|Y_k|) = 0$ *and* $E(Y_k^2) = \sigma^2$. *Let* $X_0 = 0$ *and* $X_n = \left(\sum_{k=1}^{n} Y_k\right)^2 - n\sigma^2$. *Show that* $\{X_n, n = 0, 1, \ldots\}$ *is a martingale with respect to* $\{Y_n, n = 0, 1, \ldots\}$ .

**Solution** We have

$$X_{n+1} = \left(\sum_{k=1}^{n+1} Y_k\right)^2 - (n+1)\sigma^2$$
$$= \left(\left(\sum_{k=1}^{n} Y_k\right) + Y_{n+1}\right)^2 - n\sigma^2 - \sigma^2$$
$$= X_n + 2Y_{n+1}\left(\sum_{k=1}^{n} Y_k\right) + Y_{n+1}^2 - \sigma^2 \; .$$

$$E\left[X_{n+1}/Y_0, Y_1, \ldots, Y_n\right] = E\left[(X_n + 2Y_{n+1}\left(\sum_{k=1}^{n} Y_k\right) + Y_{n+1}^2 - \sigma^2)/(Y_0, Y_1, \ldots, Y_n)\right]$$
$$= E\left[X_n/(Y_0, Y_1, \ldots, Y_n)\right]$$
$$+2E\left[Y_{n+1}\sum_{k=1}^{n} Y_k/(Y_0, Y_1, \ldots, Y_n)\right]$$
$$+E\left[Y_{n+1}^2/(Y_0, Y_1, \ldots, Y_n)\right] - \sigma^2$$
$$= X_n + 2\sum_{k=1}^{n} Y_k E(Y_{n+1}) + E(Y_{n+1}^2) - \sigma^2$$
$$= X_n \; .$$



Thus $\{X_n, n = 0, 1, \ldots\}$ is a martingale with respect to $\{Y_n, n = 0, 1, \ldots\}$ .

□

**Example 8.6.5** *Let a person start with* Rs 1. *A fair coin is tossed infinitely many times. For* $n$*th toss, if it turns up 'head', the person gets* Rs 2, *but if turn up 'tail', the person does not get any amount. Let* $Y_n$ *be his/her fortune at the end of* $n$*th toss. Prove that* $Y_n$ *is a martingale.*

**Solution** Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables each defined by

$$X_i = \begin{cases} 2, & \text{with probability } 0.5 \\ 0, & \text{with probability } 0.5 \; . \end{cases}$$

Since the game is double or nothing, his/her fortune at the end of $n$th toss is given by

$$Y_n = X_1 X_2 \cdots X_n \; (n = 1, 2, \ldots) \; .$$

Let $\mathcal{F}_n$ be the $\sigma$-field generated by $X_1, X_2, \ldots, X_n$. We note that $0 \le Y_n \le 2^n$ and $E[X_{n+1}] = 1$. Now,

$$\begin{aligned} E\left[Y_{n+1}/\mathcal{F}_n\right] &= E\left[Y_n X_{n+1}/\mathcal{F}_n\right] \\ &= Y_n E\left[X_{n+1}/\mathcal{F}_n\right] \\ &= Y_n E[X_{n+1}] \\ &= Y_n \ . \end{aligned}$$

Hence, $\{Y_n, n = 1, 2, \ldots\}$ is a martingale.

□

**Example 8.6.6** *Consider a binomial lattice model. Let $S_n$ be the stock price at period $n$ and*

$$S_{n+1} = \begin{cases} uS_n, & \text{with probability } p \\ dS_n, & \text{with probability } 1-p \ . \end{cases}$$

*Define a related process $R_n$ as*

$$R_n = ln(S_n) - n\ \left[p\ ln(u) + (1-p)\ ln(d)\right] \ .$$

*Prove that $\{ln(S_n), n = 1, 2, \ldots\}$ is not a martingale whereas $\{R_n, n = 1, 2, \ldots\}$ is a martingale with respect to $\{S_n, n = 1, 2, \ldots\}$. Also, prove that the discounted stock process $\{S_0, e^{-r}S_1, e^{-2r}S_2, \ldots\}$ is a martingale only if*

$$p = \frac{e^r - d}{u - d} \ ,$$

*where $r$ is the nominal interest rate.*



**Solution** In this binomial lattice model $\{S_0, S_1, \ldots\}$ with the natural filtration $\{\mathcal{F}_0, \mathcal{F}_1, \ldots\}$, we have

$$P(S_{n+1} = uS_n/\mathcal{F}_n) = 1 - P(S_{n+1} = dS_n/\mathcal{F}_n) = p \ .$$

Hence,

$$E(S_{n+1}/\mathcal{F}_n) = p\ uS_n + (1-p)\ dS_n = S_n[p\ u + (1-p)\ d] \ .$$

We consider the variable $ln(S_n)$ and observe that

$$E\left(ln\left(\frac{S_n}{S_{n-1}}\right)/S_{n-1}, S_{n-2}, \ldots, S_0\right) = p\ ln(u) + (1-p)\ ln(d) \ .$$

Therefore,

$$E\left(ln(S_n)/S_{n-1}, S_{n-2}, \ldots, S_0\right) = ln(S_{n-1}) + p\ ln(u) + (1-p)\ ln(d)\ . \tag{8.3}$$

Here $\{ln(S_n), n = 1, 2, \ldots\}$ is not a martingale and depending upon the values of $p, u$ and $d$ it may be either a submartingale or a supermartingale. Next, we consider the process $R_n$.

$$E(R_n/R_{n-1}, R_{n-2}, \ldots, R_0) = E\left(ln(S_n) - n\left[p\ ln(u) + (1-p)\ ln(d)\right]/R_{n-1}, R_{n-2}, \ldots, R_0\right)$$

Using equation (8.3), and noting that the history of $S_{n-1}, S_{n-2}, \ldots, S_0$ yields the history of $R_{n-1}, R_{n-2}, \ldots, R_0$ and vice-versa, we get

$$\begin{aligned} E(R_n/R_{n-1}, R_{n-2}, \ldots, R_0) &= ln(S_{n-1}) - (n-1)\left[p\ ln(u) + (1-p)\ ln(d)\right] \\ &= R_{n-1}\ . \end{aligned}$$

Therefore $\{R_n, n = 1, 2, \ldots\}$ is martingale.

Now, consider the discounted process $\{S_0, e^{-r}S_1, e^{-2r}S_2, \ldots\}$ where $r$ is the interest rate. We have

$$E(e^{-(n+1)r}S_{n+1}/\mathcal{F}_n) = p\ ue^{-(n+1)r}S_n + (1-p)\ de^{-(n+1)r}S_n\ .$$

The discounted process is a martingale only if the right hand side of the above equation is equal to $e^{-nr}S_n$. That is,

$$e^{-nr}S_n = p\ ue^{-(n+1)r}S_n + (1-p)\ de^{-(n+1)r}S_n$$

or

$$e^r = p\ u + (1-p)\ d\ .$$

Thus, the discounted process is a martingale only if

$$p = \frac{e^r - d}{u - d}.$$

□

**Remark 8.6.3** *It may be noted that in Chapter 3 the RNPM for the binomial lattice model has been derived via replicating portfolio arguments. But the approach presented here in Example 8.6.6 is based on the fact that discounted stock process under RNPM form a martingale. This approach is more general and is applicable in a large variety of derivative pricing problems.*

Now, one of the important results of martingale theory is Doob's decomposition, which connects the martingale and submartingale. This result is stated below. We shall refer to Roman [113] for its proof.

**Lemma 8.6.1 (Doob's Decomposition)** *Let $\{X_n, n = 0, 1, \ldots\}$ be a submartingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \ldots\}$. Then there exists a martingale $M = \{M_n, n = 0, 1, \ldots\}$ with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ and a stochastic process $A = \{A_n, n = 0, 1, \ldots\}$ such that:*

*(i)* $X_n = M_n + A_n,\ (n = 0, 1, \ldots)$ .
*(ii) A is an increasing stochastic process, i.e.* $A_n \leq A_{n+1},\ (n = 0, 1, \ldots)$ .
*(iii)* $A_n$ *is* $\mathcal{F}_{n-1}$*- measurable, for all* $n$ .

**Remark 8.6.4** *In Example 8.6.6,* $ln(S_n)$ *is not a martingale whereas* $R_n$ *is a martingale. This is simply an illustration (and not a consequence) of Lemma 8.6.1.*

**Example 8.6.7** *Prove that, under risk neutral probability measure Q the discounted stock price process* $\{(1 + r)^{-k} S_k\ ,\ k = 1, 2, \ldots\}$ *is a martingale.*

**Solution** This is a simple interest version of the second part of Example 8.6.1. We are proving it separately even at the cost of repetition. Let $\Omega$ be all sequences of length $n$ of heads $H$ and tails $T$. Let $S_0$ be a fixed number. Define $S_k(w) = u^j d^{k-j} S_0$ if the first $k$ elements of a given $w \in \Omega$ has $j$ occurrences of $H$ and $k - j$ occurrences of $T$. It means that if $H$ occurs then the stock price goes up by a factor $u$; if $T$ occurs then it goes down by a factor $d$. Let $\mathcal{F}_k$ be the $\sigma$-field generated by $S_0, S_1, \ldots, S_k$. Let $\overline{p} = \dfrac{(1 + r) - d}{u - d}$, $\overline{q} = \dfrac{u - (1 + r)}{u - d}$ and define $Q(w) = (\overline{p})^j (\overline{q})^{n-j}$ if $w$ has $j$ appearances of $H$ and $n - j$ appearances of $T$. We note that under probability measure $Q$, the random variables $S_{k+1}/S_k = Y_k$ (say) are independent and identically distributed with probability mass function

$$P(Y = u) = \overline{p}, \quad P(Y = d) = \overline{q}\ .$$

Since $Y_k$ i.i.d random variables, they satisfy

$$P(Y_1 = y_1, , \ldots, Y_n = y_n) = P(Y_1 = y_1) \ldots P(Y_n = y_n) = (\overline{p})^j (\overline{q})^{n-j}\ .$$

Also we may note that $Y_k$ is the factor by which the stock price goes up or down at time $k$. Since the random variable $Y_k$ is independent of $\mathcal{F}_k$, we have

$$E_Q\left[(1 + r)^{-(k+1)} S_{k+1} / \mathcal{F}_k\right] = (1 + r)^{-k} S_k (1 + r)^{-1} E_Q\left[(S_{k+1}/S_k) / \mathcal{F}_k\right]\ .$$

But

$$E_Q\left[(S_{k+1}/S_k)/\mathcal{F}_k\right] = E_Q\left[(S_{k+1}/S_k)\right] = \overline{p}\ u + \overline{q}\ d = (1+r)\ .$$

Therefore on substitution, we conclude

$$E_Q\left[(1+r)^{-(k+1)}S_{k+1}/\mathcal{F}_k\right] = (1+r)^{-k}S_k\ .$$

This proves that the process $\{(1+r)^{-k}S_k, k = 1, 2, \ldots\}$ is a martingale.

□

**Wealth Process**

Let $\Delta_k$ be the number of shares of a stock held between time $k$ and $k+1$. We assume that $\Delta_k$ is $\mathcal{F}_k$-measurable and $X_0$ is the amount of money we have started with time $t = 0$. If we have $\Delta_k$ shares between time $k$ and $k+1$, then at time $k+1$ those shares will be worth $\Delta_k S_{k+1}$, where $S_{k+1}$ is the share price at time $k+1$. The amount of cash we hold between time $k$ and $k+1$ is $X_k$ minus the amount held in stock, that is $X_k - \Delta_k S_k$. Hence, the worth of this amount at time $k+1$ is $(1+r)[X_k - \Delta_k S_k]$. Therefore, the amount of money we have at time $k+1$ is

$$X_{k+1} = \Delta_k S_{k+1} + (1+r)\left[X_k - \Delta_k S_k\right]\ .$$

When $r = 0$, this reduces to

$$X_{k+1} - X_k = \Delta_k(S_{k+1} - S_k)\ .$$



Thus,

$$X_{k+1} = X_0 + \sum_{i=0}^{k} \Delta_i(S_{i+1} - S_i)\ .$$

The stochastic process $\{X_k, k = 0, 1, \ldots\}$ is called the *wealth process*.

We shall now show that under risk neutral probability measure $Q$ the discounted wealth process is a martingale.

$$\begin{aligned} E_Q\left[X_{k+1} - X_k/\mathcal{F}_k\right] &= E_Q\left[\Delta_k(S_{k+1} - S_k)/\mathcal{F}_k\right] \\ &= \Delta_k E_Q\left[(S_{k+1} - S_k)/\mathcal{F}_k\right] \quad (\Delta_k \text{ is } \mathcal{F}_k - \text{ measurable}) \\ &= 0 \quad (S_k \text{ is a martingale})\ . \end{aligned}$$

Now writing $X_{k+1}$ as $X_k + \Delta_k(S_{k+1} - S_k)$ and noting that $r > 0$, we have

$$\begin{aligned} E_Q\left[(1+r)^{-(k+1)}X_{k+1} - (1+r)^{-k}X_k)/\mathcal{F}_k\right] &= E_Q\left[\Delta_k\left[(1+r)^{-(k+1)}S_{k+1} - (1+r)^{-k}S_k\right]/\mathcal{F}_k\right] \\ &= \Delta_k E_Q\left[(1+r)^{-(k+1)}S_{k+1} - (1+r)^{-k}S_k/\mathcal{F}_k\right] \\ &= 0 \quad (S_k \text{ is a martingale under } Q)\ . \end{aligned}$$

Hence, discounted wealth process $\{(1+r)^{-k}X_k, k = 1, 2, \ldots\}$ is a martingale.

**Definition 8.6.2 (Continuous Time Martingale)** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{X(t), t \geq 0\}$ be a stochastic process and $\{\mathcal{F}_t, t \geq 0\}$ be a filtration. The stochastic process $\{X(t), t \geq 0\}$ is said to be a martingale corresponding to the filtration $\{\mathcal{F}_t, t \geq 0\}$ if it satisfies the following conditions*

*(i) For every $t$, $E(X(t))$ exists.*
*(ii) Each $X(t)$ is $\mathcal{F}_t$-measurable.*
*(iii) For every $0 < s < t$,*

$$E(X(t)/\mathcal{F}_s) = X(s) \ . \tag{8.4}$$

**Example 8.6.8** *Prove that $\{W(t), t \geq 0\}$ is a martingale.*

**Solution** For $0 < s < t$,

$$\begin{aligned} E(W(t)/\mathcal{F}_s) &= E(W(t) - W(s) + W(s)/\mathcal{F}_s) \\ &= E(W(t) - W(s)/\mathcal{F}_s) + E(W(s)/\mathcal{F}_s) \\ &= 0 + W(s) \ \ \text{(from the property of Brownian motion)} \ . \end{aligned}$$

Therefore, $\{W(t), t \geq 0\}$ is a martingale.

□

**Example 8.6.9** *Prove that the Poisson process $\{N(t), t \geq 0\}$ is not a martingale.*

**Solution** First $E(|\, N(t) \,|) \leq \lambda t < \infty$ . For $0 < s < t$,

$$\begin{aligned} E(N(t)/\mathcal{F}_s) &= E\left((N(t) - N(s) + N(s))/\mathcal{F}_s\right) \\ &= E\left((N(t) - N(s))/\mathcal{F}_s\right) + E(N(s)/\mathcal{F}_s) \\ &= \lambda(t - s) + N(s) \ \ \text{(from the property of Poisson process)} \ . \end{aligned}$$

Hence, $E(N(t)/\mathcal{F}_s) \neq N(s)$, $\{N(t), t \geq 0\}$ is not a martingale. Note that, $\{N(t), t \geq 0\}$ is a submartingale since $E(N(t)/\mathcal{F}_s) \geq N(s)$, for $0 < s < t$.

□

**Example 8.6.10** *Show that $exp\left(W(t) - \dfrac{t}{2}\right)$ is a martingale.*

**Solution** Let $0 \leq s < t$. Since $W(t) - W(s)$ is independent of $\mathcal{F}_s$ and $W(s)$ is $\mathcal{F}_s$-measurable, we have

$$\begin{aligned} E\left(e^{W(t)}/\mathcal{F}_s\right) &= E\left(e^{W(t)-W(s)}e^{W(s)}/\mathcal{F}_s\right) \\ &= e^{W(s)}E\left(e^{W(t)-W(s)}/\mathcal{F}_s\right) \\ &= e^{W(s)}E\left(e^{W(t)-W(s)}\right) \ . \end{aligned}$$

Since $W(t)-W(s)$ has normal distribution with mean zero and variance $(t-s)$, we have

$$E\left(e^{W(t)-W(s)}\right)=e^{\frac{t-s}{2}}.$$

Hence,

$$E\left(e^{W(t)}/\mathcal{F}_s\right)=e^{W(s)}e^{\frac{t-s}{2}}.$$

This gives, for $0 \le s < t$,

$$E\left(e^{W(t)-\frac{t}{2}}/\mathcal{F}_s\right)=e^{\frac{-t}{2}}E\left(e^{W(t)}/\mathcal{F}_s\right)=e^{W(s)-\frac{s}{2}}.$$

It follows that $\exp\left(W(t)-\frac{t}{2}\right)$ is a martingale.

□

The cornerstone of martingale theory is Doob's optional sampling theorem. This states, roughly, that "stopping" a martingale at a random time $\tau$ does not alter the expected "payoff", provided the decision about when to stop is based solely on information available up to $\tau$. Such random times are called stopping times. *Stopping times* are also called as *Markov times* or *optional times.*

If $\tau$ is optional and $c>0$ is a positive constant, then $\tau+c$ is a stopping time. Before stating Doob's optional sampling theorem we give below the definitions.

**Definition 8.6.3 (Stopping Time)** *A stopping time relative to a filtration* $\{\mathcal{F}_n, n \ge 0\}$ *is a non-negative integer-valued random variable* $\tau$ *such that for each* $n$ *the event* $\tau = n \in \mathcal{F}_n$. *In continuous case, a random variable* $\gamma : \Omega \to [a,b]$ *is called a stopping time with respect to a filtration* $\{\mathcal{F}_t, a \le t \le b\}$ *if* $\{w, \gamma(w) \le t\} \in \mathcal{F}_t$ *for all* $t \in [a,b]$.

From the definition, we can think of $\gamma$ as the time to stop playing a game. The decision to stop playing the game before or at time $t$ should be determined by the information provided by $\mathcal{F}_t$ which will give the condition for $\gamma$ to a stopping time.

**Definition 8.6.4 (Hitting Time)** *Let* $\{X(t), t \ge 0\}$ *be a stochastic process. Let* $A$ *be Borel set in* $\mathbf{R}^n$. *Define*

$$\tau = inf\{t>0 : X(t) \in A\}.$$

*Then* $\tau$ *is called a hitting time of* $A$ *for the stochastic process* $\{X(t),\ t \ge 0\}$.

**Example 8.6.11** *Consider a symmetric random walk* $\{S_n, n = 0, 1, \ldots\}$*. Let* $T$ *be the first time that* $S_n = 2$*. Prove that* $T$ *is a stopping time.*

**Solution** Since $T$ is the first time that $S_n = 2$, we have

$$T = \begin{cases} \text{Min}\{n \geq 0 : S_n = 2\}, & \text{if } S_n = 2 \text{ for some } n \in N \\ \infty, & \text{otherwise} \end{cases}$$

In other words,

$$T = inf\{n \geq 0 : S_n = 2\} .$$

Further

$$\{T \leq n\} = \bigcup_{i=0}^{n} \{S_i = 2\} \in \mathcal{F}_n$$

where $\{S_i = 2\} \in \mathcal{F}_i \subseteq \mathcal{F}_n$ for $i \leq n$ because $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ is a filtration. Hence it follows that $T$ is a stopping time.

□

**Result 8.6.5** *A stopping time* $T$ *is said to be bounded if there exists a constant* $K < \infty$ *such that* $P(T \leq K) = 1$*, and is said to be finite a.s. if* $P(T < \infty) = 1$*. If* $T$ *is a bounded stopping time and* $\{X(t), t \geq 0\}$ *is a martingale, then* $E(X(T)) = E(X(0))$*. For instance, in the above example , T is a finite a.s. but not a bounded stopping time. Hence* $E(S(T)) = 2 \neq 0 = E(S(0))$ *.*



**Example 8.6.12** *Let* $\tau$ *be a stopping time. Prove that* $W(t+\tau) - W(\tau)$ *is a Brownian motion.*

**Solution:** By observing the Brownian motion, we can determine whether or not $\tau \leq t$. For a fixed $a$, the hitting time of $a$ is defined by

$$\tau = inf\{t > 0 : W(t) = a\} .$$

Since the sample paths of Brownian motion are continuous, it is easy to see that $\tau$ is a stopping time. That is,

$$\{\tau \leq t\} = \{W(s) = a \;\text{ for some } s,\; 0 \leq s \leq t\} ,$$

which depends only on $\{W(s), 0 \leq s \leq t\}$. Also, note that, if $\tau < \infty$, then $W(\tau) = a$ by continuity of sample paths. Since $\tau$ is a stopping time, $W(t + \tau) - W(\tau)$ is a Brownian motion.

□

The following theorem is an important result in martingale theory. We shall refer to Roman [113] for the proof of the above theorem.

**Theorem 8.6.1 (Doob's Optional Sampling Theorem)** *Let* $\{X_n, n = 0, 1, \dots\}$ *be a martingale with respect to* $\{Y_n, n = 0, 1, \dots\}$ *and let* $\tau$ *be a stopping time. Then for any* $n = 1, 2, \dots$

$$E(X_{\tau \bigwedge n}/Y_0, Y_1, \dots Y_{n-1}) = X_{\tau \bigwedge n-1}$$

*where* $\tau \bigwedge n = Min(\tau, n)$*. i.e. the stopped process is also a martingale.*

We shall refer to Roman [113] for the proof of the above Theorem. This theorem is also known as optional stopping theorem.

**Remark 8.6.5** *In case* $\{X_n, n = 0, 1, \dots\}$ *is a submartingale with respect to* $\{Y_n, n = 0, 1, \dots\}$ *then*

$$E(X_{\tau \bigwedge n}/Y_0, Y_1, \dots Y_{n-1}) \geq X_{\tau \bigwedge n-1} \ .$$

**Remark 8.6.6** *In case* $\{X_n, n = 0, 1, \dots\}$ *is a supermartingale with respect to* $\{Y_n, n = 0, 1, \dots\}$ *then*

$$E(X_{\tau \bigwedge n}/Y_0, Y_1, \dots Y_{n-1}) \leq X_{\tau \bigwedge n-1} \ .$$

**Example 8.6.13** *Suppose that two players A and B start respectively with Rs a and Rs b, and they bet against each other one rupee at a time by tossing a fair coin. What is the probability that player B will be broke and A ends up with all the money?*

**Solution** Let $X_n$ be player $A$ fortune at the $n$th game, and $S_n$ be his/her capital after $n$ games. We have

$$S_n = a + \sum_{i=1}^{n} X_n \ ,$$

where $X_i$ are i.i.d random variables with

$$P(X_n = \pm 1) = \frac{1}{2} \ .$$

Let

$$\tau_0 = inf\{n : S_n = 0\}$$
$$\tau_{a+b} = inf\{n : S_n = a + b\} ,$$

and $T = \tau_0 \bigwedge \tau_{a+b}$ . Therefore, $T$ is the first time that player $A$ either makes $b$ extra rupees or goes ruin i.e.

$$T = inf\{n : S_n = 0 \text{ or } S_n = b + a\} .$$

Hence, $T$ is the stopping time with respect to the martingale $\{S_n, n = 0, 1, \ldots\}$. By using optional stopping theorem, we have

$$E(S_T) = S_0 = a .$$

But

$$E(S_T) = 0 \times P(T = \tau_0) + (a + b) \times P(T = \tau_{a+b}) .$$

Hence,

$$P\,(T = \tau_{a+b}) = \frac{a}{a+b} .$$

This is exactly the probability that player $B$ goes broke and player $A$ wins all.

□



## 8.7 Summary and Additional Notes

- Conditional expectation plays a key role in derivative pricing models. In Section 8.2, the definition of conditional expectation and its properties are presented.
- In Section 8.3, $\sigma$-fields generated by a random variable, a random vector and a stochastic process are defined. Finally, $\sigma$-field generated by Brownian motion is presented.
- Filtration is presented in Section 8.4 with simple examples. Both discrete and continuous time filtration are explained with some examples.
- We provide a detailed evaluation procedure of conditional expectation with respect to the filtration. Further, their properties are presented along with some examples.
- We know that a conditional expectation is itself a random variable. The complicated conditional expectations such as conditional expectation of a random variable with respect to the filtration arise in martingale theory and the concept of martingale theory is used in a betting strategy. Kolmogorov (1932) invented the martingale theory whereas its first systematic treatment was provided by Doob (1942).

- Two important results in martingale theory namely Doob's decomposition and Doob's optional sampling theorem are presented. Optional sampling concept reveals that the expectation of a martingale remains constant. It has immediate implications concerning the pathwise behaviour of martingales, submartingales and supermartingales. We may refer to Roman [113] for some more details on martingale theory.

## 8.8 Exercises

**Exercise 8.1** *Let* $\{X_n, n \geq 1\}$ *be a sequence of independent and identically distributed random variables each having uniform distribution with values -2, -1, 0, 1, 2, 3. Let* $\{N(t), t \geq 0\}$ *be a Poisson process with parameter 2. Find the mean and variance of* $\sum_{i=1}^{N(t)} X_i$.

**Exercise 8.2** *Consider two i.i.d random variables* $X$ *and* $Y$ *each having uniform distribution between the intervals* $0$ *and* $1$. *Define* $Z = X + Y$. *Prove that* $E(X/Z) = \dfrac{Z}{2}$.

**Exercise 8.3** *Consider the successive rolling of an unbiased die. Let* $X$ *and* $Y$ *denote the number of rolls necessary to obtain a two and a three respectively. Obtain* (*i*) $E(X/Y = 2)$, (*ii*) $E(X/Y = 4)$.

**Exercise 8.4** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $X$ *be an integrable random variable and let* $\mathcal{G} \subset \mathcal{F}$ *be a* $\sigma$*-field. Prove that, the conditional expectation* $E(X/\mathcal{G})$ *exists. (Hint: Use Radon-Nikodym theorem)*

**Exercise 8.5** *Consider* $\Omega = \{a, b, c, d\}$. *Construct 4 distinct* $\sigma$*-fields* $\mathcal{F}_1, \mathcal{F}_3, \mathcal{F}_3, \mathcal{F}_4$ *such that* $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \mathcal{F}_4$.

**Exercise 8.6** *Construct an example of* $\sigma$*-fields* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *such that* $\mathcal{F}_1 \nsubseteq \mathcal{F}_2$ *and* $\mathcal{F}_2 \subset \mathcal{F}_1$.

**Exercise 8.7** *Consider a binomial model with* $t = 1, 2, 3$. *Let* $S_t$ *be the stock price at time* $t$. *Let* $\Omega = \{(u,u,u), (u,u,d), (u,d,u), (u,d,d), (d,u,d), (d,u,u), (d,d,u), (d,d,d)\}$. *Let* $\sigma$*-field* $\mathcal{F}$ *be the power set of* $\Omega$ *and* $P(w) = \dfrac{1}{8}$ *for all* $w \in \Omega$. *Let* $\mathcal{G} = \{\emptyset, \{(u,u,u), (u,u,d), (u,d,u), (u,d,d)\}, \{(d,u,d), (d,u,u), (d,d,u), (d,d,d)\}\Omega\}$. *Define a discrete random variable* $X$ *as*

$$X(w_i) = i, \quad i = 1, 2, \ldots, 8 \ .$$

*Find* $E(X/\mathcal{G})$.

**Exercise 8.8** *Let* $\{X_n, n = 0, 1, \ldots\}$ *and* $\{Y_n, n = 0, 1, \ldots\}$ *be stochastic processes. We say* $\{X_n\}$ *is a martingale with respect to* $\{Y_n\}$ *if*

*(i)* $E(| X_n |) < \infty$.
*(ii)* $E(X_{n+1}/Y_0, Y_1, \ldots, Y_n) = X_n$.

*Prove that* $\{X_n, n = 0, 1, \ldots\}$ *is a martingale with respect to* $\{Y_n, n = 0, 1, \ldots\}$ *where* $X_n = Y_1 + Y_2 + \ldots + Y_n$, $n \geq 1$, $Y_0 = 0$, $\{Y_i, i = 1, 2, \ldots\}$ *are independent random variables with* $E(Y_n) = 0$.

**Exercise 8.9** *Let* $\{S_n, n = 0, 1, \ldots\}$ *be a symmetric random walk and* $\mathcal{F}_n$ *be a filtration. Show that* $Y_n = (-1)^n cos(\pi S_n)$ *is a martingale with respect to* $\mathcal{F}_n$.

**Exercise 8.10** *Consider the tossing of an unbiased coin three times. Suppose that the toss of a head wins* Rs 1 *while an outcome of a tail loses* Rs 1. *Let* $X_n$ *denote the sum of the winnings at time* $n$. *Prove that* $\{X_n, n = 1, 2, 3\}$ *is an adapted process with respect to* $\{\mathcal{F}_i, i = 1, 2, 3\}$ *defined in Section 7.4. Further prove that* $\{X_n, n = 1, 2, 3\}$ *is a martingale with respect to* $\{\mathcal{F}_i, i = 1, 2, 3\}$.

**Exercise 8.11** $\{N(t), t \geq 0\}$ *be a Poisson process with parameter* $\lambda$. *Prove that,* $\{N(t) - \lambda t, t \geq 0\}$ *is a martingale.*

**Exercise 8.12** *Let* $X(t) = \mu t + \sigma W(t), \quad -\infty < \mu < \infty, \quad 0 < \sigma < \infty$. *Prove that* $\{X(t), t \geq 0\}$ *is a martingale for* $\mu = 0$. *Also prove that* $\{X(t), t \geq 0\}$ *is a submartingale for* $\mu > 0$, *and a supermartingale for* $\mu < 0$.

**Exercise 8.13** *Show that the maximum of two submartingales (with respect to the same filtration) is a submartingale.*

**Exercise 8.14** *Let* $Y_1, Y_2, \ldots$ *be a sequence of positive independent random variables with* $E(Y_i) = 1$ *for all* $i$. *Define* $X_0 = 1$, *and* $X_n = \prod_{i=1}^{n} Y_i$ *for* $n \geq 1$. *Show that* $\{X_n, n = 0, 1, \ldots\}$ *is a martingale with respect to its natural filtration.*

**Exercise 8.15** *Let* $Y_1, Y_2, \ldots$ *be a sequence of i.i.d random variables with* $P(Y_i = 1) = p$ *and* $P(Y_i = 0) = 1 - p$. *Define* $X_n = \prod_{i=1}^{n} \frac{Y_i}{p}$. *Prove that,* $E(X_{n+1}/Y_1, Y_2, \ldots, Y_n) = X_n$.

**Exercise 8.16** *Let* $X$ *be a random variable and let* $\phi$ *be a convex function on* $\mathbf{R}$. *Suppose both* $X$ *and* $\phi(X)$ *are integrable, and* $\mathcal{G} \subset \mathcal{F}$ *is a* $\sigma$*-field. Prove that,* $\phi(E(X/\mathcal{G})) \leq E(\phi(X)/\mathcal{G})$.

**Exercise 8.17** *Let $X_1, X_2, \ldots$ be a sequence of square integrable random variables. Show that if $\{X_n, n = 1, 2, \ldots\}$ is a martingale with respect to a filtration $\mathcal{F}_n$, then $X_n^2$ is a sub martingale with respect to the same filtration.*
*(Hint: Use Jensen's inequality with convex function $\phi(x) = x^2$)*

**Exercise 8.18** *Prove that, if $\{X_n, n = 1, 2, \ldots\}$ is a martingale, then $\mid X_n \mid$, $X_n^2$, $e^{X_n}$ and $e^{-X_n}$ are all submartingales.*

**Exercise 8.19** *Prove that, if $X_n > 0$ and is a martingale, then $\sqrt{X_n}$ and $ln(X_n)$ are supermartingales.*

**Exercise 8.20** *Prove that, if $X_n$ is a submartingale and K a constant, then Max $\{X_n, K\}$ is a submartingale, while if $X_n$ is a supermartingale, so is Min $\{X_n, K\}$.*

**Exercise 8.21** *Let $X_1, X_2, \ldots$ be i.i.d random variables each having normal distribution with zero mean and unit variance. Show that the sequence $Y_n = \exp\left((\sum_{i=1}^{n} X_i) - \frac{1}{2}n\right)$ forms a martingale.*

**Exercise 8.22** *Prove that $\{W^2(t) - t, t \geq 0\}$ is a martingale, where $\{W(t), t \geq 0\}$ is a Brownian motion.*

**Exercise 8.23** *Let $\{W(t), t \geq 0\}$ be a Wiener process. Is $\exp(\sigma W(t) - \dfrac{\sigma^2}{2}t)$ a martingale where $\sigma$ is a positive constant?*



**Exercise 8.24** *Let $\{N(t),\ t \geq 0\}$ be a Poisson process with parameter 1. Which of the following are martingales.*

*(i)* $\{N(t) - t,\ t \geq 0\}$.
*(ii)* $\{N(t)^2 - t,\ t \geq 0\}$.
*(iii)* $\{(N(t) - t)^2 - t,\ t \geq 0\}$.

**Exercise 8.25** *Let X be a integrable random variable and $\mathcal{G}_1$ and $\mathcal{G}_2$ be two sub-$\sigma$-fields of $\mathcal{F}$. Prove that, if $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1) = E(X/\mathcal{G}_1)$.*

**Exercise 8.26** *Let $\{\mathcal{F}_n, n = 1, 2, \ldots\}$ be a filtration. Let X be any random variable with $E(\mid X \mid) < \infty$. Define $X_n = E(X/\mathcal{F}_n), n = 1, 2, \ldots$ Prove that, $\{X_n,\ (n = 1, 2, \ldots)\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n,\ n = 1, 2, \ldots\}$.*

**Exercise 8.27** *Let $\{Y_n,\ (n = 0, 1, \ldots)\}$ be an arbitrary sequence of random variables. Suppose X is a random variable with $E(\mid X \mid) < \infty$. Define*

$$X_n = E(X/Y_0, Y_1, \ldots, Y_n), \ \ (n = 0, 1, \ldots)\ .$$

*Show that* $\{X_n, n = 0, 1, \ldots\}$ *forms a martingale with respect to* $\{Y_n, n = 0, 1, \ldots\}$. *We say that,* $\{X_n, n = 0, 1, \ldots\}$ *is a Doob's martingale.*

**Exercise 8.28** *Let* $\{W(t), t \geq 0\}$ *be a Weiner process and let* $\tau$ *be a stopping time. Define*

$$\tilde{W}(t) = \begin{cases} W(t), & t \leq \tau \\ 2W(\tau) - W(t), & t > \tau \ . \end{cases}$$

*Prove that* $\{\tilde{W}(t), t \geq 0\}$ *is a Weiner process.*

**Exercise 8.29** *Let* $\tau$ *is a stopping time for the Weiner process. Using the optimal sampling theorem, show that* $E(W(\tau)) = 0$. *Use this to deduce that* $P(W_{(}\tau) = a) = \dfrac{b}{a+b}$ *and* $P(W(\tau) = b) = \dfrac{a}{a+b}$.

**Exercise 8.30** *For* $a \in \mathbf{R}$, *let* $\tau_a$ *be the first time that* $\{W(t), t \geq 0\}$ *hits* $a$. *Suppose that* $a > 0 > b$. *By considering the stopped martingale* $\{(W(t))^{\tau_a \wedge \tau_b}, t \geq 0\}$, *prove that* $P(\tau_a < \tau_b) = \dfrac{-b}{a-b}$.

# 9
# Stochastic Calculus

## 9.1 Introduction

This chapter attempts to present certain introductory topics of stochastic calculus, also called *Ito's calculus*. But do we really need the apparatus of Ito's calculus in finance? The answer is YES. This is because to make any meaningful financial decision, we need to understand the dynamics of the asset under consideration. For example, to price a stock option, we need to understand the dynamics of the underlying stock. Traditionally calculus in general and differential equations in particular have played a very vital role in the mathematical modeling of any dynamical system. We expect that these topics will again come to our rescue to model the asset dynamics. But since the asset dynamics is mostly stochastic, we need a different type of calculus to take care of functions which are no where differentiable and are not of bounded variation (do not worry! we shall be defining it shortly). Ito's calculus serves our goal very well in this scenario.

When in the 19th century the German mathematician Weierstrass constructed a real-valued function which is continuous everywhere, but differentiable nowhere, (see Section 9.10) this was considered as nothing else but a mathematical curiosity. Interestingly, this "curiosity" is at the core of mathematical finance. High frequency data show that prices of exchange rates, interest rates, and liquid assets are practically continuous, but are of unbounded variation in every given time interval. In particular, they are nowhere differentiable. Therefore classical calculus is required to be extended to functions of unbounded variation, a task overlooked by mathematicians for long. This gap was bridged by the development of stochastic calculus, which can be considered as the theory of differentiation and integration of stochastic processes.

Although the subject of stochastic calculus is very broad and complex, we keep our goal very modest. We present a very brief and introductory discussion on certain topics of stochastic calculus, stochastic differential equations, Ito's integral and Ito's formula. Our discussion here is motivated by the writings of Mikosch [97], Shreve [122], Karatzas and Shreve [75], and Klebaner [78].

## 9.2 Variations of Real-Valued Functions

Before we discuss the variations of Brownian motion, we first discuss the variations of real-valued functions. Here for the sake of simplicity we consider the interval $[0,T]$ rather than a general interval $[a,b]$.

**Definition 9.2.1 (First Variation of a Real-Valued Function)** *Let* $g : [0,T] \to \mathbf{R}$ *and* $\mathbf{\Pi}$ *be the set of all partitions* $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ *of* $[0,T]$. *Then the first variation of* $g$ *over* $[0,T]$ *is defined as*

$$V_g(T) = \sup_{\pi \in \mathbf{\Pi}} \left( \sum_{i=0}^{n-1} \mid g(t_{i+1}) - g(t_i) \mid \right) .$$

Here we are using the symbol $V_g(T)$ rather than $V_g[0,T]$ since left end point 0 is fixed throughout discussion while right end point $T$ can vary.

**Definition 9.2.2 (Function of Bounded Variation)** *Let* $0 < t \leq T$. *Then* $g$ *is said to be of finite variation if* $V_g(t) < \infty$, *for all* $t$. *Further if for* $0 < t \leq T$, $V_g(t) < K$, *a constant independent of* $t$, *then* $g$ *is said to be of bounded variation on* $[0,T]$.

**Definition 9.2.3 ($p$-Variation of a Real-Valued Function)** *Let* $0 < t \leq T$. *Then the* $p$*-variation,* $p > 1$, *of the function* $g$ *in the interval* $[0,t]$ *is defined as*

$$V_g^p(t) = \sup_{\pi \in \mathbf{\Pi}} \left( \sum_{i=0}^{n-1} \mid g(t_{i+1}) - g(t_i) \mid^p \right) . \tag{9.1}$$

**Remark 9.2.1** *(Goffman [51]) In case* $g$ *is a continuous function then* $V_g(T)$ *can alternatively be expressed as*

$$V_g(T) = \lim_{\|\pi\| \to 0} \left( \sum_{i=0}^{n-1} \mid g(t_{i+1}) - g(t_i) \mid \right) ,$$

*where* $\pi$ *is an arbitrary partition* $\{0 = t_0 < t_1 < \ldots < t_n = T\}$ *and* $\| \pi \| = \underset{0 \leq k \leq n-1}{Max} (t_{k+1} - t_k)$.

In a similar manner, the $p$-variation can be expressed as

$$V_g^p(t) = \lim_{\|\pi\|\to 0}\left(\sum_{i=0}^{n-1} | g(t_{i+1}) - g(t_i) |^p\right) . \tag{9.2}$$

In our discussion here we shall mostly be using these alternate expressions of $V_g$ and $V_g^p$ because our functions of interest will be continuous.

**Result 9.2.1** *(Goffman [51]) Let the $p$-variation of the function $g$ in the interval $[0, T]$ be denoted by $V_g^p(T)$. Let $V_g^p(T) < \infty$. Then for $r < p$, $V_g^r(T) = \infty$, and for $q > p$, $V_g^q(T) = 0$.*

**Theorem 9.2.1** *Let $T > 0$ and $g : [0, T] \to \mathbf{R}$ be continuously differentiable function. Then*

*(i)* $V_g(T) = \int_0^T | g'(t) | \, dt < \infty$.
*(ii)* $V_g^2(T) = 0$.

**Proof.**

(i) We have

$$V_g(T) = \lim_{\|\pi\|\to 0}\sum_{i=0}^{n-1}\left|g(t_{i+1}) - g(t_i)\right| .$$



Since $g$ is continuously differentiable function, by applying Mean Value Theorem, we have for each subinterval $[t_i, t_{i+1}]$

$$g(t_{i+1}) - g(t_i) = g'(t_i^*)(t_{i+1} - t_i) ,$$

for some point $t_i^*$ in $(t_i, t_{i+1})$. Hence,

$$\begin{aligned} V_g(T) &= \lim_{\|\pi\|\to 0}\sum_{i=0}^{n-1} | g(t_{i+1}) - g(t_i) | \\ &= \lim_{\|\pi\|\to 0}\sum_{i=0}^{n-1} | g'(t_i^*)(t_{i+1} - t_i) | \\ &= \int_0^T | g'(t) | \, dt < \infty . \end{aligned}$$

The last integral is finite because $g'(t)$ is continuous on $[0, T]$ and therefore $| g'(t) |$ is Riemann integrable.

(ii) Next, the 2-variation (or quadratic variation) of $g$ is given by

$$
\begin{aligned}
V_g^{(2)}(T) &= \lim_{\|\pi\|\to 0} \sum_{i=0}^{n-1} \left| g(t_{i+1}) - g(t_i) \right|^2 \\
&= \lim_{\|\pi\|\to 0} \sum_{i=0}^{n-1} \left| \frac{g(t_{i+1}) - g(t_i)}{t_{i+1} - t_i} \right|^2 (t_{i+1} - t_i)^2 \ .
\end{aligned}
$$

By applying Mean Value Theorem for some $t_i^* \in (t_i, t_{i+1}),\ i = 0, 1, \ldots, n-1$, we have

$$
\begin{aligned}
V_g^{(2)}(T) &= \lim_{\|\pi\|\to 0} \sum_{i=0}^{n-1} \left| g^{'}(t_i^*) \right|^2 (t_{i+1} - t_i)^2, \\
&\le \lim_{\|\pi\|\to 0} \| \pi \| \sum_{i=0}^{n-1} \left| g^{'}(t_i^*) \right|^2 (t_{i+1} - t_i) \\
&= \lim_{\|\pi\|\to 0} \| \pi \| \lim_{\|\pi\|\to 0} \| \sum_{i=0}^{n-1} \left| g^{'}(t_i^*) \right|^2 (t_{i+1} - t_i) \\
&= \lim_{\|\pi\|\to 0} \| \pi \| \int_0^T \left| g^{'}(t) \right|^2 dt \\
&= 0 \ .
\end{aligned}
$$



Here we have used the fact that $g^{'}(t)$ is continuous on $[0, T]$ and therefore $\left| g^{'}(t) \right|^2$ is Riemann integrable.

□

**Example 9.2.1** *Consider* $g(t) = t^2$. *Determine* $V_g^p(1)$ *for* $p = 1, 2$.

**Solution** We have

$$g^{'}(t) = 2t, \quad \int_0^1 | \, g^{'}(t) \, | \, dt = 1 \ .$$

Hence

$$\int_0^1 | \, g^{'}(t) \, |^2 \, dt = 4 \int_0^1 t^2 dt = 4 \left. \frac{t^3}{3} \right|_0^1 = \frac{4}{3} \ .$$

But

$$\lim_{\|\pi\|\to 0} \| \pi \| \int_0^1 \left| g^{'}(t) \right|^2 dt = 0 \ .$$

Therefore

$$V_g(1) = 1; \quad V_g^2(1) = 0 \ .$$

Now by applying Result 9.2.1, we have

$$V_g^p(1) = \begin{cases} 1, & p = 1 \\ 0, & p > 1 \ . \end{cases}$$

□

## 9.3 Variations of Brownian Motion

Now, we discuss the first and second variations for the Brownian motion over $[0, T]$. We have already shown in Chapter 7 that sample paths of $W(t)$ are nowhere differentiable. Therefore the first variation does not make sense as it does not exist for the Brownian motion. Next we evaluate the second variation, also called the quadratic variation of the Brownian motion.

The quadratic variation for Brownian motion $\{W(t), t \geq 0\}$ over the interval $[0, T]$, denoted by $[W, W](T)$, is given by

$$[W, W](T) = V^2_{W(t)}(T) = \lim_{\|\pi\| \to 0} Q_\pi \ ,$$

where

$$Q_\pi = \sum_{i=0}^{n-1} (W(t_{i+1}) - W(t_i))^2 \ . \tag{9.3}$$

But now we face certain difficulties and these need to be addressed. Clearly, $Q_\pi$ is a function of the sample points $w \in \Omega$. Hence, the quadratic variation calculated for the Brownian motion for each partition is itself a random variable. Here we note that, the limit is to be taken over all partitions of $[0, T]$, with $\| \pi \| \to 0$ as $n \to \infty$. Since, for each partition $\pi$, $Q_\pi$ is a random variable, we need to specify the sense in which we are finding the limiting distribution of $Q_\pi$ for large $n$. In other words we need to specify the proper mode of convergence in these random variables. We shall use the convergence in mean square (convergence in $L^2$) sense as defined below.

**Definition 9.3.1** *Let $\{X_n, n \geq 1\}$ and $X$ be random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. We say that $X_n$ converges to $X$ in mean square sense (in $L^2$ sense) if*

$$\lim_{n \to \infty} E(|X_n - X|^2) = 0 \ .$$

In the case of Brownian motion, we will show that $\{Q_\pi\}$ converges to $T$ in mean square sense, i.e.

$$\lim_{\|\pi\|\to 0} E(|Q_\pi - T|^2) = 0 \ . \tag{9.4}$$

When the above result holds good, we say that the quadratic variation accumulated by the Brownian motion over the interval $[0, T]$ is $T$ almost surely and is denoted as $[W, W](T) = T$.

**Theorem 9.3.1** *Let $Q_\pi$ be defined as in (9.3). Then*

*(i)* $E(Q_\pi) = T$,
*(ii)* $Var(Q_\pi) \le 2 \parallel \pi \parallel T$ .

**Proof.**

(i) We have

$$E(Q_\pi) = \sum_{i=0}^{n-1} E\left(W(t_{i+1}) - W(t_i)\right)^2 \ . \tag{9.5}$$

Since, for fixed $i$, $W(t_{i+1}) - W(t_i)$ has normal distribution with mean zero and variance $(t_{i+1} - t_i)$, equation (9.5) gives

$$E(Q_\pi) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) = T \ .$$



(ii) Next we have

$$Var(Q_\pi) = \sum_{i=0}^{n-1} Var(W(t_{i+1}) - W(t_i))^2 \ . \tag{9.6}$$

But

$$Var(W(t_{i+1}) - W(t_i))^2 = E(W(t_{i+1}) - W(t_i))^4 - 2E(W(t_{i+1}) - W(t_i))^2 (t_{i+1} - t_i) + (t_{i+1} - t_i)^2 \ . \tag{9.7}$$

Since the fourth order moment of normal distribution with mean zero and variance $(t_{i+1} - t_i)$ is $3(t_{i+1} - i_i)^2$ (see Exercise 7.12 in Chapter 7), we get from (9.7)

$$\begin{aligned} Var(W(t_{i+1}) - W(t_i))^2 &= 3(t_{i+1} - t_i)^2 - 2(t_{i+1} - t_i)^2 + (t_{i+1} - t_i)^2 \\ &= 2(t_{i+1} - t_i)^2 \ . \end{aligned} \tag{9.8}$$

Substituting (9.8) in (9.6), we get

$$
\begin{aligned}
Var(Q_\pi) &= \sum_{i=0}^{n-1} 2(t_{i+1} - t_i)^2 \\
&\leq \sum_{i=0}^{n-1} 2 \parallel \pi \parallel (t_{i+1} - t_i) = 2 \parallel \pi \parallel T \ . \qquad (9.9)
\end{aligned}
$$

□

**Remark 9.3.1** *The above theorem tells that for the Brownian motion* $\{W(t), t \geq 0\}$*,* $[W, W](T) = T$ *for all* $T \geq 0$ *and almost surely. This is because* $Var(Q_\pi) = E\left(Q_\pi - E(Q_\pi)\right)^2 = E\left(Q_\pi - T\right)^2$*, which from (9.9) gives*

$$\lim_{\|\pi\| \to 0} E\left(Q_\pi - T\right)^2 = \lim_{\|\pi\| \to 0} Var(Q_\pi) = 0 \ .$$

*Therefore*

$$[W, W](T) = \lim_{\|\pi\| \to 0} Q_\pi = T \ ,$$

*where the limit is understood in the mean square sense because*

$$\lim_{\|\pi\| \to 0} E(|Q_\pi - T|^2) = 0 \ .$$

*Hence we write* $[W, W](T) = T$ *almost surely. Here the terminology almost surely means that there can be some paths of the Brownian motion for which the assertion* $[W, W](T) = T$ *is not true. But the 'the set of all such paths' has zero probability. Though we write* $[W, W](T) = T$*, we must realise that it is to be understood in the sense as described above.*

**Remark 9.3.2** *In view of Theorem 9.2.1, we have*

$$V_W^p(T) = \begin{cases} \infty, & p = 1 \\ T, & p = 2 \\ 0, & p > 2 \ . \end{cases}$$

*This concludes that, the Brownian motion* $\{W(t), t \geq 0\}$ *is of unbounded variation and finite quadratic variation for every* $t > 0$.

**Remark 9.3.3** *We know that for the given Brownian motion* $\{W(t), t \geq 0\}$*,* $[W, W](T) = T$*, i.e.*

$$\lim_{\|\pi\| \to 0} \left( \sum_{i=0}^{n-1} \left(W(t_{i+1}) - W(t_i)\right)^2 \right) = T \ . \qquad (9.10)$$

*Also for* $0 < T_1 < T_2$, $[W,W](T_2) - [W,W](T_1) = T_2 - T_1$, *the Brownian motion accumulates* $(T_2 - T_1)$ *units of quadratic variation over the interval* $[T_1, T_2]$. *Since this is true for every interval, we infer that the Brownian motion accumulates quadratic variation at rate one per unit time. This last statement we write informally as*

$$dW(t)\ dW(t) = dt\ , \tag{9.11}$$

*and remember that the* $dt$ *in (9.11) is in fact* $1.dt$.

The mathematical justification of the formula (9.11) is essentially equation (9.10). At this stage we shall also like to know some formula for $dW(t)\ dt$ and $dt\ dt$. To determine these, we need to compute the cross variation of $W(t)$ and $t$ and also the quadratic variation of $t$ itself. We have the below given theorem in this regard.

**Theorem 9.3.2** *Let* $\{W(t), t \geq 0\}$ *be the given Brownian motion and* $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ *be a partition of* $[0,T]$. *Then*

*(i)*

$$\lim_{\|\pi\| \to 0} \left( \sum_{i=0}^{n-1} \left(W(t_{i+1}) - W(t_i)\right)(t_{i+1} - t_i) \right) = 0\ ,$$

*(ii)*

$$\lim_{\|\pi\| \to 0} \left( \sum_{i=0}^{n-1} (t_{i+1} - t_i)^2 \right) = 0\ .$$



**Proof.**

(i) We observe that

$$\left|\left(W(t_{i+1}) - W(t_i)\right)(t_{i+1} - t_i)\right| \leq \operatorname*{Max}_{0 \leq i < n} \left|W(t_{i+1}) - W(t_i)\right| (t_{i+1} - t_i)\ \ .$$

Therefore

$$\left| \sum_{i=0}^{n-1} \left(W(t_{i+1}) - W(t_i)\right)(t_{i+1} - t_i) \right| \leq \operatorname*{Max}_{0 \leq k \leq n} \left|W(t_{k+1}) - W(t_k)\right| .T\ . \tag{9.12}$$

Since $W(t)$ is continuous, the R.H.S of (9.12) goes to zero as $\| \pi \| \to 0$.

(ii) We note that

$$\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2 \leq \left(\operatorname*{Max}_{0\leq k\leq n-1}(t_{i+1}-t_i)\right).\sum_{i=0}^{n-1}(t_{i+1}-t_i) = \parallel \pi \parallel T \ ,$$

which goes to zero as $\parallel \pi \parallel \to 0$.

□

**Remark 9.3.4** *In view of Theorem 9.3.2 and in analogy with (9.10) and (9.11), we can informally write*

$$dW(t)\ dt = 0\ ,\quad dt\ dt = 0\ .$$

*These relations together with $dW(t)\ dW(t) = dt$ become very handy in the working of stochastic calculus. We shall like to emphasize the word 'informally' here because actual derivation is rather involved and uses law of large numbers.*

## 9.4 The Riemann and The Riemann-Stieltjes Integrals

In this section, we discuss the notion of ordinary integral or Riemann integral and also consider an extension, the Riemann-Stieltjes integral. At the end of this section, we explain why classical integrals may fail when the integrand or the integrator are Brownian sample paths.

For the sake of consistency, we continue to work with $[0, T]$; however it is so very natural to change it, if desired, to any arbitrary $[a, b]$.

### Riemann Integral

Let $g$ be a real-valued function defined on $[0, T]$. We assume that $g$ is continuous in $[0, T]$. Consider the integral

$$S = \int_0^T g(t)\ dt\ . \tag{9.13}$$

Let $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ be an arbitrary partition of the interval $[0, T]$. For each subinterval $[t_i, t_{i+1}]$, we set $M_i = \operatorname*{Max}_{t_i\leq t\leq t_{i+1}} g(t)$ and $m_i = \operatorname*{Min}_{t_i\leq t\leq t_{i+1}} g(t)$. Then the upper Riemann sum is defined as

$$S_\pi^+ = \sum_{i=0}^{n-1} M_i(t_{i+1}-t_i)\ ,$$

and the lower Riemann sum is defined as

$$S_\pi^- = \sum_{i=0}^{n-1} m_i(t_{i+1} - t_i) \ .$$

Let the maximum step size of the partition $\pi$ be defined as

$$\| \pi \| = \operatorname*{Max}_{0 \le k \le n-1} (t_{k+1} - t_k) \ .$$

When the number $n$ of partition points goes to infinity and the length of the longest subinterval $t_{k+1} - t_k$ goes to zero (i.e. $\| \pi \| \to 0$), the upper Riemann sum $S_\pi^+$ and the lower Riemann sum $S_\pi^-$ converge to the same limit, which we call $\int_0^T g(t)dt$. Equivalently the Riemann integral can also be defined as

$$S = \int_0^T g(t) \ dt = \lim_{\|\pi\| \to 0} \sum_{i=0}^{n-1} g(t_i^*)(t_{i+1} - t_i) \ , \tag{9.14}$$

where $t_i^*$ is arbitrary point in $[t_i, t_{i+1}]$.

In the above definition we must note that the function is Riemann integrable provided the limit does not depend on which partition $\pi$ of $[0, T]$ has been taken and which point $t^*$ in $[t_i, t_{i+1}]$ has been chosen. Thus irrespective of the way the interval $[0, T]$ has been partitioned and the point $t^*$ in $[t_i, t_{i+1}]$ has been chosen, the limit of the sum (9.14), as $\| \pi \| \to 0$, remains the same. We can show that all continuous functions and all piecewise continuous bounded functions are Riemann integrable.

### Riemann-Stieltjes Integral

Suppose that $g$ and $f$ are real-valued functions defined on $[0, T]$. We assume that on $[0, T]$, $g$ is continuous and $f$ is monotonically nondecreasing. We aim to define the integral

$$S = \int_0^T g(t) \ df(t) \ . \tag{9.15}$$

As for the Riemann integral, we consider an arbitrary partition $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ of $[0, T]$. For each subinterval $[t_i, t_{i+1}]$, we set $M_i = \operatorname*{Max}_{t_i \le t \le t_{i+1}} g(t)$ and $m_i = \operatorname*{Min}_{t_i \le t \le t_{i+1}} g(t)$. The corresponding upper and lower Riemann-Stieltjes sums are respectively

$$S_\pi^+ = \sum_{i=0}^{n-1} M_i(f(t_{i+1}) - f(t_i)) \ ,$$

and

$$S_\pi^- = \sum_{i=0}^{n-1} m_i(f(t_{i+1}) - f(t_i)) \ .$$

When $\| \pi \| \to 0$, the upper Riemann-Stieltjes sum $S_\pi^+$ and the lower Riemann-Stieltjes sum $S_\pi^-$ converge to the same limit, which we call $\int_0^T g(t)df(t)$. Equivalently the Riemann-Stieltjes integral can also be defined as

$$S = \int_0^T g(t)\ df(t) = \lim_{\|\pi\|\to 0} \sum_{i=0}^{n-1} g(t_i^*)\left(f(t_{i+1}) - f(t_i)\right) \ , \tag{9.16}$$

where $t_i^*$ is arbitrary point in $[t_i, t_{i+1}]$.

We note that if $g$ is continuous on $[0, T]$, then $g$ is Riemann-Stieltjes integrable with respect to every $f$ which is monotonically nondecreasing on $[0, T]$.

The Riemann-Stieltjes integral has application in probability theory to find the expectation of a random variable $X$



$$E(X) = \int_{-\infty}^{\infty} t\ dF(t) \ ,$$

where $F(t)$ is the cumulative distributive function of $X$. For instance, if $g(t) = t$ and $F(t)$ is given by

$$F(t) = \begin{cases} 0, & t < 0 \\ p + (1-p)(1 - e^{-\lambda t}), & 0 \le t < T \\ 1, & t \ge T \ , \end{cases}$$

where $0 < p < 1$, $\lambda > 0$ and $T > 0$, then

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} tdF(t) \\
&= \int_{-\infty}^{0} tdF(t) + 0 \times P(X=0) + \int_{0}^{T} tdF(t) + T \times P(X=T) + \int_{T}^{\infty} tdF(t) \\
&= 0 + 0 + \int_{0}^{T} t\lambda(1-p)e^{-\lambda t}dt + T(1-p)e^{-\lambda T} + 0 \\
&= \frac{(1-p)}{\lambda}\int_{0}^{\lambda T} ye^{-y}dy + T(1-p)e^{-\lambda T} \\
&= \frac{(1-p)}{\lambda}(1 - \lambda Te^{-\lambda T} - e^{-\lambda T}) + T(1-p)e^{-\lambda T} \\
&= \frac{1}{\lambda}(1-p)\left(1 - e^{-\lambda T}\right) .
\end{aligned}
$$

## 9.5 Stochastic Integral and its Properties

In the last section we have presented the classical Riemann and Riemann-Stieltjes integrals. In this section we wish to extend the concept of 'integration' of a stochastic process $\{X(t), t \geq 0\}$ with respect to a given Brownian motion $\{W(t), t \geq 0\}$. But then the meaning of the word 'integration' is to be understood appropriately.

Let $\{X(t), t \geq 0\}$ be a stochastic process which is adapted to the natural filtration $\{\mathcal{F}_t, t \geq 0\}$ of Wiener process $\{W(t), t \geq 0\}$, i.e. $X(t)$ is $\mathcal{F}_t$-measurable. Let us assume $E\left(\int_0^T X^2(t)\ dt\right) < \infty$.

We next consider a partition $\pi$ of $[0, T]$ where $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$, and form the sum $\sum_{i=0}^{n-1} X(t_i)\,(W(t_{i+1}) - W(t_i))$. Now if we take the limit of this sum as $\| \pi \| \to 0$, then in analogy with the procedure discussed in Section 9.4 we can write

$$
I(T) = \int_0^T X(s)\ dW(s) = \lim_{\|\pi\| \to 0} \sum_{i=0}^{n-1} X(t_i)\,(W(t_{i+1}) - W(t_i)) \ . \tag{9.17}
$$

Here it must be noted that at present equation (9.17) is purely informal because we have not yet attached any meaning to the 'limiting process' of the sum of random variables involved in the R.H.S of this equation. In defining stochastic integral we use the mean square convergence as in Definition 9.3.1. Once we agree to this mode of convergence, then (9.17) is well defined. We therefore take (9.17) as the definition of *Stochastic integral* or *Ito integral* of the stochastic process $\{X(t), t \geq 0\}$ with respect to the Brownian motion $\{W(t), t \geq 0\}$.

**Remark 9.5.1** *Looking at the partial sums involved in (9.17) and (9.12) (or (9.11)), we find one major difference. In (9.11) and (9.12) we were free to choose any $t_i^*$ in the $i^{th}$ subinterval $[t_i, t_{i+1}]$ and evaluate the integrand $g$ at $t_i^*$. But here we are constrained to choose the left end point $t_i$ of the subinterval $[t_i, t_{i+1}]$ and take $t_i^*$ to be $t_i$ only. This will require evaluation of the integrand $X(t)$ at $t_i$ and not at any other point $t_i^*$ of our choice in $[t_i, t_{i+1}]$. This has been done so as to make Ito integral $I(t) = \int_0^t X(s)\, dW(s)$, $0 < t \leq T$, a martingale.*

### Properties of Ito integral

The stochastic integral $I(t)$, $0 < t \leq T$, satisfies the following properties.

(i) $E(I(t)) = 0$.

(ii) $E\left(\int_0^t X(s)\, dW(s)\right)^2 = E\left(\int_0^t X^2(s)\, ds\right)$. This property is called Ito isometry.

(iii) Let $\{X^{(1)}(t), t \geq 0\}$ and $\{X^{(2)}(t), t \geq 0\}$ be stochastic processes having stochastic integrals with respect to $\{W(t), t \geq 0\}$. Let $\alpha$ and $\beta$ be constants. Then

$$\int_0^t \left[\alpha X^{(1)}(s) + \beta X^{(2)}(s)\right]\, dW(s) = \alpha \int_0^t X^{(1)}(s)\, dW(s) + \beta \int_0^t X^{(2)}(s)\, dW(s)\ .$$



This is called the linearity property of the Ito integral.

(iv)
$$\int_0^t X(s)\, dW(s) = \int_0^{t_1} X(s)\, dW(s) + \int_{t_1}^t X(s)\, dW(s)$$
for $0 \leq t_1 \leq t$.

(v) The process $I(t)$ has continuous sample path.

(vi) For each $t$, $I(t)$ is $\mathcal{F}_t$- measurable.

(vii) $[I, I](t) = \int_0^t X^2(s)\, ds$.

(viii) The process $I(t) = \int_0^t X(s)\, dW(s)$, $t \in [0, T]$, is a martingale with respect to the natural Brownian filtration $\mathcal{F}_t$ $(0 \leq t \leq T)$.

This last property is very important in applications. It essentially follows because in the definition of Ito integral, in the subinterval $[t_i, t_{i+1}]$ we have taken $t_i^*$ as the left end point. So this choice of $t_i^*$ very crucial to make $I(t)$ a martingale.

We are not proving above results, and shall refer to Karatzas and Shreve [75] for the same.

**Remark 9.5.2** *The Ito integral does not have the monotonicity property. Thus* $X(t) \leq Y(t)$ *does not necessarily mean* $\int_0^T X(t)\, dW(t) \leq \int_0^T Y(t)\, dW(t)$. *We may take* $X(t) = 0$ *and* $Y(t) = 1$. *Then* $\int_0^T 1.dW(t) = W(T)$ *and* $\int_0^T 0.dW(t) = 0$. *But* $W(T)$ *is smaller than 0 with probability* $\frac{1}{2}$.

We shall now have certain examples to illustrate some of the points discussed above.

**Example 9.5.1** *Evaluate the Ito integral*

$$\int_0^T W(s)\, dW(s)\ .$$

**Solution** Let $\pi = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ be an arbitrary partition of $[0, T]$. We have

$$\int_0^T W(s)\, dW(s) = \lim_{\|\pi\| \to 0} \sum_{i=0}^{n-1} W(t_i)\, (W(t_{i+1}) - W(t_i))\ . \tag{9.18}$$

But, for each $i$, $W(t_i)$ and $W(t_{i+1}) - W(t_i)$ are independent random variables and are having normal distributions. Hence, the right hand side terms within the summation are nothing but the sum of independent random variables. Hence, the integral is nothing but the limit of sum of such random variables. Now, we have

$$\begin{aligned} Q_\pi &= \sum_{i=0}^{n-1} (W(t_{i+1}) - W(t_i))^2 \\ &= \sum_{i=0}^{n-1} \Big( W^2(t_{i+1}) - W(t_i)^2 - 2W(t_i)(W(t_{i+1}) - W(t_i)) \Big) \\ &= W^2(T) - W^2(0) - 2 \sum_{i=0}^{n-1} W(t_i)(W(t_{i+1}) - W(t_i))\ , \end{aligned}$$

i.e.

$$\sum_{i=0}^{n-1} W(t_i)\, (W(t_{i+1}) - W(t_i)) = \frac{1}{2} \Big[ W^2(T) - W^2(0) - Q_\pi \Big]\ . \tag{9.19}$$

Now taking limit as $\| \pi \| \to 0$ and using $E(Q_\pi) = T$, we get

$$\int_0^T W(s)\, dW(s) = \frac{W^2(T) - T}{2}\ .$$

We see that, unlike the Riemann-Stieltjes integral, we have an extra term $T/2$, which arises on account of the finite quadratic variation of the Brownian motion. □

**Remark 9.5.3** *In Example 9.5.1, if we form the sum* $\sum_{i=0}^{n-1} W(t_i^*)\,(W(t_{i+1}) - W(t_i))$ *with* $t_i^* = \dfrac{t_i + t_{i+1}}{2}$, *and take its limit as* $\| \pi \| \to 0$, *then we obtain this limit as* $0.5W^2(T)$. *But can we say that the Ito integral* $\int_0^T W(s)\, dW(s)$ *equals* $0.5W^2(T)$*? This is not correct because we desire Ito integral to be a martingale, and* $0.5W^2(T)$ *is not a martingale. However when we take* $t_i^* = t_i$, *the left end point of the interval* $[t_i, t_{i+1}]$, *then the limit is* $(W^2(T) - T)/2$ *which is a martingale. The requirement that* $I(t)$ *should be a martingale tells us to choose* $t_i^*$ *as* $t_i$.

**Remark 9.5.4** *There is a more compelling practical reason to choose* $t_i^*$ *as* $t_i$ *(the left point of* $[t_i, t_{i+1}]$*). Let* $t > 0$ *and* $\pi = \{0 = t_0 < t_1 < \ldots < t_n = t\}$. *We can think of* $t_0, t_1, \ldots, t_{n-1}$ *as the trading dates in the asset and* $\Delta(t_0), \Delta(t_1), \ldots, \Delta(t_{n-1})$ *as the position (number of shares) taken in the asset at each trading date and held to the next trading date. If* $I(t)$ *denote the gain from trading at each limit* $t$, *then*

$$I(t) = \begin{cases} \Delta(t_0)\,[W(t) - W(0)] = \Delta(0)W(t), & 0 \le t \le t_1 \\ \Delta(t_0)W(t_1) + \Delta(t_1)\,[W(t) - W(t_1)]\,, & t_1 \le t \le t_2 \\ \vdots & \\ \sum_{i=0}^{k-1} \Delta(t_i)\,[W(t_{i+1}) - W(t_i)] + \Delta(t_k)\,[W(t) - W(t_k)]\,, & t_k \le t \le t_{k+1} \\ \vdots & \end{cases}$$

*Obviously the process* $I(t)$ *defined above becomes our Ito integral* $\int_0^t \Delta(s)\, dW(s)$ *of the simple process* $\Delta(t)$.

**Remark 9.5.5** *In the Ito integral* $\int_0^t X(s)\, dW(s)$, *we can have a financial interpretation of integrand and integrator. The integrand represents a position in an asset and integrator represents the price of that asset. Since we need to decide the position at the beginning of each interval, we have to take* $t_i^*$ *as* $t_i$, *rather an arbitrary point in* $[t_i, t_{i+1}]$. *Thus there are theoretical as well as practical reasons for choosing* $t_i^*$ *as the left end point of the interval* $[t_i, t_{i+1}]$.

**Example 9.5.2** *Evaluate*

$$\int_0^t W(1)\ dW(s), \quad 0 \leq t \leq 1\ .$$

**Solution** Note that, $W(1)$ is not adapted to the filtration $\sigma\{W(s), 0 < s \leq t\},\ 0 \leq t \leq 1$, because it depends on future events. Hence, this Ito integral does not exist. This example shows that, assumption of the integrand adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$ is needed to have existence of the Ito integral.

□

**Example 9.5.3** *Let $X(t)$ and $Y(t)$ be suitable processes so that $I_1 = \int_0^T X(t)\ dW(t)$ and $I_2 = \int_0^T Y(t)\ dW(t)$ exist. Show that $E(I_1.I_2) = \int_0^T E\left(X(t)Y(t)\right)\ dt$.*

**Solution** We know that $I_1 I_2 = \dfrac{1}{2}\left[(I_1 + I_2)^2 - I_1^2 - I_2^2\right]$. Now we use the isometry property and get the result.

□

## 9.6 Ito-Doeblin Formula and its Variants

In many situations we are interested in 'differentiating' an expression of the form $f(W(t))$. For example, if we assume that the stock price at any time $t$ is of the form $S(t) = f(W(t))$, where $f$ is a twice continuously differentiable real-valued function, then to study the dynamical behavior of $S(t)$ we need to 'differentiate' $f(W(t))$. For the real-valued differentiable functions $f$ and $g$, we have the standard chain rule

$$\frac{d}{dt}\left(f(g(t)) = f'(g(t))\ g'(t)\right.,$$

i.e.

$$d\left(f(g(t)) = f'(g(t))\ dg(t).\right.$$

But do we have similar result for $f(W(t))$? The answer is provided by the Ito-Doeblin formula.

**1. Ito-Doeblin Formula for Brownian Motion: First Version**

Let $f$ be at least twice continuously differentiable function of $x$ and $\{W(t), t \geq 0\}$ be a Weiner process. Then

$$df(W(t)) = f'(W(t))\, dW(t) + \frac{1}{2} f''(W(t))\, dt\,, \tag{9.20}$$

or equivalently

$$f(W(t)) = f(W(0)) + \int_0^t f'(W(u))\, dW(u) + \frac{1}{2}\int_0^t f''(W(u))\, du\,. \tag{9.21}$$

Here the first integral in (9.21) is an Ito integral whereas the second integral in (9.21) is a Reimann integral.

**Remark 9.6.1** *Borrowing ideas from the classical Taylor's series we can informally write*

$$\begin{aligned} df(W(t)) &= f(W(t) + dW(t)) - f(W(t)) \\ &= f'(W(t))\, dW(t) + \frac{1}{2} f''(W(t))\, (dW(t))^2 + \ldots\,. \end{aligned}$$

*But then as per our understanding*

$$\begin{aligned} (dW(t))^2 &= (dW(t))\,(dW(t)) = dt \\ (dW(t))^3 &= (dW(t))^2\,(dW(t)) = dt\,(dW(t)) = 0\,, \end{aligned}$$

*and*

$$(dW(t))^4 = (dW(t))^2\,(dW(t))^2 = dt\,(dt) = 0\,.$$

*This is true for other higher order terms as well. Though it is not a proof of Ito-Doeblin formula (9.20), it certainly convinces about its validity.*

**Example 9.6.1.** Find $\int_0^T W(t)\, dW(t)$ using the Ito-Doeblin formula.

**Solution** Consider the first version of the Ito-Doeblin formula

$$df(W(t)) = f'(W(t))\, dW(t) + \frac{1}{2} f''(W(t))\, dt\,,$$

or equivalently

$$f(W(T)) - f(W(0)) = \int_0^T f'(W(t))\, dW(t) + \frac{1}{2}\int_0^T f''(W(t))\, dt\,.$$

Taking motivation from the fact that $\int_0^t x\,dx = t^2/2$, we choose $f(x) = x^2/2$. This gives $f'(x) = x$ and $f''(x) = 1$. Hence the above Ito-Doeblin formula gives

$$\frac{W^2(T)}{2} - 0 = \int_0^T W(t)\,dW(t) + \frac{1}{2}\int_0^T dt\ .$$

Therefore

$$\int_0^T W(t)\,dW(t) = \frac{W^2(T) - T}{2}\ . \tag{9.22}$$

□

**Remark 9.6.2** *Here we may note that, the integral* $\int_0^T W(t)\,dW(t)$ *cannot be defined path by path in Riemann-Stieltjes sense, because sample path of Brownian motion is of unbounded variation on each time interval. In the formula*

$$\int_0^T W(t)\,dW(t) = \frac{W^2(T) - T}{2},\quad t \geq 0\ ,$$

*there is an additional term of* $-\frac{T}{2}$*. This is because the local increment of the Wiener process over an increment of length* $\Delta t$ *is of the size of its standard deviation* $\sqrt{\Delta t}$*. However for a smooth continuously differentiable function* $f(t)$*, the second term in (9.22) is zero, as it should be.*

**Example 9.6.1** *Show that*

$$e^{W(t)} = 1 + \int_0^t e^{W(s)}\,dW(s) + \frac{1}{2}\int_0^t e^{W(s)}\,ds\ .$$

**Solution** We take $f(x) = e^x$ and then use the Ito-Doeblin formula.

□

### 2. Ito-Doeblin Formula for Brownian Motion: Second Version

Let $f(t,x)$ have continuous partial derivatives of at least second order and $\{W(t), t \geq 0\}$ be the given Weiner process. Then

$$df(t,W(t)) = f_t(t,W(t))\,dt + f_x(t,W(t))\,dW(t) + \frac{1}{2}f_{xx}(t,W(t))\,dt\ ,$$

or equivalently

$$f(t, W(t)) - f(0, W(0)) = \int_0^t \left[ f_t(u, W(u)) + \frac{1}{2} f_{xx}(u, W(u)) \right] du + \int_0^t f_x(u, W(u))\, dW(u)\,.$$

This formula can again be justified by considering the classical Taylor's expansion for a function of two variables. In particular we may take

$$\begin{aligned} f(t+\Delta t, W(t+\Delta t)) - f(t, W(t)) = &\left[ \frac{\partial f(t, W(t))}{\partial t}\, dt + \frac{\partial f(t, W(t))}{\partial x}\, dW(t) \right] \\ &+ \frac{1}{2} \left[ \frac{\partial^2 f(t, W(t))}{\partial t^2}\, (dt)^2 + 2 \frac{\partial^2 f(t, W(t))}{\partial t \partial x}\, dt\, dW(t) \right. \\ &\left. + \frac{\partial^2 f(t, W(t))}{\partial x^2}\, dW(t)\, dW(t) \right] + \ldots \end{aligned}$$

But since we have $dW(t)\, dW(t) = dt$, $dW(t)\, dt = 0$ and $dt\, dt = 0$, therefore,

$$\begin{aligned} f(t+\Delta t, W(t+\Delta t)) - f(t, W(t)) = &\left[ \frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2} \frac{\partial^2 f(t, W(t))}{\partial x^2} \right] dt \\ &+ \frac{\partial f(t, W(t))}{\partial x}\, dW(t), \end{aligned}$$

where

$$\frac{\partial f(t, W(t))}{\partial x} = \left.\frac{\partial f(t, x)}{\partial x}\right|_{x=W(t)}, \ \frac{\partial f(t, W(t))}{\partial t} = \left.\frac{\partial f(t, x)}{\partial t}\right|_{x=W(t)}, \ \frac{\partial^2 f(t, W(t))}{\partial x^2} = \left.\frac{\partial^2 f(t, x)}{\partial^2 x}\right|_{x=W(t)}.$$

**Example 9.6.2** *Using second version of the Ito-Doeblin formula, find* $\int_0^T W(t)\, dW(t)$.

**Solution** For any $T > 0$, we have

$$f(T, W(T)) - f(0, W(0)) = \int_0^T \left[ f_t(u, W(u)) + \frac{1}{2} f_{xx}(u, W(u)) \right] du + \int_0^T f_x(u, W(u))\, dW(u)\,.$$

Choose $f(t, x) = \dfrac{x^2}{2}$. Then $f_x(t, x) = x$, $f_t(t, x) = 0$ and $f_{xx}(t, x) = 1$. Substituting, we get

$$\frac{W^2(T)}{2} - 0 = \int_0^T \left(0 + \frac{1}{2}\right) du + \int_0^T W(u)\, dW(u)$$

Hence,

$$\int_0^T W(t)\, dW(t) = \frac{W^2(T)}{2} - \frac{T}{2}\,.$$

□

**Example 9.6.3** *Find the stochastic differential* $dX(t)$ *of* $X(t) = e^{W(t)-\frac{t}{2}}$.

**Solution** We take $f(t,x) = e^{x-\frac{t}{2}}$ and then use the second version of the Ito-Doeblin formula. This gives

$$dX(t) = df(t, W(t)) = f(t, W(t))\, dW(t) = X(t)\, dW(t)\,.$$

□

## 9.7 Stochastic Differential Equation

In this section we introduce the concept of stochastic differential equations. For this let us first consider the initial value problem (IVP)

$$\frac{dx(t)}{dt} = f(t, x(t)), \quad t \in [0,T], \quad x(0) = x_0\,, \tag{9.23}$$

where $f : [0,T] \times \mathbf{R} \to \mathbf{R}$ is a continuous function. A continuously differentiable function $x : [0,T] \to \mathbf{R}$ is called a solution of the given IVP if $x(0) = x_0$ and $x(t)$ satisfies the ordinary differential equation (ODE) (9.23) for all $t \in [0,T]$. We note that the IVP (9.23) is equivalent to the integral equation given by

$$x(t) = x_0 + \int_0^t f(s, x(s))\, ds\ . \tag{9.24}$$

Let us recollect that $f$ satisfies a Lipschitz condition if there exists a constant $k > 0$ such that

$$|\, f(t,x) - f(t,y)\,| \;\; \le \;\; k\,|\, x - y\,| \quad \text{for all } t \in [0,T], \quad x, y \in \mathbf{R}\ .$$

Then it is well known that for such a function $f$ the IVP (9.23)(or the corresponding integral equation (9.24)) has a unique solution. Further this unique solution $x(t)$ can be obtained by applying the standard Picard's method

$$x_{n+1}(t) = x_0 + \int_0^t f(s, x_n(s))\, ds, \quad (n = 0, 1, \ldots)\ ,$$

giving

$$x(t) = \lim_{n\to\infty} x_n(t) \ .$$

Suppose we now introduce randomness in the above IVP (9.23). This randomness could either be introduced in $x(0)$ or in $f$. If $x(0) = x_0$ not fixed, but rather a random variable, then for each $w \in \Omega$, the IVP (9.23) can be solved. The solution in this case will depend on $\Omega$. Thus the solution $x(t)$ will be a stochastic process $\{X(t,w), t \in [0,T], w \in \Omega\}$. Similarly, if the function $f$ depends on $w \in \Omega$, then again the solution $\{X(t,w), t \in [0,T], w \in \Omega\}$ is a stochastic process. Such differential equations are known as *random differential equations*. Here we observe that, in both these cases of introducing randomness, we solve IVP's for each $w \in \Omega$.

We now discuss the *stochastic differential equations*. These are different from random differential equations. Here we introduce uncertainties in the equation (9.23) by introducing an additive term, which is the 'derivative' of Brownian motion $W(t)$. Symbolically we write

$$\frac{dX(t)}{dt} = b(t, X(t)) + \sigma(t, X(t))\,\frac{dW(t)}{dt} \quad ,(0 \le t \le T)\ ,$$

where $b : [0,T] \times \mathbf{R} \to \mathbf{R}$ and $\sigma : [0,T] \times \mathbf{R} \to \mathbf{R}$ are two given functions. The above equation can also be symbolically written

$$dX(t) = b(t, X(t))\ dt + \sigma(t, X(t))\ dW(t)\ . \tag{9.25}$$

A naive interpretation of (9.25) tells us that the change $dX(t) = X(t+dt) - X(t)$ is caused by a change $dt$ of time, with factor $b(t, X(t))$ in combination with a change $dW(t) = W(t+dt) - W(t)$ of Brownian motion with factor $\sigma(t, X(t))$. Now borrowing analogy from classical ODE we also symbolically express (9.25) in terms of integral equation

$$X(t) = X(0) + \int_0^t b(s, X(s))\ ds + \int_0^t \sigma(s, X(s))\ dW(s)\ ,(0 < t \le T)\ . \tag{9.26}$$

Here the initial condition $X(0)$ and the coefficient functions $b(t,x)$ and $\sigma(t,x)$ are given. The differential equation (9.25) is termed as the stochastic differential equation and (9.26) is termed as *stochastic integral equation*. We shall now discuss the types of possible solution of SDE (9.25) or its equivalent integral equation version (9.26).

We must note that so far stochastic differential equation (9.25) and stochastic integral equation (9.26) are only symbolic. They will make sense only after we define them in a proper mathematical way. In this context we shall like to emphasize that mathematically only stochastic integral equation has a meaning, because we have defined a stochastic integral in strict mathematical sense. The stochastic differential equation shall always remain a notational convenience whose interpretation shall always be in terms of the corresponding stochastic integral equation.

There are two types of solution concepts for stochastic differential equations.

**1. Strong solution**

A strong solution to the SDE (9.25) is a stochastic process $\{X(t); t \in [0,T]\}$ which satisfies the following

(i) $\{X(t), t \in [0,T]\}$ is adapted to the Brownian motion, i.e. at time $t$ it is a function of $W(s),\ s \leq t$.
(ii) The integrals in (9.26) are well defined and $\{X(t); t \in [0,T]\}$ satisfies the same.
(iii) $\{X(t); t \in [0,T]\}$ is a function of the underlying Brownian sample path and of the coefficients $b(t,x)$ and $\sigma(t,x)$.

Thus a strong solution is an explicit function $f$ such that $X(t) = f(t, W(s) : s \leq t)$.

A strong solution to (9.26) is based on the path of the underlying Brownian motion. The solution $\{X(t); t \in [0,T]\}$ is said to be unique strong solution if given any other solution $\{Y(t), t \in [0,T]\}$, $P(X(t) = Y(t)) = 1$ for all $t \in [0,T]$

**2. Weak solution**

For a weak solution, the path behaviour is not essential. Hence we are only interested in the distribution of X. Thus weak solutions are sufficient to determine the expectation, variance and covariance functions of the process. In this case we do not have to know the sample paths of X.

A strong or weak solution X of the given SDE is called a *diffusion*. We may note that Brownian motion is also a diffusion process because in (9.26) we can take $b(t,x) = 0$ and $\sigma(t,x) = 1$.

We now have the following existence theorem.

**Theorem 9.7.1 (Existence Theorem)** *Let* $E(X^2(0)) < \infty$ *and* $X(0)$ *be independent of* $\{W(t), t \geq 0\}$. *Let for all* $t \in [0,T]$ *and* $x, y \in \mathbf{R}$, $b(t,x)$ *and* $\sigma(t,x)$ *be continuous and satisfy Lipschitz condition with respect to second variable, i.e.*

$$| b(t,x) - b(t,y) | + | \sigma(t,x) - \sigma(t,y) | \quad \leq \quad K \, | x - y | \quad .$$

*Then the SDE (9.26) has a unique strong solution* $\{X(t), 0 \leq t \leq T\}$.

**Example 9.7.1** *Consider the SDE*

$$dX(t) = X(t)\ dW(t),\ \textit{with}\ X(0) = 1\ .$$

*Find the strong solution using Ito-Doeblin formula.*

**Solution** We have $b(t,x) = 0$ and $\sigma(t,x) = x$. Hence, Lipschitz condition is satisfied. Now we use the second version of Ito-Doeblin formula with $f(t,x) = e^{(x-\frac{t}{2})}$. This gives

$$\frac{\partial f(t,x)}{\partial t} = \frac{-1}{2}\ e^{x-\frac{t}{2}},\ \frac{\partial f(t,x)}{\partial x} = e^{x-\frac{t}{2}} \ \text{ and } \ \frac{\partial^2 f(t,x)}{\partial x^2} = e^{x-\frac{t}{2}}\ .$$

On substituting these expressions in the second version of Ito-Doeblin formula, we get

$$\begin{aligned} dX(t) &= df(t,W(t)) \\ &= f_t(t,W(t))\ dt + f_x(t,W(t))\ dW(t) + \frac{1}{2} f_{xx}(t,W(t))\ dt \\ &= \left(-\frac{t}{2}X(t) + \frac{t}{2}X(t)\right)\ dt + X(t)\ dW(t) \\ &= X(t)\ dW(t)\ . \end{aligned}$$

Hence, the required strong solution is $X(t) = e^{W(t)-\frac{t}{2}}$.

□

**Example 9.7.2** *A stochastic process* $\{S(t), t \geq 0\}$ *is governed by*

$$dS(t) = a\ S(t)\ dt + bS(t)\ dW(t)$$

*where* $a$ *and* $b$ *are constants. Find the SDE of*

*(i)* $\sqrt{S(t)}$.
*(ii)* $ln(S(t))$.

**Solution**

(i) Choose $f(x) = x^{1/2}$, then

$$f_t = 0; \quad f_x = \frac{1}{2\sqrt{x}}; \quad f_{xx} = -\frac{1}{4x^{3/2}}\ .$$

now applying Ito-Doeblin formula, we get

$$
\begin{aligned}
d(\sqrt{S(t)}) &= f_t\ dt + f_x\ dS(t) + \frac{1}{2} f_{xx}\ dS(t)\ dS(t) \\
&= 0 + \frac{1}{2\sqrt{S(t)}} dS(t) - \frac{1}{8\sqrt{x^{3/2}}} dS(t)\ dS(t) \\
&= \frac{1}{2\sqrt{S(t)}} [a\ S(t)\ dt + bS(t)\ dW(t)] - \frac{1}{8\sqrt{(S(t))^{3/2}}} b^2 (S(t))^2 dt \\
&= \left(\frac{a}{2} - \frac{b^2}{8}\right)\ \sqrt{S(t)}\ dt + \frac{b}{2}\sqrt{S(t)}\ dW(t)\ .
\end{aligned}
$$

(ii) Choose $f(x) = ln(x)$, then

$$
f_t = 0; \quad f_x = \frac{1}{x}; \quad f_{xx} = -\frac{1}{x^2}\ .
$$

Apply Ito-Doeblin formula, we get

$$
\begin{aligned}
d(ln(S(t))) &= f_t\ dt + f_x\ dS_t + \frac{1}{2} f_{xx}\ dS_t\ dS_t \\
&= 0 + \frac{1}{S(t)} dS(t) - \frac{1}{2(S(t))^2} dS(t)\ dS(t) \\
&= \frac{1}{S(t)} [a\ S(t)\ dt + bS(t)\ dW(t)] - \frac{1}{2(S(t))^2} b^2 (S(t))^2 dt \\
&= \left(a - \frac{b^2}{2}\right)\ dt + b\ dW(t)\ .
\end{aligned}
$$

□

**Definition 9.7.1 ( Ito process)** *Let* $\{W(t), t \geq 0\}$ *be a Brownian motion and let* $\{\mathcal{F}_t, t \geq 0\}$ *be the associated natural filtration. An Ito process is a stochastic process* $\{X(t), t \geq 0\}$ *of the form*

$$
X(t) = X(0) + \int_0^t \Delta(u)\ dW(u) + \int_0^t \Theta(u)\ du\ ,
$$

*where* $X(0)$ *is non random,* $\Delta(u)$ *and* $\Theta(u)$ *are adapted processes.*

A stochastic differential equation form of the Ito process $\{X(t), t \geq 0\}$ is

$$
dX(t) = \Delta_t\ dW(t) + \Theta_t\ dt\ .
$$

All stochastic processes except those that have jumps are actually Ito processes.

**Another Extension of Ito-Doeblin Formula**
Let $X = \{X(t), t \geq 0\}$ be an Ito process and $f(t, x)$ be a function whose second order partial derivatives are continuous. Then

$$f(t, X(t)) - f(0, X(0)) = \int_0^t f_t(s, X(s))\, ds + \int_0^t f_x(s, X(s))\, dX(s) + \frac{1}{2}\int_0^t f_{xx}(s, X(s))\, d[X, X](s)\ .$$

Equivalently

$$df(t, X(t)) = f_t(t, X(t))\, dt + f_x(t, X(t))\, dX(t) + \frac{1}{2} f_{xx}(t, X(t))\, dX(t)\, dX(t)\ .$$

We may further rewrite above expression as

$$f(t, X(t)) = f(0, X(0)) + \int_0^t f_t(s, X(s))\, ds + \int_0^t f_x(s, X(s))\, \Delta(s) dW(s) + \int_0^t f_x(s, X(s))\, \Theta(s) ds + \frac{1}{2}\int_0^t f_{xx}(s, X(s))\, \Delta^2(s) ds\ .$$

This is because $dX_t = \Delta(t)\, dW(t) + \Theta(t)\, dt$ and $dX(t)\, dX(t) = \Delta^2(t)\, dt$.



**Example 9.7.1.** Find the stochastic differential of $W^3(t)$ and show that $W^3(t)$ is an Ito process.

**Solution** Using the first version of Ito-Doeblin formula, we get

$$dW^3(t) = 3W(t)\, dt + 3W^2(t)\, dW(t)\ .$$

Now using the equivalent integral equation with the condition $W(0) = 0$, we have

$$W^3(t) = 3\int_0^t W(s)\, ds + 3\int_0^t W^2(s)\, dW(s)\ .$$

Since $3W(s)$ is adapted process, the integral $\int_0^t 3W(s)\, ds$ exist and is finite. Also, $3W^2(s)$ is mean square integrable, i.e. $E\int_0^t (3W^2(s))^2\, ds < \infty$. By using the Definition 9.7.1, $W^3(t)$ is an Ito process.

□

**Example 9.7.3** *Show that $tW(t)$ is an Ito process and find the stochastic differential of $tW(t)$.*

**Solution** We take $f(t, x) = tx$ and then use the second version of Ito-Doeblin formula to get

$$tW(t) = \int_0^t W(s)\ ds + \int_0^t s\ dW(s)$$

i.e.

$$\int_0^t s\ dW(s) = t\ W(t) - \int_0^t W(s)\ ds\ .$$

Since $W(t)$ is adapted process and $f(t) = t$ is a mean square integrable, for any $t > 0$, $tW(t)$ is an Ito process. The stochastic differential of $tW(t)$ is

$$d(tW(t)) = W(t)\ dt + tdW(t)\ .$$

□

Recall that, we had proved earlier that Brownian motion $\{W(t), t \geq 0\}$, $W(0) = 0$, a martingale and have the continuous sample path with $[W, W](t) = t$. But we now have a natural converse question. If the above stated properties hold for a process then will the process be a Brownian motion? The below given theorem provides an answer to the question.

**Theorem 9.7.2 (Levy's Theorem)** *Let $\{M(t), t \geq 0\}$ be a martingale with respect to a filtration $\{\mathcal{F}_t, t \geq 0\}$. Suppose $M(0) = 0$ and $M(t)$ has continuous paths. Further let $[M, M](t) = t$ for all $t \geq 0$. Then $\{M(t), t \geq 0\}$ is a Brownian motion with associated filtration $\{\mathcal{F}_t, t \geq 0\}$.*

**Proof.** We know that Brownian motion is a martingale and increments $W(t)-W(s)$ for $s < t$ are normally distributed with mean 0 and variance $t - s$. Here we are already given that $M(t)$ is a martingale. Therefore if we prove that $M(t)$ is normally distributed with mean 0 and variance $t$, then it is a Brownian motion. For this we apply Ito-Doeblin formula to $M(t)$. Therefore for any function $f(t, x)$ whose partial derivatives exist and are continuous, we have

$$df(t, M(t)) = f_t(t, M(t))\ dt + f_x(t, M(t))\ dM(t) + \frac{1}{2} f_{xx}(t, M(t0)\ dt\ . \tag{9.27}$$

Here $dM(t)\ dM(t) = dt$ since $[M, M](t) = t$. Equation (9.27) in integral form is

$$f(t, M(t)) = f(0, M(0)) + \int_0^t [f_t(t, M(0)) + \frac{1}{2} f_{xx}(s, M(0))]\ ds + \int_0^t f_x(s, M(s))\ dM(s)\ . \tag{9.28}$$

Since $M(t)$ is a martingale, $I(t) = \int_0^t f_x(s, M(s))\, dM(s)$ is also a martingale. Also $I(0) = 0 = E(I(t))$. Now taking expectation on both sides in (9.28) we get

$$Ef(t, M(t)) = f(0, M(0)) + E\left[\int_0^t \left(f_t + \frac{1}{2}f_{xx}\right)\, ds + 0\right] . \qquad (9.29)$$

If we now prescribe $f(t, x) = \exp(ux - \frac{1}{2}u^2 t)$, for fixed $u$, then

$$f_t = -\frac{1}{2}u^2 f(t, x), \quad f_x = uf(t, x), \quad f_{xx} = u^2 f(t, x), \quad \text{and} \quad f_t + \frac{1}{2}f_{xx} = 0 .$$

Hence from (9.29) we get $E(f(t, M(t))) = f(0, M(0))$, which gives

$$E(\exp(u\, M(t) - \frac{1}{2}u^2 t)) = 1$$

i.e.

$$E(e^{uM(t)}) = e^{\frac{1}{2}u^2 t} .$$

By using the uniqueness property of moment generating function, we conclude that $M(t)$ is normally distributed with mean 0 and variance $t$.

□



## 9.8 Some Important SDE's and their Solutions

In this section we discuss some important SDE's which are frequently encountered in practice.

### 1. Geometric Brownian Motion

Let $S(t)$ be the stock price at time $t$. Let $-\infty < \mu < \infty$ be the constant growth rate of the stock and $\sigma > 0$ be the volatility. Let us consider the SDE

$$dS(t) = \mu S(t)\, dt + \sigma S(t)\, dW(t), \quad S(0) \text{ is known.}$$

We are interested in finding the strong solution of $S(t)$, if it exists. For this we first verify that conditions of Theorem 9.7.1 are satisfied. But this is true because $\mu$ and $\sigma$ are constants. Now we assume that $S(t) = f(t, W(t))$ and make use of the second version of Ito-Doeblin formula. This gives

$$df(t) = f_t\, dt + f_x\, dW(t) + \frac{1}{2}f_{xx}\, dt ,$$

where $f_t = \dfrac{\partial f(t,x)}{\partial t}$, $f_x = \dfrac{\partial f(t,x)}{\partial x}$ and $f_{xx} = \dfrac{\partial^2 f(t,x)}{\partial x^2}$. Now, on comparing with the given SDE, we get

$$f_x = \sigma f \tag{9.30}$$

$$f_t + \frac{1}{2} f_{xx} = \mu f \ . \tag{9.31}$$

Now solving equation (9.30), we get $f(t,x) = e^{\sigma x} k(t)$ , for some function $k(t)$. From here we get $f_t = k'(t)\ e^{\sigma x}$ and $f_{xx} = \sigma^2 e^{\sigma x}\ k(t)$ which on substituting in equation (9.31), we get

$$k'(t)\ e^{\sigma x} = \left(\mu - \frac{\sigma^2}{2}\right)\ e^{\sigma x}\ k(t)\ .$$

Solving the above equation, we get

$$k(t) = S(0)\ e^{\left(\mu - \frac{\sigma^2}{2}\right) t} \ .$$

Therefore, the required solution is

$$S(t) = S(0)\ e^{\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma\ W(t)} \ .$$

Here we may observe that for fixed $t$, $S(t)$ follows lognormal distribution. Hence, it can be verified that

(i) $E(S(t)) = E(S(0))\ e^{\mu t}$.
(ii) $E((S(t))^2) = E(S^2(0))\ e^{(2\mu+\sigma^2)t}$.

Further, we observe

(i) If $\mu > \dfrac{\sigma^2}{2}$ then $S(t) \to \infty$ as $t \to \infty$ almost surely.
(ii) If $\mu < \dfrac{\sigma^2}{2}$ then $S(t) \to 0$ as $t \to \infty$ almost surely.
(iii) If $\mu = \dfrac{\sigma^2}{2}$ then $S(t)$ will fluctuate between arbitrary large and arbitrary small value as $t \to \infty$.

**2. Ornstein-Uhlenbeck Process**

We consider another SDE

$$dX(t) = -\mu\ X(t)\ dt + \sigma\ dW(t), \quad X(0) = x_0\ , \tag{9.32}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are constants. This equation is often referred to as Langevin equation. This equation is useful to model the velocity at time $t$ of a free particle that performs a Brownian motion. If we set $dt = 1$ then

$$X(t+1) - X(t) = -\mu\ X(t) + \sigma\ (W(t+1) - W(t))$$

i.e.

$$X(t+1) = \hat{\mu}\ X(t) + \sigma\ (W(t+1) - W(t))\ ,$$

where $\hat{\mu} = 1 - \mu$. This model can be considered as a discrete analogue of the solution of the Langevin equation (9.32).

We derive the solution of (9.32) using Ito-Doeblin formula. Let $f(t, x) = e^{\mu t}x$. Then

$$\frac{\partial f}{\partial t} = \mu e^{\mu t}x, \qquad \frac{\partial f}{\partial x} = e^{\mu t}x, \qquad \frac{\partial^2 f}{\partial x^2} = 0\ .$$

Now using the second version of the Ito-Doeblin formula, we get



$$X(t) = e^{-\mu t}\ x_0 + \sigma \int_0^t e^{-\mu(t-s)}\ dW(s)\ . \tag{9.33}$$

This stochastic process $\{X(t), t \geq 0\}$ is called a Ornstein-Uhlenbeck process. Equation (9.33) represents the strong solution of the SDE (9.32).

Several other important SDE's, particularly those occurring in interest rate modeling, will be presented in Chapter 11.

## 9.9 Summary and Additional Notes

- An example of a function which is continuous everywhere but nowhere differentiable was given by Weierstrass in 1872. This famous example is

$$f(t) = \sum_{n=1}^{\infty} \frac{\cos(3^n t)}{2^n}\ , \quad (0 \leq t \leq 2\pi)\ .$$

- This chapter gives a brief description of certain basic concepts and important results of stochastic calculus keeping finance in view. By ignoring the various technical conditions that are required to make our definitions rigorous, these concepts are discussed.
- This chapter starts with variation of real-valued functions and then variation of Brownian motion in Sections 9.2 and 9.3 respectively.
- In Section 9.5, stochastic integral is introduced and its properties are discussed with few examples. For results concerning the existence of the general Ito stochastic integral we may refer to Mikosch [97].
- Kiyosi Ito developed Ito formula in 1951 and that is how these calculus rules were referred in earlier texts and research papers. Independently the same was studied by Wolfgang Doeblin before him, although Doeblin's work remained secret, hidden away in the safe of the French Academy of Science. In May 2000, the sealed envelope sent in February 1940 by Doeblin was finally opened. In recognition of the Doeblin's work, the Ito's formula is now referred as Ito-Doeblin formula. Interested readers may refer to Vassiliou [144] for these historical details.
- Ito-Doeblin formula with two versions are presented and illustrated with some examples in Section 9.6. The proofs of various versions of the Ito formula can be found in standard textbooks on stochastic calculus, for example, Karatzas and Shreve [75] or Shreve [122].
- The counter part of stochastic integral, stochastic differential equation is presented in Section 9.7. SDE's which admit an explicit solution are the exception from the rule. Therefore using numerical techniques, the approximation of the solution to a SDE can be obtained. Such an approximation is called a numerical solution. Numerical solutions allow us to simulate the sample paths which is the basis for Monte Carlo techniques. In this text book, we have discussed only exact and closed form solutions of stochastic differential equations and Monte Carlo simulation will be discussed in a later chapter.
- Some important SDE's which are frequently occurring in finance are presented with their solutions in Section 9.8.
- We may want to allow for the possibility that a stock price can experience sudden jumps. It may be useful to have models that incorporate jump processes and can be studied using Ito'rule for jump processes. The topic of jump process is not presented in this text book. Option pricing with jumps is studied in Merton [93].
- Stochastic calculus for processes with jumps is very important in financial applications. But as this chapter is only introductory in nature, we have avoided any

discussion on this important topic because it requires a high degree of mathematical maturity. Interested readers may refer to Shreve [122] and Karatzas and Shreve [75] in this regard.

- In many economic processes, volatility of the stock may itself be a stochastic process changing randomly over time. This flexibility produces more realistic models for pricing options. To study the models with stochastic volatility, interested readers may refer to Hull and White [66] and Heston [62].

## 9.10 Exercises

**Exercise 9.1** *Let* $f : [-1, 1] \to \mathbf{R}$ *be given by*

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \ . \end{cases}$$

*Is* $f$ *continuous? Is* $f$ *of bounded variations? What will be your answer if for* $x \neq 0$, $f(x)$ *is taken as* $x^2 \sin\left(\frac{1}{x}\right)$*?*

**Exercise 9.2** *Let* $Y_1, Y_2, \ldots$ *be independent random variables each taking two values +1 and -1 with equal probabilities. Define* $X(0) = 0$ *and* $X(n) = \sum_{j=1}^{n} Y_j$, $(n = 1, 2, \ldots)$. *This stochastic process* $\{X(n), n = 0, 1, 2 \ldots\}$ *is a symmetric random walk. Show that, the quadratic variation of* $X(n)$ *up to* $k$ *is* $k$, *i.e.* $[X(n), X(n)](k) = k$.

**Exercise 9.3** *For a Poisson process* $\{X(t), t \geq 0\}$ *with rate* $1$, *find*

*(i)* $E\left[\int_0^t X(s)\, dW(s)\right]$

*(ii)* $Var\left[\int_0^t X(s)\, dW(s)\right]$ .

**Exercise 9.4** *Find the stochastic differentials of* $\sin(W(t))$ *and* $\cos(W(t))$.

**Exercise 9.5** *Show that the process* $\{X(t), t \geq 0\}$ *given by*

$$X(t) = -1 + e^{W(t)} - \frac{1}{2} \int_0^t e^{W(s)}\, ds$$

*is a martingale.*

**Exercise 9.6** *Let* $h(s)$ *be a real-valued function which is differentiable and such that* $\int_0^t h^2(s) ds < \infty$.

*(i) Show that* $h(s)$ *is Ito-integrable.*
*(ii) Use Ito formula to prove the identity*

$$\int_0^t h(s)\ dW(s) = h(t)\ W(t) - \int_0^t h'(s)\ W(s)\ ds\ .$$

*(iii) Find the distribution of* $\int_0^t h(s)\ dW(s)$.

**Exercise 9.7** *Consider the SDE of the form*

$$dX(t) = \mu\ dt + \sigma\ dW(t), \quad X(0) = x\ .$$

*Find a deterministic function* $A(t)$ *such that* $\exp(X(t) + A(t))$ *is a martingale.*

**Exercise 9.8** *Consider the SDE of the form*

$$dX(t) = -\mu X(t)\ dt + \sigma\ dW(t)\ ,$$

*where* $X(0)$, $\mu$ *and* $\sigma > 0$ *are constants. Find the strong solution of the above SDE. Also, find the distribution of* $X(t)$.

**Exercise 9.9** *Prove that*

$$I(t) = \int_0^t X(s)\ dW(s)$$

*is a martingale.*

**Exercise 9.10** *Prove that*

$$W(T) = \int_0^T dW(t)$$

*is an Ito process.*

**Exercise 9.11** *Consider the SDE of the form* $dX(t) = X(t)\ dW(t)$ *with* $X(0) = 1$. *Prove that its solution* $X(t) = e^{W(t)-\frac{t}{2}}$ *is an Ito process.*

**Exercise 9.12** *Find the stochastic differential of* $W^2(t)$ *and show that* $W^2(t)$ *is an Ito process.*

**Exercise 9.13** *Using the first version of Ito-Doeblin formula, to evaluate* $\int_0^T W^2(t) dW(t)$.

**Exercise 9.14** *Are the random variables* $\int_0^T t\, dW(t)$ *and* $\int_0^T W(t)\, dt$ *independent? Also, find the mean and variance of these random variables.*

**Exercise 9.15** *An option is called digital option if the pay-off is 1 for* $S(T) > S(0)$ *at the time of exercise* $T$*, and zero otherwise. Find the arbitrage free price of a digital option (European) with strike price* $K = S(0)$*. You may assume that the stock price follows the SDE.*

$$dS(t) = r\ S(t)\ dt + \sigma\ S(t)\ dW(t)\ ,$$

*where* $r$ *is the interest rate and* $W(t)$ *is the Brownian motion under risk neutral probability measure.*

**Exercise 9.16** *Consider the SDE*

$$dX(t) = c(t)\ X(t)\ dt + \sigma(t)\ X(t)\ dW(t), \quad t \in [0, T]\ .$$

*Using the second version of Ito-Doeblin formula, prove that, the solution is*

$$X(t) = X(0)\ \exp\left\{\int_0^t \left(c(s) - \frac{1}{2}\sigma^2(s)\right) ds + \int_0^t \sigma(s)\ dW(s)\right\}, \quad t \in [0, T]\ .$$

**Exercise 9.17** *Consider the SDE*

$$dX(t) = A(t)\ dt + B(t)\ dW(t), \quad X(0) = x\ ,$$

*where* $A(t)$ *and* $B(t)$ *are two time-dependent functions. Find* $A(t)$ *such that* $Z(t) = \exp(X(t))$ *is an exponential martingale?*

**Exercise 9.18** *Let* $\mu_n(t)$ *be the* $n$*-th order moment about the origin for the Brownian motion* $\{W(t), t \geq 0\}$*. Using Ito-Doeblin formula, prove that*

$$\mu_n(t) = \frac{1}{2}n(n-1) \int_0^t \mu_{n-2}(t), \quad (n = 2, 3, \ldots)\ .$$

*Also, deduce that* $\mu_4(t) = 3t^2$.

**Exercise 9.19** *Consider the SDE*

$$dX(t) = -\frac{X(t)}{1-t}\ dt + dW(t), \quad 0 \leq t < 1\ ,$$

*with* $X(0) = 0$*. Prove that it's solution*

$$X(t) = (1-t) \int_0^t \frac{1}{1-s}\ dW(s), \quad 0 \leq t < 1$$

*is a Brownian bridge, between time 0 and time 1.*

**Exercise 9.20** *Consider the SDE*

$$X(t) = x_0 + \int_0^t sgn(X(s))\ dW(s)\ .$$

*Prove that, it has a weak solution, but does not have a strong solution. Further, prove that, when* $x_0 = 0$*, the weak solution of* $X(t)$ *is implicitly given by*

$$W(t) = \int_0^t sgn(X(s))\ dX(s)\ .$$

**Exercise 9.21** *Consider the SDE*

$$dX(t) = X(t)dt + dW(t)\ ,$$

*with initial condition* $X(o) = c$*. Obtain the strong solution of* $X(t)$*. Prove that,*

$$X(t) = c\ e^t + e^t \int_0^t e^{-s}\ dW(t)\ .$$

**Exercise 9.22** *Let* $Q(t) = \ln S(t)$ *and* $dQ(t) = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma\ dW(t)$*. Find* $dS(t)$*.*

# 10
# Black-Scholes Formula Revisited

## 10.1 Introduction

It is seldom a surprise to find in literature that the approach taken by mathematicians to study concepts of finance is quite different from the one adopted by financial economists. The two groups speak different languages and take their own routes to develop theories for financial mathematics, but arrive at the same conclusions. Nothing can illustrate our point better than the theory of derivative pricing. We have already seen that one can resort to any of the several methods for derivative pricing. For instance, for European options pricing, no matter whether we follow the binomial lattice approach or the CRR approach, we arrive at the same Black-Scholes (BS) formula. The two approaches have been fairly simple to understand, and hence favorite with many.

In this chapter we aim to introduce the readers to a fascinating though complex world of *change of probability measure* (one can recollect how the method of substitution works so well in the classical Riemann integral calculus). We shall also see how this new concept in stochastic integration help us to derive the BS formula for option pricing. But wait! why should we derive the BS formula again and that too with an altogether new theory? We would like to emphasize here that our aim is not to derive the BS formula only, but more importantly, to get familiarize with a more formal mathematical approach in financial instruments pricing. On first reading the initial part of this chapter one may feel like asking, what is really going on! We agree with you that the discussion is more mathematical than financial but we all agree that it is useful sometimes, even for financial economists, to recognize and appreciate what the mathematicians are saying.

Our primary aim is to present two important theorems from stochastic calculus namely, the Girsanov's theorem and the Feynman-Kac theorem. It is worth mentioning that both the theorems provide glimpse of confluence of mathematical theory with financial theory. The theorems initially appear to be a somewhat complex rules but enable us to perform change of probability measure in stochastic processes. The same will subsequently be applied to obtain the BS formula. The two theorems provide a rigorous analytical framework to build not only the BS formula but many other financial theories. They also open the gates to look for advance theories in probability measure theory and partial differential equations from the perspective of financial concepts. We urge the readers to go through the entire chapter to appreciate the presented approach.

## 10.2 Change of Probability Measure

We initiate our discussion by first describing how a random variable behaves when an original probability measure is changed to an equivalent probability measure, in fact we need to define what we meant by an equivalent probability measure.

In order to facilitate understanding of otherwise a difficult concept, we first consider a very simple situation as follows.

Let $\Omega$ be a finite sample space, specifically say $\Omega = \{\omega_1, \omega_2, \omega_3\}$. A $\sigma$-field $\mathcal{F}$ of $\Omega$ is simply the power set $P(\Omega)$. The cardinality of $\mathcal{F}$ is 8, and we write it as $\mathcal{F} = \{\emptyset, A_1, A_2, A_3, B_1, B_2, B_3, \Omega\}$, where $A_i = \{\omega_i\}$, $i = 1, 2, 3$, and $B_1 = \{\omega_1, \omega_2\}$, $B_2 = \{\omega_1, \omega_3\}$, $B_3 = \{\omega_2, \omega_3\}$. In order to speak of change of measure we need to consider at least two probability measures defined on the same measure space $(\Omega, \mathcal{F})$. Let us denote these probability measures by $P$ and $Q$. Thus, we have two probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega, \mathcal{F}, Q)$. We aim to spell out what we mean by changing from one measure $P$ to another measure $Q$. Suppose we define $P$ and $Q$ respectively as $P(\emptyset) = 0,\ P(A_i) = p_i,\ p_i > 0,\ i = 1, 2, 3,\ \sum_{i=1}^{3} p_i = 1$, and $Q(\emptyset) = 0,\ Q(A_i) = q_i,\ q_i > 0,\ i = 1, 2, 3,\ \sum_{i=1}^{3} q_i = 1$. It is worth noting at the outset that both $P$ and $Q$ agree for the null event $\emptyset$, and sure event $\Omega$. Let us now explore a relationship between $P$ and $Q$ on other events in $\mathcal{F}$. We note that

$$Q(A_i) = q_i = \frac{q_i}{p_i} P(A_i),\ i = 1, 2, 3. \tag{10.1}$$

$$\begin{aligned} Q(B_1) &= q_1 + q_2 = \frac{q_1}{p_1} p_1 + \frac{q_2}{p_2} p_2 \\ &= \Big( \frac{q_1}{p_1} \frac{p_1}{p_1 + p_2} + \frac{q_2}{p_2} \frac{p_2}{p_1 + p_2} \Big) P(B_1). \end{aligned} \tag{10.2}$$

Similar expressions can be obtained for $Q(B_2)$ and $Q(B_3)$.

We now define a non-negative random variable $Z$ on $\Omega$ as, $Z(\omega_i) = \dfrac{q_i}{p_i}$, $i = 1, 2, 3$. Then $Z$ can be written as

$$Z(\omega) = \sum_{i=1}^{3} \frac{q_i}{p_i} 1_{A_i}(\omega),$$

where $1_{A_i}$ is the indicator function of the set $A_i$ (which takes value 1 only when $\omega = \omega_i$ else value 0). With this $Z$, note that we can rewrite (10.1) and (10.2) as follows

$$Q(A_i) = E_P(Z1_{A_i}) = E_P(Z/A_i)P(A_i),\ i = 1, 2, 3,$$
$$Q(B_i) = E_P(Z1_{B_i}) = E_P(Z/B_i)P(B_i),\ i = 1, 2, 3,$$

where $E_P$ denotes the expectation operator with respect to measure $P$, and $E_P(Z/C)$ denotes the conditional expectation of $Z$ given $C \in \mathcal{F}$ with respect to $P$.

Since $E_P(Z) = E_P(Z/\Omega) = 1$, we have, in general,

$$Q(C) = E_P(Z1_C) = E_P(Z/C)P(C), \text{ for any } C \in \mathcal{F}.$$

Now, let us define a natural random variable $Y$ on $\Omega$ as $Y(\omega_i) = y_i$, $i = 1, 2, 3$. Obviously $Y$ has two alternative probability distributions on $\Omega$ depending on probability measure $P$ or $Q$. We can now give an expression for the expectation of $Y$ under measure in terms of and $P$, $Q$, $Z$ as follows.



$$E_Q(Y) = \sum_{i=1}^{3} Y(\omega_i)Q(\omega_i) = \sum_{i=1}^{3} y_i q_i = \sum_{i=1}^{3} y_i \frac{q_i}{p_i} p_i = E_P(ZY).$$

Finally, we have, $E_Q(Y) = E_P(ZY)$, for any random variable $Y$ on $\Omega$. The random variable $Z$ can be thought of as a constructed random variable such that $Z(\omega) = \dfrac{Q(\omega)}{P(\omega)}$, $\omega \in \Omega$. Note that we have specifically assumed $P(\omega) > 0$ in the above discussion to avoid division by zero. However, it is not hard to overcome this point as we can also write $Z(\omega)P(\omega) = Q(\omega)$. Note that if $Z(\omega) > 1$ (or $Z(\omega) < 1$) for some $\omega \in \Omega$, then $Q(\omega) > P(\omega)$ (or $Q(\omega) < P(\omega)$), that is, the probability $P(\omega)$ is revised upwards (or downwards). Also, for $Q$ to be a probability measure, we must have, $\sum_{\omega \in \Omega} Z(\omega)P(\omega) = 1$. Thus the random variable $Z$ is acting as a revision factor such that $E_P(Z) = 1$.

We now have a set-up to build on and extend the above idea to an arbitrary set $\Omega$, which may be uncountably infinite. But before that, we define equivalent measures.

Let $(\Omega, \mathcal{F}, P)$ be a measure space and $Z$ a non-negative $\mathcal{F}$-measurable stochastic variable such that $\int_\omega Z(\omega)dP(\omega) = 1$. We define a new probability measure $Q$ on $\Omega$ by $dQ(\omega) = Z(\omega)dP(\omega)$, that is

$$\int_\Omega X(\omega)dQ(\omega) = \int_\Omega X(\omega)Z(\omega)dP(\omega),$$

for all $\mathcal{F}$-measurable functions $X$ such that the integrals exist. In this case, $(\Omega, \mathcal{F}, Q)$ is also a measure space, but in general $Q(A) \neq P(A),\ A \in \mathcal{F}$. Note that if $P(A) = 0$, then $1_A(\omega)Z(\omega) = 0$ almost surely with respect to $P$, implying, $Q(A) = 0$. Opposite is not true. For instance, if $Z(\omega) = 0$ for $\omega \in A$ where $P(A) > 0$ then $Q(A) = 0$ despite $P(A) \neq 0$. But if $Z(\omega) > 0,\ \forall\, \omega \in \Omega$, then it works both ways.

**Definition 10.2.1** (**Equivalent measure**) *Let $(\Omega, \mathcal{F})$ be a measurable space, that is, $\mathcal{F}$ is a $\sigma$-field on $\Omega$. Consider two probability measure $P$ and $Q$ on $\mathcal{F}$. We say that $Q$ is absolutely continuous with respect to $P$ (denoted by $Q \ll P$) if*

$$P(A) = 0 \Rightarrow Q(A) = 0,\ \forall\, A \in \mathcal{F}.$$

*If $P \ll Q$ and $Q \ll P$, then $P$ and $Q$ are called equivalent measures.*

The objective of change of probability measure is typically to obtain a more favorable probability measure to work with. The vital question is that when an equivalent measure exists? The answer lies in the following theorem known as the Radon-Nikodym theorem. We skip its proof but interested readers can find the same in [67].

**Theorem 10.2.1** *Let $P$ be a probability measure on measurable space $(\Omega, \mathcal{F})$, and let $Q$ be a finite measure on $(\Omega, \mathcal{F})$. If $Q \ll P$ then there exists a nonnegative random variable $Z$ such that $Q(A) = E_P(Z1_A),\ \forall\, A \in \mathcal{F}$.*

The variable $Z$ is $P$-unique almost surely and it is referred to as the *Radon-Nikodym derivative* of $Q$ with respect to $P$. If there is no confusion in symbol, we denote it by $Z = \dfrac{dQ}{dP}$.

We now make some important observations regarding the Radon-Nikodym theorem.

(i) Since $Q$ is a finite measure, that is $Q(\omega)$ is finite, it can be normalized by $\dfrac{Q(A)}{Q(\Omega)}$. So, $Q$ is a probability measure on $\Omega$. The normalization also implies that $E_P(Z) = 1$.

(ii) $Z$ is $P$-unique almost surely means that if there is another version of $Z$, say $Z^*$, then we have $P(\{\omega \in \Omega : Z(\omega) = Z^*(\omega)\}) = 1$. That is the set where $Z$ and $Z^*$ do not agree has a measure zero with respect to $P$.

(iii) The Radon-Nikodym states the existence of $Z$ without providing an explicit expression for $Z$. In a practical situation, $Z$ is estimated by a series of observations or simulation of sufficiently large sample size.

(iv) The expectations of any random variable $X : \Omega \to \mathbf{R}^n$ under the two probability measures are related by

$$\int_{\omega \in A} f(\omega) dQ(\omega) = \int_{\omega \in A} f(\omega) Z(\omega) dP(\omega), \; A \in \mathcal{F}.$$

The above theorem is significant for financial engineers as it provides a framework to switch from real market probability measure $P$ to a risk neutral probability measure (RNPM) $Q = \widetilde{P}$ world. Though finding a RNPM $\widetilde{P}$ itself is challenging, yet it is its very existence that ensures arbitrage free market. The latter assumption helps to develop theories for general equilibrium pricing of products and services in markets. The same had been witnessed in first few chapters of the book.

Let us illustrate the concept of the Radon-Nikoym derivative through some simple examples.

**Example 10.2.1** *Let $X$ be a random variable on a $(\mathbf{R}, \mathcal{B}, P)$ where $\mathcal{B}$ is a Borel field of $\mathbf{R}$. Let the density function of $X$ with respect to $P$ be $f_X$. We assume that $f_X(x)$ is not zero for any $x \in \mathbf{R}$. Consider a transformation of $X$ to a random variable $Y$ according to $Y = g(X)$, where $g(\cdot)$ is a strictly increasing differentiable function on $\mathbf{R}$. Find the Radon-Nikodym derivative.*

**Solution** From basic probability theory, we know

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)). \end{aligned}$$

Differentiate both sides with respect to $y$, we get

$$f_Y(y) = f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))} \,,$$

where $f_Y$ is the density function of $Y$. Here we have used that $f_Y(y) = \dfrac{d}{dy}(F_Y(y))$. Define a random variable $Z$ as follows

$$Z = \frac{f_Y(g(X))g'(X)}{f_X(X)}.$$

Then $Z \geq 0$, $E_p(Z) = 1$, and

$$\int_A f_Y(y)dy = \int_A Z f_X(x)dx.$$

Thus $Z$ is a Radon-Nikodym derivative transforming $X$ into $Y$.

□

**Example 10.2.2** *Suppose a random variable $X$ on $(\mathbf{R}, \mathcal{B}, P)$, where $\mathcal{B}$ is a Borel field of $\mathbf{R}$, has an exponential distribution according to $P$ measure with parameter $\lambda$. The density function of $X$ is $f_X(x) = \lambda e^{-\lambda x}$, with mean $\frac{1}{\lambda}$. Let a random variable $Z(x) = \xi e^{kx}$ be the Radon-Nikodym derivative of $X$ such that the distribution of $X$ in the new probability $Q$ remains exponential. Find $Z$.*

**Solution** We wish $Z$ to be non-negative and $E(Z) = 1$. For these two to happen, $\xi \geq 0$, and in fact

$$\xi = \frac{1}{M_X(k)} = \frac{\lambda - k}{\lambda},$$

where $M_X(k)$ is the moment generating function of $X$ evaluated at $k$. Thus, $\lambda \geq k$. Also, note that

$$\begin{aligned} Z(x)f_X(x) &= \frac{e^{kx}}{M_X(k)}\lambda e^{-\lambda x} \\ &= e^{kx}\Big(\frac{\lambda - k}{\lambda}\Big)\lambda e^{-\lambda x} \\ &= (\lambda - k)e^{-(\lambda - k)x}. \end{aligned}$$

Thus, $X$ remains an exponentially distributed random variable under $Q$-measure with parameter $\lambda - k$.

□

**Example 10.2.3** *Consider two measure spaces $(\mathbf{R}, \mathcal{B}, P)$ and $(\mathbf{R}, \mathcal{B}, Q)$, where $\mathcal{B}$ is a Borel field of $\mathbf{R}$. Let $X$ be a $\mathcal{N}(0,1)$ random variable on $(\mathbf{R}, \mathcal{B}, P)$ with induced probability measure given by*

$$\mu_X^P(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \quad A \in \mathcal{B}.$$

*For an arbitrary constant $\theta$, construct an induced measure $\mu_X^Q$, so that $(X+\theta) \sim \mathcal{N}(0,1)$ random variable with respect to $\mu_X^Q$.*

**Solution** Note $(X+\theta) \sim \mathcal{N}(0,1)$ means $X \sim \mathcal{N}(-\theta,1)$. So, we wish $X$ to be $\mathcal{N}(-\theta,1)$ with respect to $\mu_X^Q$. Let $f^P(x)$ be a probability density function of $X$ with respect to probability measure $P$. Then,

$$\mu_X^P(A) = \int_A f^P(x)dx, \quad A \in \mathcal{B}.$$

Applying Theorem 10.2.1, we have

$$\begin{aligned}\mu_X^Q(A) &= \int_A Z(x)d\mu_X^P(x)\\ &= \int_{\mathbf{R}} Z(x)1_A(x)d\mu_X^P(x)\\ &= \int_{\mathbf{R}} Z(x)1_A(x)f^P(x)dx\\ &= \int_A Z(x)f^P(x)dx.\end{aligned}$$

Now, $\mu_X^Q(x) = \int_A d\mu_X^Q(x) = \int_A f^Q(x)dx$, where $f^Q(x)$ is a probability density function of $X$ with respect to measure $Q$. From above two relations, it is clear that

$$\int_A f^Q(x)dx = \int_A Z(x)f^P(x)dx, \quad A \in \mathcal{B}.$$

The Radon-Nikodym derivative is given by

$$Z(x) = \frac{f^Q(x)}{f^P(x)} = \frac{e^{-(x+\theta)^2/2}}{e^{-x^2/2}} = e^{-\theta x - \frac{1}{2}\theta^2}.$$

The new measure is thus defined as follows

$$\mu_X^Q(A) = \int_A e^{-\theta x - \frac{1}{2}\theta^2} d\mu_X^P(x), \quad A \in \mathcal{B}.$$

□

Example 10.2.3 plays motivating role in developing theory for change of measure, in general. What we have seen is that if a random variable $X \sim \mathcal{N}(0,1)$ with respect to probability measure $P$ then $(X+\theta) \sim \mathcal{N}(0,1)$ with respect to an equivalent probability measure $Q$, for any constant $\theta$, where $Q(A) = \int_A Z(x)dP(x)$.

We now wish to see how the above description of change of measure on random variable can be extended to a stochastic processes. In the section to follow we shall throw more light on this aspect. We shall see how a stochastic process with a non-zero drift can be transformed to another stochastic process with either drift zero or decreasing.

## 10.3 Girsanov Theorem

**Definition 10.3.1 (Radon-Nikoym derivative process)** *Let $Z$ be a random variable on space $(\Omega, \mathcal{F}, P)$ such that $Z(\omega) > 0$ almost surely with respect to $P$, and $E_P(Z) = \int_\Omega Z(\omega)dP(\omega) = 1$. The Randon-Nikodym derivative process $\{Z(t), 0 \leq t \leq T\}$ is defined by*

$$Z(t) = E_P(Z/\mathcal{F}_t), \quad t \in [0, T].$$

Define a new probability measure $Q$ on $(\Omega, \mathcal{F})$ by $dQ = ZdP$, that is,

$$Q(A) = \int_{\omega \in A} Z(\omega)dP(\omega) = E_P(Z(\omega)1_A(\omega)) = E_P(E_P(Z/\mathcal{F}_\omega)1_A(\omega)), \quad A \in \mathcal{F}.$$

We simply write it as

$$Q(A) = E_P(E_P(Z/\mathcal{F})1_A), \quad A \in \mathcal{F}.$$

The much awaited Girsanov theorem is stated as follows. The result was first proved by R. H. Cameron and W. T. Martin in the 1940's and by I. V. Girsanov in 1960.



**Theorem 10.3.1** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{W(t), 0 \leq t \leq T\}$ be a Brownian motion with associated filtration $\{\mathcal{F}_t, 0 \leq t \leq T\}$. Let $\{\theta(t), 0 \leq t \leq T\}$ be an adapted measurable process adapted to the filtration $\{\mathcal{F}_t, 0 \leq t \leq T\}$. Define*

$$Z(t) = e^{-\int_0^t \theta(t)dW(t) - \frac{1}{2}\int_0^t \theta^2(t)dt}, \quad t \in [0, T].$$

*Let $Z = Z(T)$. A sufficient condition, known as the Novikov's condition, for $Z$ to be a martingale is that*

$$E\left(e^{\frac{1}{2}\int_0^T \theta^2(t)dt}\right) < \infty.$$

*Assume $Z$ is a $P$-martingale. Define a new measure $Q$ on $(\Omega, \mathcal{F})$ as*

$$Q(A) = E_P(ZI_A).$$

*Then the process $\{\widetilde{W}(t), 0 \leq t \leq T\}$, defined by*

$$\widetilde{W}(t) = W(t) + \int_0^t \theta(t)dt,$$

*is a Brownian motion on $(\Omega, \mathcal{F}, Q)$.*

**Proof.** We first observe that $Z(t)$ is a martingale under $P$. because $Z(t)$ is a particular case of the generalized Brownian motion defined by

$$Z(t) = Z(0)e^{\left(\int_0^t \sigma(t)dW(t) - \int_0^t (\alpha(t) - \frac{1}{2}\sigma^2(t))dt\right)},$$

with $\alpha(t) = 0,\ Z(0) = 1,$ and the generalized Brownian motion is a martingale.

Consider $0 \leq s \leq t \leq T$. We have

$$E_P(Z(t)/\mathcal{F}_s) = E_P(E_P(Z(T)/\mathcal{F}_t)/\mathcal{F}_s) = E_P(Z/\mathcal{F}_s) = Z(s),$$

and $E_P(Z) = E_P(Z(T)) = 1$. Thus $\{Z(t), 0 \leq t \leq T\}$ is the Radon-Nikodym process. Moreover, if $\{Y(t), 0 \leq t \leq T\}$ is any $\mathcal{F}_t$-measurable process, then

$$\begin{aligned} E_Q(Y) &= E_P(YZ) \\ &= E_P(Y(t)Z(t)/\mathcal{F}_t) \\ &= E_P(YE_P(Z(t)/\mathcal{F}_t)) \\ &= E_P(YZ(t)). \end{aligned}$$

Hence, $Q$ is equivalent measure to $P$. The probability measure $Q$ is called an equivalent martingale measure.

Next we show that $\widetilde{W}(t)Z(t)$ is a martingale under $P$.

$$\begin{aligned} d(\widetilde{W}(t)Z(t)) &= Z(t)d\widetilde{W}(t) + \widetilde{W}(t)dZ(t) + dZ(t)d\widetilde{W}(t) \\ &= Z(t)(dW(t) + \theta(t)dt) + \widetilde{W}(t)(-\theta(t)Z(t)dW(t)) \\ &\quad +(-\theta(t)Z(t)dW(t))(dW(t) + \theta(t)dt) \\ &= Z(t)dW(t) - \theta(t)\widetilde{W}(t)Z(t)dW(t) \\ &= (1 - \theta(t)\widetilde{W}(t))Z(t)dW(t), \end{aligned}$$

where in the second last relation we have used that $dW(t)dW(t) = dt$ and $dW(t)dt = 0$. Therefore,

$$\widetilde{W}(t)Z(t) = \widetilde{W}(0)Z(0) + \int_0^t Z(t)dW(t) - \int_0^t \theta(t)\widetilde{W}(t)Z(t)dW(t).$$

The two Ito integrals on the right hand side of the above expression are martingale, so the process $\widetilde{W}(t)Z(t)$ is a martingale under $P$. Therefore, for any $0 \leq s \leq t \leq T$, we have,

$$E_P(\widetilde{W}(t)Z(t)/\mathcal{F}_s) = \widetilde{W}(s)Z(s).$$

Now, for $0 \leq s \leq t \leq T$,

$$\begin{aligned}
E_Q(\widetilde{W}(t)/\mathcal{F}(s)) &= E_Q(1_A\widetilde{W}(t)), \quad A \in \mathcal{F}_s \\
&= E_P(1_A\widetilde{W}(t)Z(t)) \\
&= E_P(1_A E_P(\widetilde{W}(t)Z/\mathcal{F}_s)) \\
&= E_P((1_A \frac{1}{Z(s)} E_P(\widetilde{W}(t)Z(t)/\mathcal{F}_s))Z(s)) \\
&= E_Q(1_A \frac{1}{Z(s)} E_P(\widetilde{W}(t)Z(t)/\mathcal{F}_s)) \\
&= \frac{1}{Z(s)} E_P(\widetilde{W}(t)Z(t)/\mathcal{F}_s) \\
&= \frac{1}{Z(s)} \widetilde{W}(s)Z(s) \\
&= \widetilde{W}(s).
\end{aligned}$$

Hence, $\widetilde{W}(t)$ is a martingale under probability measure $Q$.

Invoking Levy's theorem (Theorem 9.7.2), with notices that $d\widetilde{W}(t)d\widetilde{W}(t) = dt$, and $\int_0^t \theta(t)$, being a Riemann integration, is a continuous function of $t$, it follows that $\widetilde{W}(t)$ is a Brownian motion. The process $\{\widetilde{W}(t), 0 \leq t \leq T\}$ is adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$.

□

## 10.4 Discounted Portfolio Process

We attempt to build a general theory for derivative security which shall be used in the subsequent section to derive BS formula for option pricing. Our discussion in this section is centered around the Girsonov's theorem.

Let the stock having a price $S(t)$ per unit follows a generalized geometric Brownian motion with a constant mean return $\mu$ and a constant volatility $\sigma > 0$. The price process is governed by the linear SDE

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad t \in [0, T]. \tag{10.3}$$

We also let $\beta(t)$ to be the price of some risk-free asset which satisfies the following ordinary differential equation

$$d\beta(t) = r\beta(t)dt, \tag{10.4}$$

where $r$ is the constant risk-free interest rate.

Suppose at time $t$ we take a portfolio consisting of $a(t)$ shares of stock and $b(t)$ shares of risk-free asset. Let $V(t)$ be the value of this portfolio at $t$, that is,

$$V(t) = a(t)S(t) + b(t)\beta(t),\ t \in [0, T].$$

Then,

$$dV(t) = a(t)dS(t) + b(t)d\beta(t).$$

The discounted price of one share of stock is $\widetilde{S}(t) = e^{-rt}S(t),\ t \in [0, T]$. Apply the Ito's Lemma given at (9.20) on it, we have,

$$\begin{aligned} d\widetilde{S}(t) &= -re^{-rt}S(t)dt + e^{-rt}dS(t) \\ &= -re^{-rt}S(t)dt + e^{-rt}S(t)(\mu dt + \sigma dW(t)) \\ &= \widetilde{S}(t)((\mu - r)dt + \sigma dW(t)) \\ &= \sigma\widetilde{S}(t)d\widetilde{W}(t) \end{aligned} \tag{10.5}$$

where we denote by $\widetilde{W}(t) = W(t) + \dfrac{\mu - r}{\sigma}t,\ t \in [0, T]$. A natural interpretation for $\dfrac{\mu - r}{\sigma}$ is that the numerator is the expected return minus the risk-free rate, or the *risk premium* while the denominator is risk. The ratio is the risk premium per unit of risk and is called the *market price of risk*. In other words, the ratio reflects the additional expected return necessary to induce investors to take risk. Refresh that in the study of general market equilibrium such as the capital asset pricing model (CAPM), the appropriate risk is the systematic risk beta. It is important to note that in (10.5), the drift term is completely removed. When we remove the drift, what we are doing is removing the risk premium and the risk-free rate.

From the Girsanov's theorem (Theorem 10.2.1), there exists an equivalent measure $Q,\ Q(A) = \int_A Z(T)dP,\ A \in \mathcal{F}$, and $Z$ is a Radon-Nikodym derivative of $Q$ with respect to $P$, which turns $\widetilde{W}(t)$ into the Brownian motion with respect to $Q$. Note that, $E_P(W(t)) = 0$, on account of $W(t)$ being a Brownian motion under $P$, hence, $E_P(\widetilde{W}(t)) = \dfrac{\mu - r}{\sigma}t$. Thus, $\widetilde{W}(t)$ is not a Brownian motion under $P$ but it is a Brownian motion under $Q$. Furthermore, solution of equation (10.5), given by,

$$\widetilde{S}(t) = \widetilde{S}(0)e^{\frac{-\sigma^2}{2}t + \sigma\widetilde{W}(t)}, \quad t \in [0, T],$$

is a martingale under $Q$. This measure $Q$ is nothing else but the risk neutral probability measure (RNPM) (the probability measure in which the discounted price process becomes a martingale). We shall be denoting $Q$ by $\widetilde{P}$, as a convention. To summarize, we have to adjust the drift of the stock price process by changing the probability measures such that we obtain a martingale.

**Theorem 10.4.1** *There exist two probability measures namely the Wiener measure $P$ and the risk neutral measure $\widetilde{P}$, such that the stock price $S(t)$ at time $t$, has the following properties:*
*(i) In probability measure $P$, $S(t) = S(0)e^{\mu t+\sigma W(t)}$.*
*(ii) In probability measure $\widetilde{P}$, $S(t) = S(0)e^{(r-\frac{\sigma^2}{2})t+\sigma \widetilde{W}(t)}$, and $\{\widetilde{W}(t), t \geq 0\}$ is a Wiener process.*

Consider the discounted portfolio value process

$$\widetilde{V}(t) = e^{-rt}V(t) = e^{-rt}\left(a(t)S(t) + b(t)\beta(t)\right).$$

Then,

$$\begin{aligned}
d\widetilde{V}(t) &= -re^{-rt}V(t)dt + e^{-rt}dV(t) \\
&= -re^{-rt}\left(a(t)S(t) + b(t)\beta(t)\right)dt + e^{-rt}(a(t)dS(t) + b(t)d\beta(t)) \\
&= a(t)e^{-rt}\left(-rS(t)dt + dS(t)\right) + b(t)e^{-rt}\left(-r\beta(t)dt + d\beta(t)\right) \\
&= a(t)d\widetilde{S}(t),
\end{aligned}$$



where the last relation follows on account of (10.4) and definition of $\widetilde{S}(t)$. Thus, the increment in the value of the discounted portfolio process $\{V(t), 0 \leq t \leq T\}$ is coming from the discounted stock price process $\{S(t), 0 \leq t \leq T\}$. Also, note that $\widetilde{V}(0) = V(0)$. We thus have

$$\begin{aligned}
\widetilde{V}(t) &= V(0) + \int_0^t a(t)d\widetilde{S}(t) \\
&= V(0) + \sigma \int_0^t a(t)\widetilde{S}(t)d\widetilde{W}(t).
\end{aligned}$$

Since $\widetilde{W}(t)$ is a Brownian motion under $\widetilde{P}$ and $a(t)\widetilde{S}(t)$ is an adapted process to $\mathcal{F}_t$, $t \in [0, T]$, hence $\widetilde{V}(t)$ constitutes a martingale. Therefore,

$$\widetilde{V}(t) = E_{\widetilde{P}}\left(\widetilde{V}(T)/\mathcal{F}_t\right), \quad t \in [0, T],$$

implying

$$V(t) = e^{rt} E_{\widetilde{P}}\left(e^{-rT} V(T)/\mathcal{F}_t\right), \quad t \in [0, T]. \tag{10.6}$$

It is easy to note that $V(0) = E_{\widetilde{P}}\left(e^{-rT} V(T)\right)$, because $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

## 10.5 Risk Neutral Pricing Formula

Recall, how, in earlier part of this book, we have used risk neutral probability measure (RNPM) in replication and thereafter pricing a derivative security. The basic philosophy behind replication is that there exists a portfolio which can hedge a short position on the derivative security on the stock priced at $S(t)$. The second fundamental theorem of asset pricing guarantees the existence of such a replicating (hedging) portfolio.

For the sake of completeness we state, with a partial proof, the aforementioned theorem.

**Theorem 10.5.1** *Suppose the market model has a risk neutral probability measure (RNPM) (in other words, market follows no arbitrage principle). The market is complete (that is, every derivative security can be hedged by some portfolio) if and only if the RNPM is unique.*

**Proof.** Suppose the market is complete. Let there be two RNPMs in the market, say, $\widetilde{P}_1$ and $\widetilde{P}_2$. Define the value of the derivative security as

$$X(T) = \frac{1}{D(T)} 1_A, \quad A \in \mathcal{F}_T = \mathcal{F},$$

where $\{D(t), 0 \leq t \leq T\}$ is a discounted price process (a stochastic process analogous to $e^{-rt}$) adapted to filtration $\{\mathcal{F}_t, 0 \leq t \leq T\}$. Under assumption of complete market, this portfolio can be hedged; there exists a portfolio $V(t),\ t \in [0, T]$, such that $X(T) = V(T)$. Also, the discounted value of portfolio, $\widetilde{V}(t) = D(t)V(t)$ is a martingale under RNPM. We encourage readers to verify the same for a general discount process, $D(t) = \exp(-\int_0^t R(t)dt)$, where $\{R(t), 0 \leq t \leq T\}$ is the interest rate process adapted to filtration $\{\mathcal{F}_t, 0 \leq t \leq T\}$, on the similar lines as explained in previous section with a particular choice of constant rate of interest $R(t) = r,\ t \in [0, T]$.

Continuing our discussion, we have,

$$E_{\widetilde{P}_1}(\widetilde{V}(T)) = V(0),$$
$$E_{\widetilde{P}_2}(\widetilde{V}(T)) = V(0).$$

Consequently, because of the portfolio hedging, we have

$$E_{\widetilde{P}_1}(\widetilde{X}(T)) = E_{\widetilde{P}_1}(1_A) = V(0)$$
$$E_{\widetilde{P}_2}(\widetilde{X}(T)) = E_{\widetilde{P}_2}(1_A) = V(0).$$

Finally yielding that $\widetilde{P}_1(A) = \widetilde{P}_2(A)$, for all$A \in \mathcal{F}$, thereby means, $\widetilde{P}_1 = \widetilde{P}_2$. Thus in a complete market RNPM is unique.

□

From now onwards, we shall assume that a short position in the derivative security can be hedged by some portfolio.

Let $X(T)$ be a payoff on maturity from a derivative security written on a stock whose price process is $\{S(t), 0 \leq t \leq T\}$, and let $V(t),\ t \in [0, T]$, be the value of the associated hedging portfolio. That is, we can construct a portfolio comprising shares of the underline stock and risk-free asset. By hedging mechanism (or replicating), we have,

$$X(T) = V(T).$$

For fair price of the derivative security (no arbitrage), we must have,

$$X(t) = V(t), \quad \text{for all } t \in [0, T). \tag{10.7}$$

From (10.6), we have

$$e^{-rt} V(t) = E_{\widetilde{P}}(e^{-rT} V(T)/\mathcal{F}_t), \quad t \in [0, T],$$

which, in view of (10.7), yields

$$X(t) = e^{rt} E_{\widetilde{P}}(e^{-rT} X(T)/\mathcal{F}_t), \quad t \in [0, T]. \tag{10.8}$$

The above is called the *risk neutral pricing formula*. The applicability of the risk neutral pricing formula is much wider than only the BS formula as many of the assumptions made in say CRR model had been dropped here and we are working in a more general scenario for any derivative security which can be hedged.

Furthermore, the price of derivative security at $t = 0$ is

$$X(0) = E_{\widetilde{P}}(e^{-rT} X(T)).$$

We urge you to take a pause and carefully examine the above formula. Is it not somewhat a familiar expression that we have came across earlier too, for instance, in Chapter 3; except that there we have used notations $V(0)$ for what is $X(0)$ here and $V_{RP}(t)$ for what is $V(t)$ in the above discussion. The readers must appreciate that all the hard mathematics concepts finally lead to a path we are so familiar with, giving us a hope that we are close to deriving the BS formula. The discussion to follow is the ultimate in this context.

Consider a European call option derivative security on the stock. The payoff, to the holder of call, on maturity is $X(T) = C(T) = (S(T) - K)^+$, $K$ is the strike price of call.

$$\begin{aligned} C(t) &= e^{rt} E_{\widetilde{P}}(e^{-rT} C(T)/\mathcal{F}_t), \quad t \in [0, T] \\ &= E_{\widetilde{P}}(e^{-r(T-t)}(S(T) - K)^+/\mathcal{F}_t). \end{aligned} \tag{10.9}$$

Since $S(t)$ is a random process following the generalized Brownian motion under $\widetilde{P}$, we have,

$$S(t) = S(0) e^{\sigma \widetilde{W}(t) + (r - (\sigma^2/2))t}, \quad t \in [0, T].$$

In other words,

$$S(T) = S(t) e^{\sigma(\widetilde{W}(T) - \widetilde{W}(t)) + (r - (\sigma^2/2))(T-t)}, \quad t \in [0, T].$$

Let $\varsigma = T - t$. Then,

$$C(t) = E_{\widetilde{P}}(e^{-r\varsigma}(S(t) e^{\sigma(\widetilde{W}(T) - \widetilde{W}(t)) + (r - (\sigma^2/2))\varsigma} - K)^+/\mathcal{F}_t).$$



Now, $\widetilde{W}(T) - \widetilde{W}(t)$ is the increment in the Brownian motion for remaining time $\varsigma$, so it is independent of $\mathcal{F}_t$. Moreover $S(t)$ is $\mathcal{F}_t$-measurable. Thus, the conditional expectation in immediate above formula is superfluous. Furthermore, $\widetilde{W}(T) - \widetilde{W}(t) \sim \mathcal{N}(0, \tau)$. Setting

$$-Y = \frac{\widetilde{W}(T) - \widetilde{W}(t)}{\sqrt{T - t}} \sim \mathcal{N}(0, 1),$$

we get,

$$\begin{aligned} C(t) &= E_{\widetilde{P}}\left(e^{-r\varsigma}(S(t) e^{\sigma \sqrt{\varsigma}(-Y) + (r - (\sigma^2/2))\tau} - K)^+\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-r\varsigma}(S(t) e^{\sigma \sqrt{\varsigma}(-y) + (r - (\sigma^2/2))\tau} - K)^+ e^{-\frac{1}{2}y^2} dy. \end{aligned}$$

From now onwards, mimic the argument used in the proof presented in Section 4.3, Chapter 4 to get the Black-scholes formula for European Option. We skip its redoing here.

## 10.6 Feynman-Kac Theorem and BS Formula

Let us now turn our attention to present another famous and significant theorem in stochastic of finance. The Feynman-Kac theorem (formula), named after R. Feynman and M. Kac, establishes a link between parabolic partial differential equation (PDE) and stochastic processes.

**Theorem 10.6.1** *Let the stochastic process* $\{X(t), 0 \leq t \leq T\}$ *satisfy the following SDE*

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{10.10}$$

*where* $\mu(t, X(t))$ *and* $\sigma(t, X(t))$ *are smooth functions on* $[0, T] \times \mathbf{R}$*, called the drift and diffusion functions respectively. In addition, let the initial condition on* $X(t)$ *be* $X(0) = x$*, for some* $x \in \mathbf{R}$*. Then the solution of the following PDE*

$$g_t(t, x) + \mu(t, x)g_x(t, x) + \frac{1}{2}\sigma^2(t, x)g_{xx}(t, X) - rg(t, x) = 0, \tag{10.11}$$

*subject to the boundary condition* $g(T, X(T) = x) = h(x),\ x \in \mathbf{R}$*, is a function* $g : [0, T] \times \mathbf{R} \to \mathbf{R}$ *given by*

$$g(t, x) = E(e^{-r(T-t)}h(X(T))/X(t) = x). \tag{10.12}$$

The Feynman-Kac theorem can be used in both directions. That is, if we know that $X(t)$ follows the SDE (10.10) and we are given the PDE with boundary condition, then we can always obtain the solution $g(t, x)$ as described in (10.12). On the other hand if we know that the solution $g(t, x)$ is given by (10.12) and that $X(t)$ follows the process in (10.10), then we are assured that $g(t, x)$ satisfies the PDE (10.11).

The generator of the process in (10.10) is defined as the operator

$$\mathcal{A} = \mu(t, x(t))\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t, x(t))\frac{\partial^2}{\partial x^2}.$$

The PDE (10.11) can then be rewritten as follows

$$\frac{\partial g}{\partial t} + \mathcal{A}g - rg = 0.$$

Let us now examine how we can apply the above theorem to derive the BS formula for a derivative security.

Let the stock price $S(t)$ be driven by the process

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t).$$

As explained in (10.3) and (10.5), the risk neutral process is given by

$$dS(t) = rS(t)dt + \sigma S(t)d\widetilde{W}(t),$$

where $\widetilde{W}(t) = W(t) + \dfrac{\mu - r}{\sigma}t.$

Suppose a derivative is written on this stock. Let $V(t, S(t))$ be the price of this security at any time $t \in [0, T)$, and $V(T, S(T))$ be its payoff on maturity. Note $V : [0, T] \times \mathbf{R}_+ \to \mathbf{R}_+,\ \mathbf{R}_+$ is the set of non negative reals. By Ito's Lemma, we have

$$dV(t) = dV(t, S(t)) = \left(\frac{\partial V}{\partial t} + rS(t)\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 V}{\partial x^2}\right)dt + \sigma S(t)\frac{\partial V}{\partial x}d\widetilde{W}(t). \quad (10.13)$$

Suppose the derivative security can be hedged. We replicate the derivative security by taking a portfolio in which we take $a(t)$ shares of underline stock and $b(t)$ shares of a risk-free asset whose price is governed by an ordinary differential equation (10.4). Then, we will have,

$$\begin{aligned} dV(t) &= a(t)dS(t) + b(t)r\beta(t)dt \\ &= a(t)(rS(t)dt + \sigma S(t)d\widetilde{W}(t)) + rb(t)\beta(t)dt. \end{aligned} \quad (10.14)$$

From (10.13) and (10.14), it follows that

$$\frac{\partial V}{\partial x} = a(t) \quad \text{and} \quad \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 V}{\partial x^2} = rb(t)\beta(t).$$

Now using $b(t)\beta(t) = V(t) - a(t)S(t) = V(t) - \dfrac{\partial V}{\partial x}S(t)$, we finally get that the derivative price satisfies the following Black- Scholes PDE

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial S^2} - rV = 0. \quad (10.15)$$

The generator of the process given by $\mathcal{A} = rS\frac{\partial}{\partial S} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial S^2}$. By the Feynman-Kac theorem, the time $t$ value of the derivative is the solution of

$$V(t, S(t)) = e^{-r(T-t)}E_{\widetilde{P}}(h(S(T))/\mathcal{F}_t), \quad (10.16)$$

where $\widetilde{P}$ is a RNPM, and $h(S(T))$ is payoff of the derivative security on maturity. For instance, $h(T) = (S(T)-K)^+$ for European call option maturing at $T$ with strike price $K$. Note that (10.16) is same as (10.8) except that in (10.8) the notation has been $X(t)$ for denoting derivative price on stock whose price is $S(t)$. Now onwards, we can follow the discussion of Section 10.5 after relation (10.8) to get the BS formula for price of European call option. We leave it for readers to recognize that all approaches for derivative pricing finally lead to a similar result.

It is also a good opportunity to have a closer look at the Black-Scholes (BS) PDE (10.16). We would like to mention here that the PDE like (10.16) has motivated a community of numerical analysts to play a serious role in financial mathematics. Though, this PDE has a closed form solution which is nothing but the BS formula, there are many other PDE's which naturally arise in financial theory which require efficient numerical methods to come up with their analytic solutions.

We sketch below the method by which the closed form solution of (10.16) can be obtained.

Introduce the variables $\tau = \nu(T-t)$ and $\xi = \alpha(ln(\frac{x}{K}) + \beta(T-t))$, where the constants $\alpha$, $\beta$, and $\nu$ will be appropriately chosen later. Define

$$C(t,x) = e^{-r(T-t)}y(\tau,\xi).$$

Then,

$$C_t = \frac{\partial C}{\partial t} = re^{-r(T-t)}y + e^{-r(T-t)}y_t$$
$$= re^{-r(T-t)}y + e^{-r(T-t)}(-\alpha\beta y_\xi - \nu y_\tau).$$

$$xC_x = x\frac{\partial C}{\partial c} = xe^{-r(T-t)}y_x = e^{-r(T-t)}\alpha y_\xi.$$

$$C_{xx} = \frac{\partial C_x}{\partial x} = \frac{\partial}{\partial x}(e^{-r(T-t)}\frac{\alpha}{x}y_\xi).$$

The latter on some simplification yields

$$x^2C_{xx} = e^{-r(T-t)}(\alpha^2 y_{\xi\xi} - \alpha y_\xi).$$

We left the above for the readers to verify. Since C satisfies the BS PDE,

$$C_t + rxC_x + \frac{1}{2}\sigma^2x^2C_{xx} - rC = 0,$$

we must have,

$$-\alpha\beta y_\xi - \nu y_\tau + \nu\alpha y_\xi + \frac{1}{2}\sigma^2(\alpha^2 y_{\xi\xi} - \alpha y_\xi) = 0. \tag{10.17}$$

Choose $\beta$ and $\nu$ such that $\nu - \beta = \frac{1}{2}\sigma^2$. Then, the coefficient of $y_\xi$ in (10.17) is zero. Thus, (10.17) reduces to the following equation

$$-\nu y_\tau + \frac{1}{2}\alpha^2\sigma^2 y_{\xi\xi} = 0.$$

Choose $\nu = \alpha^2\sigma^2,\ (\alpha \neq 0)$. Then the above PDE becomes the following heat equation.

$$y_\tau = \frac{1}{2}y_{\xi\xi}.$$

Let us take a particular instance when $C(t,x)$ is the payoff of a European call option with strike $K$ and maturity $T$.

Now, we started with, $C(t,x) = e^{-r(T-t)}y(\tau,\xi)$. In order to satisfy $C(T,x) = (x-K)^+$, we must have that, when $t = T,\ \tau = \nu(T-t) = 0$ and $\xi = \alpha ln(\frac{x}{K})$. Moreover, we can choose any $\alpha \neq 0$, so, we take it to be equal to 1. Then, $x = Ke^\xi$, and hence we get the boundary condition $y(0,\xi) = (Ke^\xi - K)^+$. Thus, pricing the European call option with strike $K$ and maturity $T$ is equivalent to solving the following heat equation with boundary condition.

$$y_\tau = \frac{1}{2}y_{\xi\xi}\,, \quad y(0,\xi) = (Ke^\xi - K)^+.$$

We leave it here itself; as to solve the heat equation one needs an appropriate background and that is beyond the scope of our present work. But we ensure the readers that it will indeed be the BS formula obtained in Chapter 4.

## 10.7 Summary and Additional Notes

- In this chapter, we have introduced the concept of change of probability measure in probability measure spaces. The idea is extremely powerful and used in many occasions in financial theories besides alone the classical derivative pricing. In fact the Girsanov's theorem studied in this chapter has been also applied in literature to price exotic options and insurance premium risks. The Girsanov's theorem has been extended to more general classes of processes. Further, several variants of this theorem have appeared in literature that include its multidimensional version. The theorem has also been proved in the Hilbert space setting. To keep the theory simple, we have focussed our discussion on a basic version of the Girsanov's theorem. Of course there is no dearth of literature. We encourage you to search on a web crawler and many research articles of interest will appear.

- We have discounted the stock price and then worked with the discounted stock price to change its probability measure to RNPM. We then evaluate the option price by applying the probability distribution of the discounted stock price to the option payoff on maturity. In this way the option price is its expected payoff at expiration without discounting. Mathematicians prefer to convert the stock price to a martingale, requiring that the discounting be done beforehand.
  On the other hand, without discounting the stock price, we could have worked directly with the SDE $dS(t) = rS(t)dt + \sigma S(t)d\widetilde{W}(t),$ instead of $dS(t) = \mu S(t)dt + \sigma S(t)dW(t)$ (see, (10.3)). We would then have first evaluated the expected option payoff at expiration and do the discounting thereafter only. The financial economists prefer to follow this path. That is because it is in line with the idea that the price of any asset is its expected future value discounted to the present at an appropriate rate.
  Whichever approach one adopts, the appreciating point is that fundamental process of taking the expectation would not have altered except the simple linear adjustment by $e^{-rt}$. The same is depicted in (10.9).
- The other important highlight of the chapter is application of the Feynman-Kac theorem in derivative pricing. Again, this theorem also has many variants. It has been proved for multidimensional case as well as in variety of settings. We again urge the interested readers to explore the abstract world of this theorem on web.
- The chapter also emphasize that be it economists, financial engineers, hard-core mathematicians, optimization researchers, numerical analysts, or real practitioners, all have significant role to play in developing the subject of finance. The beautiful confluence of many streams of mathematics into one is indeed remarkable. All this make 'finance' awesome.

## 10.8 Exercises

1. Let $X$ be a random variable such that $X \sim \mathcal{N}(0,1)$ with respect to a probability measure $P$. Suppose $Q$ is another probability measure such that $E_Q(X) = E_P(XG)$, where $G = e^{-\gamma X - \frac{1}{2}\gamma^2}$. Prove that $X \sim \mathcal{N}(-\gamma, 1)$ with respect to $Q$.
   (Note: Changes of probability measures can be used to shift the means of random variables.)
2. (a)Let $X \sim \mathcal{N}(\mu, 1)$. Define another random variable $Y = X + \theta$. What is the new probability measure $Q$ such that $Y$, under $Q$, has the same distribution

as $X$ under the original probability measure?

(b)Do the same exercise in part (a) but with $X \sim \mathcal{N}(\mu, \sigma^2)$.

(Hint for (b): Take $Z = \exp\Big(\dfrac{-\theta(x-\mu) - 0.5\theta^2}{\sigma^2}\Big)$).

(Note: Question 2 is an equivalent exercise of Question 1 above).

3. Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ under probability measure $P$. Suppose the Radon-Nikodym derivative of another probability measure $Q$ is given by

$$\frac{dQ}{dP}(X) = \frac{\sigma_1}{\sigma_2}\exp\Big(\frac{(X-\mu_1)^2}{2\sigma_1^2} - \frac{(X-\mu_2)^2}{2\sigma_2^2}\Big).$$

Show that $X \sim \mathcal{N}(\mu_2, \sigma_2^2)$ under $Q$.
(Hint: Try to work out the moment generating function of $X$ under $Q$. Also see, Klebaner [78].)

4. Let $\{N(t), 0 \le t \le T\}$ be a Poisson process with rate 1 under $P$. For a constant $\lambda > 0$ define $Q$, by

$$\frac{dQ}{dP} = \exp((1-\lambda)T + N(T)ln(\lambda)).$$

Prove that $\{N(t), 0 \le t \le T\}$ is a poisson process under $Q$ with rate $\lambda$. Also see, Klebaner [78].)

5. Let $\{W(t), t > 0\}$ be a Brownian motion on probability measure space $(\Omega, \mathcal{F}, P)$. Find probability measure(s) $Q$ on $(\Omega, \mathcal{F})$ that is mutually absolutely continuous with respect to $P$, and under which the following $Y(t)$ becomes martingale
$(a)\ Y(t) = 2dt - 3dW(t) \qquad (b)\ Y(t) = 4dt + dW(t), \qquad (c)\ Y(t) = -2dW(t).$

6. Let $\{W(t), t > 0\}$ be a Brownian motion with respect to a probability measure $P$ and associated filtration $\{F(t), t > 0\}$. If $d\widehat{W(t)} = \theta dt + dW(t)$, then prove that there exists a probability measure $Q$ such that $\widehat{W(t)}$ is a Brownian motion with respect to $Q$.
(Hint: Take $Z(t) = e^{-\theta t - \frac{1}{2}\theta^2 t}$, in the Girsanov theorem.)
(Note: The above result is a particular case of the Girsanov theorem when the drift term is a constant $\theta$. However, we encourage the readers to rework the proof independently.)

7. Let $Q$ be a risk-neutral probability measure for the investor investing in rupees. Suppose the dollar/rupee exchange rate $Y(t)$ obeys a stochastic differential equation

$$dY(t) = \mu(t)Y(t)dt + \sigma(t)Y(t)dW(t).$$

If the risk-less rates of return for dollar investors and rupee investors are $r_d(t)$ and $r_{rp}(t)$, then describe the exchange rate $Y(t)$ under $Q$? (Hint: Note the drift is $r_d(t) - r_{rp}(t)$, then use the Girsanov theorem.)

8. Let $W(t) = (W_1(t), W_2(t)), t \leq T\}$ , be a 2-dimensional Brownian motion on probability measure space $(\Omega, \mathcal{F}, P)$. Find an equivalent probability measure $Q$ on $(\Omega, \mathcal{F})$ under which the following $Y(t)$ becomes a martingale

$$Y(t) = \begin{pmatrix} 2 \\ 4 \end{pmatrix} dt + \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \ t \leq T.$$

(Hint: Explore the multidimensional version of the Girsanov's theorem and use it.)

# 11
# Interest Rates and Interest Rate Derivatives

## 11.1 Introduction

Up till now we have assumed that the market interest rate is constant and observable. Be it a general concept building or a discussion on option pricing theory, European or American, the market interest rate is taken to remain the same from the time of purchase/investment till the maturity of the instrument. This has greatly simplified the discussion on pricing and hedging of these instruments. In practice, however, it cannot be the case when options and derivatives are written either on interest rates or specifically on securities whose values depend on interest rates (e.g., bond options, swaps, caps, floors etc.). Any fluctuation in interest rates can bring risk to investment on these instruments, and it is exactly what an investor is looking to hedge. It thus makes no sense to assume a constant interest rate. In fact pricing and hedging interest rate derivatives require models that can describe the evolution of the entire term structure of interest rates. This provides a good reason to study and understand what and how the interest rate theory is build.

One thing that we have to understand is that besides equity derivatives (like the one we had studied in option pricing theory or portfolio optimization), there are several other types of derivatives available in the market. Among them, the interest rate derivatives are the popular ones. The interest rate derivatives market constitutes the largest derivatives market in the world. An interest rate derivative is a derivative where the underlying asset is the right to pay or receive a notional amount of money at a given interest rate. Interest rate theory is fairly complex and requires greater insight so much so that it is generally covered only in Appendix of few introductory textbooks on financial mathematics till late. Except when either the text is aimed for advance level learning/modelling or it primarily focus on interest rate theory, there are not much references available to report at a very basic level. The difficulties are mainly due to the very nature of the fixed income instruments, implementation and calibration of various models, and above all the interest rate theory is not standardized yet. There is no well-accepted standard general model that can capture the dynamic of interest rates like the Black-Scholes model for equities.

This chapter is not intended to present the entire contents of otherwise a vast theory on interest rates, but rather to briefly provide the fragments of it so that the readers can appreciate the significance of the concept as a whole. Of course, we cite here few good books and notes [21, 24, 87, 92, 127, 146, 149] on interest rates that have been published in the last decade or so.

## 11.2 Bond Price

A good starting point is to understand the meaning of bond (or fixed securities) and some notions related to bond interest.

A government bond is a bond issued by a national government denominated in the country's own currency. Bonds issued by national governments in foreign currencies are normally referred to as sovereign bonds. The first ever government bond was issued by the English government in 1693 to raise money to fund a war against France. Bonds are also issued by public authorities, credit institutions, companies and supranational institutions in the primary markets. The most common process of issuing bonds is through underwriting. In underwriting, one or more securities firms or banks, forming a syndicate, buy an entire issue of bonds from an issuer and re-sell them to investors.

**Definition 11.2.1 (Bond)** *A bond is a contract/certificate in which the issuer promises to pay the holder a sequence of interest payments for a specified period of time, and to repay a definite amount at a specified terminal date. The promised final amount is called the face value of the bond, and the terminal date of the contract is called the maturity or redemption date. The sequence of payments at regular interval until the redemption date are done through coupons. A* zero coupon bonds *(ZCB) are the ones which makes no payment (in other words no coupons) before the redemption date.*

Thus a bond is like a loan; the issuer is the borrower (debtor), the holder is the lender (creditor), and the coupon is the interest. Bonds provide the borrower with external funds to finance long-term investments, or, in the case of government bonds, to finance current expenditure. Bonds can be looked at as long term debt investment and because of a relative low risk of default they do not usually come with a high interest rate and would largely be for investment rather than speculation purpose.

Unless otherwise stated, by a bond we shall mean a coupon bond. Whenever we wish to talk about a zero coupon bond we shall be stating it explicitly as ZCB. We shall be making two assumptions in the sequel. One that the issuer of the bond fulfills all commitments made at the issue date of bond, and second that the ZCBs are traded for all maturities $T$ in the market.

We shall be using the notation $B(t, T)$ to denote the price of a $T$-maturity bond at time $t$, $0 \leq t \leq T$.

Let the redemption date of a ZCB be $T$, and assume that its face value is Rs 1, that is, $B(T, T) = 1$. Note that $B(T, T)$ might be less than 1 if the issuer of the $T$-bond defaults. If the market rate of interest (continuously compounding) $r$ is constant, then

$$B(t, T)e^{r(T-t)} = 1,$$

equivalently,

$$B(t, T) = e^{-r(T-t)}.$$

If the face value of a ZCB is $F$ then $B(t, T) = Fe^{-r(T-t)}$.

**Example 11.2.1** *A ZCB is issued by the government on August 16, 2010 at an interest rate* 2.53% *payable semiannually. The bond will mature on August 16, 2040, with a face value Rs* 100*. Find the price of the bond.*

**Solution** This is a 30-year bond with $F =$ Rs 100 and $r = 0.0253/2 = 0.01265$, $T = 60$ (because of semiannual payments). Hence,

$$B(0, 60) = 100e^{-60*0.01265} = \text{Rs } 46.81.$$

Thus, one unit of the ZCB will be trading at Rs 46.81 on the issue date. So if an investor purchases 500 units of this bond on August 16, 2010 then he has invested Rs 23405 and will get Rs 50000 after 30 years.

□

**Example 11.2.2** *Consider the same data as in Example 11.2.1. (a) Suppose an investor wishes to purchase the bond on August 16, 2015, then what price he has to pay for the bond? (b) If the same bond is purchased on December 1, 2015, then find the bond price?*

**Solution** ($a$) The main difference is in the purchase date which is after $5$ years of issue of bond. Because of semiannual payment, $t = 10$, and thus

$$B(10, 60) = 100e^{-50*0.01265} = \text{Rs } 53.13.$$

($b$) There are 106 days between August 16, 2015 and December 1, 2015 (assuming that the bond can be traded on all 365 days of the year). Therefore, $t = 2(5 + (106/365)) = 10.581$, where the factor of $2$ is because of semiannual payment and $5$ is the number of years from the issue of the bond to August 16, 2015. So,

$$B(t, 60) = 100e^{-(60-10.581)*0.01265} = \text{Rs } 53.52.$$

□

Let us now talk about the bond which pays coupons at regular intervals.

**Definition 11.2.2 (Coupon Rate)** *It is the rate at which the bond pays interest on its face value to the holder at regular time intervals until the redemption date.*

We shall denote the coupon rate by $c$. Let $n$ coupons be paid at regular time intervals from the date of purchase of bond, say $t$, to the redemption date $T$. Let $h = (T - t)/n$. We assume that $c$ and $r$ are measured per coupon-payment period.

The following figure illustrates the cash flow on a coupon paying bond to its holder.

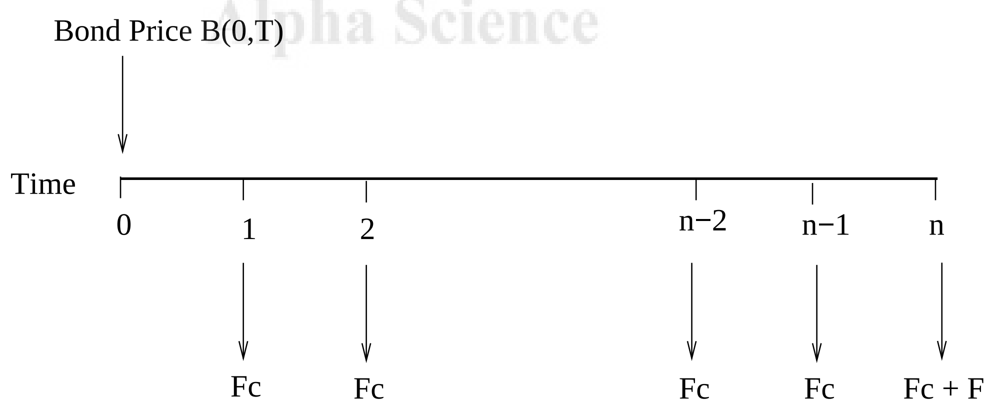



**Fig. 11.1.** Cash flow of a coupon paying bond

**Example 11.2.3** *Assume that a company issues a* $2$*-year bond with a coupon rate of* $10\%$ *payable semiannually and face value of Rs* $100$*. Find the cash flow of the bondholder.*

**Solution** The bondholder receives the following cash flows.

| *Year* | 0.5 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|
| *CashFlow*(*Rs*) | 5 | 5 | 5 | 105 |

□

Observe that, for a coupon paying bond with face value $F$ at redemption time $T$, the bond price is described by

$$B(t,T) = Fc(e^{-hr} + e^{-2hr} + \cdots + e^{-nhr}) + Fe^{-r(T-t)}, \tag{11.1}$$

Simplifying the expression, we get

$$B(t,T) = F\ c\Big(\frac{1-e^{-r(T-t)}}{e^{hr}-1}\Big) + Fe^{-r(T-t)},\ \ h = \frac{T-t}{n}.$$

Although we have taken the same coupon rate $c$ throughout the life of the bond but it may differ with the time intervals. It is worth to note from (11.1) that the price of a $T$-bond with face value $F$ paying coupons at the rate $c_i$ at dates $t_i,\ (i = 1,\dots n),\ 0 < t_1 < \cdots < t_{n-1} < t_n,\ t_n = T$, can be expressed in terms of the prices of $t_i$-ZCBs as follows.

$$B(0,T) = \sum_{i=1}^{n} \hat{c}_i\ B(0,t_i),$$



where $\hat{c}_i = c_i,\ (i = 1,\dots,n-1)$, and $\hat{c}_n = c_n + F$. The left side $B(0,T)$ means price of $T$-(coupon) bond and the right side $B(0,t_i)$ means price of the $t_i$-ZCB. In turn it amounts to say that if we can develop and analyze the price dynamics of ZCBs of all maturities $T$, then we can translate the same to analyze the dynamics of (coupon) bonds.

**Example 11.2.4** *A bond is issued by the government on August 16, 2010 at an interest rate* $8.53\%$ *and coupon rate* $7.5\%$ *payable semiannually. The bond will mature on August 16, 2040, with a face value Rs* $100$. *Find the price of the bond.*

**Solution** The bond is a 30-years bond with $c = 0.075/2 = 0.0375,\ r = 0.0853/2 = 0.04265,\ F = \text{Rs } 100,\ T = 60,\ n = 60.$ Hence, using the aforementioned expression, the bond price on date of issue is given by

$$B(0,60) = 100(0.0375)\Big(\frac{1-e^{-(60)(0.04265)}}{e^{0.04265}-1}\Big) + 100e^{-60*0.04265} = \text{Rs } 87.11.$$

□

**Example 11.2.5** *A Rs* 100 *face value 10-year bond and a coupon rate of* 8% *payable semiannually is purchased by an investor at the price of Rs* 98*. Find the annual market rate of interest on the bond.*

**Solution** We need to find $r$ given that $T = 20$ (because of semiannual data), $B(0, 20) = \text{Rs}\,98,\ F = \text{Rs}\,100,\ c = 0.04$ (again because of semiannual coupon). Thus,

$$98 = (100)(0.04)\left(\frac{1 - e^{-20r}}{e^r - 1}\right) + 100e^{-20r}.$$

Set $e^r = \xi$. Then $r = ln\xi$, and $\xi$ is a solution of the following equation.

$$49\xi^{21} - 51\xi^{20} - 50\xi + 52 = 0,\ \xi \neq 1.$$

A numerical technique can be applied to solve the above equation for $\xi$. We can verify that $\xi$ is close to 1.0415 and so $r$ is close to 4.066%. Thus, the interest rate is 8.132% per annum convertible semiannually.

□

Sometime instead of coupon rate, coupon payments of the bond are specified. For instance, suppose it is specified that a bond makes $n$ coupon payments per year for $T$ years in the amount $\frac{C}{n}$, and pays $F$ at maturity. Let $r$ be the annualized interest rate and $h = \frac{T}{n}$. Then the price of the bond is given by



$$B(0, T) = \frac{C}{n}\sum_{i=1}^{n} e^{-ihr} + Fe^{-rT}.$$

It can be observed that bond prices have close knit relationship with the interest rate $r$ (that is $B(0, T)$ can be treated as a function of $r$, although it is not explicitly specified notational in $B(0, T)$), and that is precisely what we are aiming to investigate in sections to follow. Some bonds have greater sensitivity to changes in interest rates. This risk is measure by computing how its price is likely to change when market interest rates go up or down. A bond's modified duration, a figure derived from several factors, measures this risk. The change in the bond price for a unit change in the yield, $r$, is described by

$$\frac{dB}{dr} = -h\left(\frac{C}{n}\sum_{i=1}^{n} ie^{-ihr} + nFe^{-rnh}\right),\quad h = \frac{T}{n}.$$

The percentage change in the bond price for a unit change in the interest is called *modified duration* of the bond. Thus, the modified duration of bond is

$-\frac{1}{B(0,T)}\frac{dB}{dr}$. Modified duration is stated in years. For example, a 3-year duration means the bond will decrease in value by 3% if interest rate rises 1% and increase in value by 3% if interest rate falls by 1%. For more details on this aspect, we refer to [87].

There are many other risks involved in investing in bonds, including default risk (when the bond issuer is unable to make interest payments and/or redemption repayment), market risk (the risk that the bond market as a whole declines), currency risk for foreign investors, interest rate risk (if the prevailing interest rate rises, the price of the bond will fall making the bond less attractive), inflation risk, to name a few. Some examples do exist where a government has defaulted on its domestic currency debt, such as the one in Russia in 1998, and very recently (2012) in Greece, however such examples are rare. Therefore our already assumed condition that there is no default payment on part of the issuer does not seem out of place. In the section to follow we shall be talking more about interest rate risk on bonds.

## 11.3 Term Structure of Interest

The real markets interest rates on bonds are not constants. Instead, real markets have *yield curve* which is a function of not only the redemption time $T$ but also the time of purchase $t$ of the bond.



**Definition 11.3.1 (Bond Yield)** *A rate that gives face value* 1 *at maturity* $T$ *on investment of* $B(t,T)$ *in a ZCB at time* $t$ *is called bond yield from* $t$ *to* $T$. *It is denoted by* $Y(t,T)$. *The continuously compounding bond yield* $Y(t,T)$ *is given by*

$$B(t,T) = e^{-Y(t,T)(T-t)},$$

*that is,*

$$Y(t,T) = -\frac{1}{T-t} ln(B(t,T)).$$

Note that $Y(t,T) > 0$ since $B(t,T) < 1$ for $t < T$.

In case if the face value of a ZCB is $F$ then, $Y(t,T) = -\frac{1}{T-t} ln\Big(\frac{B(t,T)}{F}\Big)$. Bond yield is an example of a *long rate of interest*.

**Definition 11.3.2 (Yield to Maturity (YTM))** *It is the yield earned on a bond when the bond is held until maturity, assuming that all coupons and principal payments shall be made on schedule.*

For a $T$-year annual coupon bond with face value $F$ and purchase price $B(t,T)$, the YTM, denoted by $Y_m$, is the solution of the following equation

$$B(t,T) = F\ c\ \sum_{i=1}^{n_r} e^{-ihY_m} + Fe^{-(T-t)Y_m}, \quad h = \frac{T-t}{n_r}. \tag{11.2}$$

where $c$ is the coupon rate and $n_r$ is the remaining number of coupon payment periods from date of purchase $t$ to date of maturity $T$ of the bond. Note that the YTM, $Y_m$, of a ZCB is same as the bond yield $Y(t,T)$. However, for a (coupon) bond, the YTM is not the same as $Y(t,T)$.

**Definition 11.3.3 ( Term Structure and Spot Rates)** *The function* $Y(t,T)$ *of two variables* $t$ *and* $T$, $t < T$, *corresponding to ZCBs, is called the term structure of interest rates. The yields* $Y(0,T)$, $T > 0$, *described by the current bond prices of* $T$*-ZCBs are called spot rates.*

Let us present an example to clarify the difference between spot rates and YTM.

**Example 11.3.1** *Consider a Rs* 100 *face value* 2*-year bond which pays annual coupon at the rate of* 5%. *Suppose the spot rates are* $Y(0,1) = 0.08$ *and* $Y(0,2) = 0.1$. *Find the YTM of the bond.*

**Solution** First we compute the bond price with the data $F = \text{Rs } 100$, $T = 2$, $c = 0.05$.

$$\begin{aligned} B(0,2) &= Fc \sum_{i=1}^{2} e^{-iY(0,i)} + Fe^{-2Y(0,2)} \\ &= (100)(0.05)e^{-0.08} + (100)(0.05)e^{-(2)(0.1)} + 100e^{-(2)(0.1)} \\ &= \text{Rs } 90.58. \end{aligned}$$

The YTM of the bond is the interest, say $Y_m$, is computed using (11.2) as follows.

$$B(0,2) = 5e^{-Y_m} + 5e^{-2Y_m} + 100e^{-2Y_m}$$

that is,

$$90.58 = 5e^{-Y_m} + 105e^{-2Y_m}.$$

After some simplification we get $Y_m = 0.099495$ or the YTM is 9.95%.

□

In practice, the yield rates or term structure on bonds, $Y(t, T)$, $0 < t \leq T \leq \hat{T}$, where $\hat{T}$ is some hypothetical maximum time horizon such that all ZCBs trading in markets will mature at or before $\hat{T}$, are not observable while the current (purchase) price of the bonds can be observed from the market. What we can do then is to solve for the YTM that equates the discounted future cash flows (coupon payments and redemption value) to the price of the bond.

**Example 11.3.2** *Consider a Rs* 100 *face value* 10*-year bond issued on August 16, 2010 which pays coupon semiannually at a rate* 10% *on August 16 and February 16 of each year. An investor purchase a bond at a price Rs* 102.54 *on December 31, 2017. Find the YTM of an investor.*

**Solution** There are 183 days between two coupon dates and 46 days between the date of purchase December 31, 2017, and the date of the next coupon on February 16, 2018 (assuming 365 trading days in an year). An investor will also receive the next 6 coupon payments starting on February 16, 2018. A little insight in the problem will give the YTM $Y_m$ of an investor as a solution of the following equation

$$102.54 = ((100)(0.05)(1 + e^{-Y_m} + e^{-2Y_m} + e^{-3Y_m} + e^{-4Y_m} + e^{-5Y_m}) + 100e^{-5Y_m})e^{-\frac{46}{183}Y_m},$$

that is,

$$(105\xi^6 - 100\xi^5 - 5)\xi^{0.25137} + 102.54(1 - \xi) = 0, \quad \xi = e^{-Y_m}, \ \xi \neq 1.$$

The above can be solved for $Y_m$ by some numerical technique. We can verify that $\xi$ is close to 0.95. The YTM per annum is thus close to $2Y_m * 100\% = 10.26\%$.

□

The term structure of interest rates is also known as **yield curve**. It is a very common bond valuation method. The yield curve is constructed by interpolating the market data of yield to maturities and the corresponding maturity dates of finitely many benchmark fixed-income bonds. The yield curve measures the market's expectations of future interest rate. The *short-term interest rate* can be thought of as shortest maturity yield or perhaps the overnight rate offered by the market. The exact shape of the curve can be different at any point in time. Any change in the shape of a normal yield curve is an indication that the investors need to change their outlook on economy.

Under normal market conditions, wherein investors believe that that there will be no significant changes in the economy, such as in inflation rates, and that the

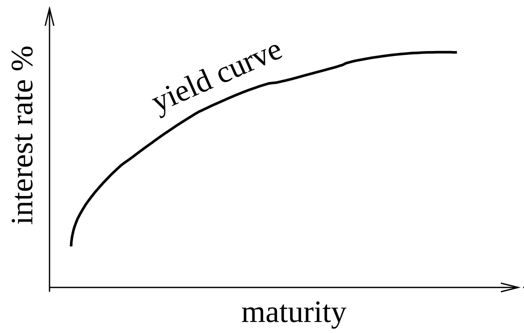



**Fig. 11.2.** Normal yield curve

economy will continue to grow at a normal rate, the yield curve generally looks like the one shown in Fig. 11.2.

A *flat yield curve* usually occurs when there are mixed indications in the market. There are speculations that short-term interest rates will rise and simultaneously there are signals that long-term interest rates will fall. If the initial term structure is flat, then the yields $Y(0, T)$ may be independent of $T$ as shown in Fig. 11.3(a).

The market expects long-term fixed income securities to offer higher yields than short-term fixed income securities. This is a normal expectation because short-term instruments generally hold less risk than long-term instruments. Sometimes, however, abnormal conditions in the economic environment can result in short-term interest rates rising above that offered by long-term fixed income investments. The result is a *negative yield curve* depicted in Fig. 11.3(b).

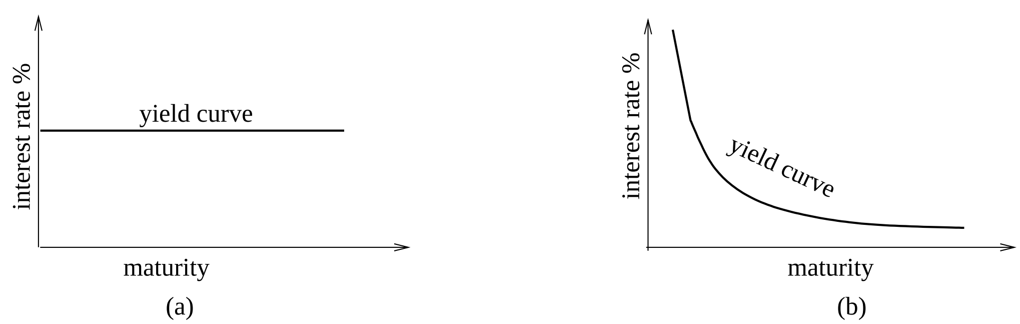



**Fig. 11.3.** (a) Flat yield curve; (b) Negative yield curve

To keep the matter simple for understanding, in the example to follow, we assume that the bond yield $Y(t,T)$ is independent of bond maturity time $T$. So, we denote it by $Y(t)$.

**Example 11.3.3** *Consider a Rs* 100 *face value* 2*-year ZCB which is currently trading at Rs* 90*. Analyze how the ZCB price gets affected with a change in the bond yield.*

**Solution** We first compute the present yield value $Y(0)$ of the ZCB under consideration.

$$B(0,2) = 100e^{-2Y(0)} \text{ with } B(0,2) = \text{Rs } 90,$$

yields

$$Y(0) = \frac{ln(0.9)}{2} = 0.0527.$$

Thus the bond yield at time $t = 0$ is $Y(0) = 5.27\%$.

Suppose after one year, the bond yield $Y(1)$ changes to 6%. It amounts to assuming that $Y(1,T) = 0.06$, for all $T > 1$. Then

$$B(1,2) = 100e^{-0.06} = 94.18,$$

which means that one unit of the same bond would be trading at a cost of Rs 94.18 if bond yield becomes 6%. The logarithmic return on the bond is



$$ln\left(\frac{94.18}{90}\right) = 0.0454 \text{ or } 4.54\%.$$

On the other hand if the bond yield reduces to 4% after 1 year. Then the price of the bond becomes

$$B(1,2) = 100e^{-0.04} = 96.08,$$

which means that one unit of the bond will be trading at Rs 96.08. The logarithmic return, in this case, is

$$ln\left(\frac{96.08}{90}\right) = 0.0654 \text{ or } 6.54\%.$$

In case if the yield $Y(1)$ remains the same, that is, 5.27%, then it can easily be seen that $B(1,2) = \text{Rs } 94.87$ with logarithmic return 5.265%.

□

From above discussion we can realize that an investor in a ZCB is better off if the bond yield decreases.

Next suppose we wish to have a logarithmic return on investment to be say 7%. Then the bond price $B(1,2)$ and the corresponding bond yield $Y(1)$ should respectively be

$$B(1,2) = B(0,2)e^{0.07} = 0.9653 \text{ or Rs } 96.53\,,$$
$$Y(1) = -ln(0.9653) = 0.0354 \text{ or } 3.54\%\,.$$

Let us ask another question. Can we get a yield of 11% on the bond after one year? Suppose yes; then $B(1,2) = 90e^{-0.11} = 100.46$, which is not possible as $B(1,2) \ngtr$ Rs 100, the face value of the bond.

When the term structure is not flat, the premium of a bond with respect to its face value may vary with the time to maturity even for bonds with the same coupon rate of interest. The same is illustrated in the following example.

**Example 11.3.4** *Consider a Rs* 100 *face value bond with coupon rate of* 4.0% *per annum paid semiannually. The spot rates of interest in the market are given as follows.*

| *Year* | *0.5* | *1* | *1.5* | *2* | *2.5* | *3* | *3.5* | *4* |
|---|---|---|---|---|---|---|---|---|
| *Spot Rate (in %)* | *3* | *3* | *3.5* | *3.5* | *4* | *4* | *4.5* | *4.5* |



*Find the price of the bond if it matures in (a) 6 months (b) 2 years, and (c) 4 years.*

**Solution** (a) Note that $T = 1$ and $Y(0,1) = 0.03/2 = 0.015$. Thus,

$$B(0,T) = Fce^{Y(0,1)} + Fe^{Y(0,1)} = 102e^{-0.015} = \text{Rs } 100.48.$$

(b) Here, $T = 4$ and $Y(0,i) = (i^{th} \text{ spot rate})/2$.

$$\begin{aligned} B(0,T) &= Fc\sum_{i=1}^{T} e^{-iY(0,i)} + Fe^{Y(0,T)T} \\ &= 2(e^{-0.015} + e^{-(2)(0.015)} + e^{-(3)(0.0175)} + e^{-(4)(0.0175)}) + 100e^{-(4)(0.0175)} \\ &= \text{Rs } 96.92. \end{aligned}$$

(c) In this case $T = 8$. So,

$$
\begin{aligned}
B(0,T) &= Fc\sum_{i=1}^{T} e^{-iY(0,i)} + Fe^{Y(0,T)T} \\
&= 2(e^{-(3)(0.015)} + e^{-(7)(0.0175)} + e^{-(11)(0.02)} + e^{-(15)(0.0225)}) + 100e^{-(8)(0.0225)} \\
&= \text{Rs } 90.24.
\end{aligned}
$$

Note that we have a premium bond for the case of $(a)$ with the price higher than the face value while for the cases $(b)$ and $(c)$, we have a discount bond with the price lower than the face value.

□

## 11.4 Forward Rates

We begin this section with an example to familiarize the readers with another very important type of interest rate closely related to term structure of ZCB. This will also highlight the relationship between spot rates of different time maturity.

We have already seen that the yield vary through time primarily because of expected variation in the inflation rates in the markets. This in turn causes bond price fluctuations even if the bonds have same maturity value.

**Example 11.4.1** *Consider two ZCBs. Bond $A$ is a 1-year bond and bond $B$ is a 2-year bond. They both have face values Rs 100. The one-year spot rate is 5% and the two-year spot rate is 8%. Find the two bonds prices.*

**Solution** We can easily calculate the present value for bond $A$ and bond $B$ as follows.

$$
\begin{aligned}
B_A(0,1) &= 100e^{-0.05} = \text{Rs } 95.12 \\
B_B(0,2) &= 100e^{-(2)(0.08)} = \text{Rs } 85.21.
\end{aligned}
$$

□

Through the above example we would like to note something important about the relationship between one-year and two-year spot rates. If one invests in a 2-year ZCB that yields 8%, his wealth at the end of two years is the same as if he/she received yields of 5% over the first year and an 11% over the second year because $100e^{(2)(0.08)} = 100(e^{0.05})(e^{0.11})$. This hypothetical yield of 11% for the second year is called the **forward rate**. Thus, we can think of an investor with a 2-year ZCB as getting the one-year spot rate of 5% and locking in 11% over the second year.

**Definition 11.4.1 (Forward Rate)** *The rate at time $t$, denoted by $f(0,t,T)$, such that the present price of a ZCB with maturity $T$ can be generated at time $T$ by locking a present price of a ZCB with maturity $t$ at time $t$. That is, the continuously compounded forward rate for $[t,T]$ prevailing currently (at time zero) satisfies*

$$B(0,T) = B(0,t)e^{-(T-t)f(0,t,T)}.$$

In other words,

$$f(0,t,T) = -\frac{1}{T-t}(ln(B(0,T)) - ln(B(0,t)))$$

$$= \frac{TY(0,T) - tY(0,t)}{T-t}. \qquad (11.3)$$

Since $B(t,t) = 1$, hence

$$f(t,t,T) = -\frac{1}{T-t}ln(B(t,T)) = Y(t,T).$$

Let us see some more examples and interpretation to clarify what we mean by forward rate.



**Example 11.4.2** *If the one-year spot rate is 8% and the two-year spot rate is 9%, what is $f(0,1,2)$? Interpret the result from an investor perspective.*

**Solution** From (11.3), we have $f(0,1,2) = 2(0.09) - 0.08 = 0.10$ or 10%. Consider an individual investing in a 2-year ZCB yielding 9%. Equivalently, it is same as if an investor receives 8% over the first year and simultaneously locks in 10% over the second year. □

Forward rate can also be viewed as an interest rate which is specified at a current time $t$ for a loan that will occur at a specified future date $T$. Forward interest rates also include a term structure which shows the different forward rates offered to loans of different maturities.

**Example 11.4.3** *Suppose you wish to take a loan of Rs 100000 for your child admission one month from now, and you expect to have means to repay the loan along with interest on loan after 6 months from now. Assume that the market spot rates for 1 month and 6 months are respectively 0.35% and 0.55% . What interest rate does the bank will offer you to construct this loan?*

**Solution** You can arrange Rs 100000 by choosing the following strategy. Compute the discounted value of Rs 100000 for 1 month from now, that is, $(100000)e^{-0.0035} =$

Rs 99650.61. So, you can take a loan of Rs 99650.61 today for the period of 6 months. Invest this loan in purchasing 996.5061 units of 1-month ZCB of Rs 100 face value. This will give $(996.5061)(100e^{0.0035})$ = Rs 100000 after 1 month from today, and this can be used for child's admission purpose. Now think of repaying the loan after 6 months from today. The principal along with interest comes out to be $(99650.61)e^{(6)(0.0055)}$ = Rs 102993.94.

Instead suppose you decided to wait and take a loan of Rs 100000 after 1-month from now. In order to maintain no arbitrage position, the interest rate that the bank should charge on your loan to be paid after 5 months is $\frac{1}{5}ln\Big(\frac{102993.94}{100000}\Big) = 0.0059$. It amounts to saying that an interest 7.08% (per annum) will be charged by the bank on your loan. Observe that the same interest rate can be obtained directly from (11.3).

□

To clarify how the forward rates are computed in real markets, we construct a *forward rate agreement (FRA)*. A prototypical FRA is an over the counter (OTC) contract involving three time instants say $t,\ T$ and $T+\tau,\ t < T < (T+\tau)$, where $t$ is the current time, $T$ is the expiry time, and $(T+\tau)$ is the maturity time. Suppose that today is day $t$, and that at time $T$ we want to lend Rs 1 for the period $\tau$ , earning the implied forward rate $f(t,T,T+\tau)$ over the interval from $T$ to $T+\tau$. In other words, we want to accomplish on day $t$ the position that allows cash going out on day $T$ and coming in on day $T+\tau$. To meet this cash flow, we need to borrow on day $t$ with a $T$-day maturity (to generate a cash outflow on day $T$) and lend with a $T+\tau$-day maturity (to generate a cash inflow on day $T+\tau$). Moreover, we want that the borrowing and lending be equal on day $t$ so that there is no initial cash flow. In FRA, the same is achieved, by setting up the following portfolio at time $t$.



(a) Take a short position of one unit of $T$-maturity ZCB having face value 1.

(b) Take a long position by purchasing $\frac{B(t,T)}{B(t,T+\tau)}$ units of $T+\tau$-maturity ZCB having face value Rs 1.

The value of the portfolio at $t$ is $B(t,T) - \frac{B(t,T)}{B(t,T+\tau)}B(t,T+\tau) = 0$. At time $T$, close the short position of Rs 1 in $T$-maturity bond. At later time $T+\tau$, receive an amount $\frac{B(t,T)}{B(t,T+\tau)}$ from the long position in $T+\tau$-maturity ZCB. In other words, under no arbitrage condition, the discounted value of $\frac{B(t,T)}{B(t,T+\tau)}$ at time $T$ must be equal to Rs 1. The yield $Y$ that can explain this payment is the forward rate for $[T, T+\tau]$ prevailing at time $t$ and is given by

$$\frac{B(t,T)}{B(t,T+\tau)} = 1 * e^{((T+\tau)-T)f(t,T,T+\tau)}, \tag{11.4}$$

which means,

$$f(t,T,T+\tau) = -\frac{1}{\tau}(lnB(t,T+\tau) - lnB(t,T)).$$

We define instantaneous forward rate prevailing at time $t$ for investing at time $T$ as follows.

$$\begin{aligned} f(t,T) &= -\lim_{\tau \downarrow 0} \frac{lnB(t,T+\tau) - lnB(t,T)}{\tau} \\ &= -\frac{\partial}{\partial T} lnB(t,T). \end{aligned}$$

If we know $f(t,T)$ for all $0 < t \leq T$, we can easily recover $B(t,T)$ for all values of $0 < t \leq T$ as

$$B(t,T) = e^{-\int_t^T f(t,s)\,ds}.$$

**Definition 11.4.2** *The function $f(0,T)$, $T \geq 0$, of variable $T$, is called the* ***initial forward rate curve****, while the interest rate $f(t,t)$ is called* ***instantaneous short rate*** *that we can lock in at time $t$ to borrow at time $t$.*

Thus instantaneous short rate can be thought of a market rate at which the money can be borrowed for a very short duration which could be overnight charges to be repaid later. Suppose a bank in India requires an amount say $S$ to clear their liabilities due for next day, and presently it is not in a position to arrange for $S$ on its own. Then the bank can take an overnight loan $S_1$ from say RBI (central bank of India) at an instantaneous rate (viz., 1 day lending rate) offered by the RBI which makes $S_1$ grows to $S$ overnight. The bank can repay the amount $S_1$ along with its interest to RBI at some later date on the borrowing rate offered by RBI.

Now, if instead of continuously compounding forward interest rate, we take simple interest rate to explain the heretofore payment of FRA in (11.4). Let the simple interest rate be $L(t,T)$. Then, by the similar argument leading to (11.4), we have

$$\frac{B(t,T)}{B(t,T+\tau)} = 1 + \tau L(t,T),$$

equivalently,

$$L(t, T) = \frac{1}{\tau}\left(\frac{B(t, T)}{B(t, T+\tau)} - 1\right).$$

For $0 \leq t < T,\ L(t, T)$ is the *forward LIBOR*, the interest rate locked at time $t$ for investment over the time $[T, T+\tau]$. Also, $L(T, T)$ is called *spot LIBOR* (or simply LIBOR), and $\tau$ is *tenor* of LIBOR and is usually taken as 0.25 or 0.5 year.

Many large financial institutions trade with each other deposits for maturities ranging from just overnight to one year at a given currency. These are traded on market interest rates. The most commonly used market interest is LIBOR (London Interbank Offered Rate). The LIBOR is the rate at which financial institutions are willing to lend, on average (actually there are fifteen different LIBOR rates for fifteen maturities: overnight, one week, 2 weeks, one month, and so on). It is an average indicative quote of the interbank lending market. It is calculated by Thomson Reuters for ten currencies (including USD, AUD, GBP, DKK, EUR, CAD, JPY, to name a few), and published daily by the British Bankers Association. On the other hand, the LIBID (London Interbank Bid Rate) is the rate that these financial institutions are prepared to pay to borrow money, on average. Normally, LIBID $<$ LIBOR. The LIBOR is a fundamental point of reference to financial institutions. Moreover, many fixed income instruments like, forward rate agreements or mortgage rates, are indexed to the LIBOR. The following is a small example of EUR LIBOR interest rates having maturity 1 day

| Date | Rate (in %) |
|---|---|
| 9 Aug 2011 | 1.18875 |
| 10 Aug 2011 | 0.98750 |
| 11 Aug 2011 | 0.87250 |
| 12 Aug 2011 | 0.85250 |
| 15 Aug 2011 | 0.84500 |



When reference is made to the Indian interest rate this often refers to the MIBOR (Mumbai Inter-Bank Offered Rate) and MIBID (Mumbai Inter-Bank Bid Rate). The MIBOR was launched on June 15, 1998 by the Committee for the Development of the Debt Market, as an overnight rate. The National Stock Exchange of India launched the 14-day MIBOR on November 10, 1998, and the one month and three month MIBORs on December 1, 1998. Since the launch, MIBOR rates have been used as benchmark rates for the majority of money market deals made in India.

Besides bonds, there are various other interest rate derivative (debt) instruments in the market. Some are easy to understand and model while some others involves complex financial intricacies. The interest rate derivative market is enormous and it is not possible for us to cover all aspects of it in one chapter. Still we would like to briefly touch upon *interest rate swaps*, one of the largest and fastest growing derivative instruments. A "plain vanilla" interest rate swap is a contract between two parties, often called counter-parties, in which they agree to exchange their interest payments of two different kind on a predefined principal amount, on a periodic basis over the fixed time period. Typically payments made by one counter-party is based on a fixed interest rate for the term of the contract, while payments made by other counter-party is based on floating interest rate for the same term. In doing so, the principal amount is not physically exchanged rather interest payments are exchanged on notional principal. It is a convention to designate a fixed-rate counter-party as the buyer of the swap and the floating-rate counter-party is the seller of the swap. This type of contract is based on the needs and estimates of the level and changes in interest rates during the period of the swap contract.

For example, when we say a 3-year 8% fixed for six-month LIBOR floating Rs 10 lakh swap we mean a fixed-rate party is required to pay 8% fixed-rate interest on a notional principal of 10 lakh to a floating-rate party in exchange for a variable-rate interest that depends on a pre-specific six-month LIBOR rate on 10 lakh, and the transaction is to be settled every six months. Fig. 11.4 explains this kind of transaction.



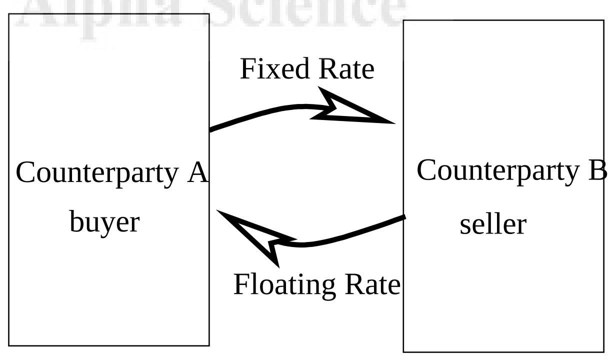



**Fig. 11.4.** Vanilla interest rate swap

**Example 11.4.4** *A swap buyer company A and the swap seller company B enter into a* 5*-year swap on January 10, 2001, on terms that company A pays fixed interest* 3.25% *annually on the notional principal of Rs* 10 *lakh and company B* *pays floating interest equal 3-month MIBOR annually on a notional principal of Rs 10 lakh. What is the cash flow between the parties.*

**Solution** The company $A$ has to pay 3.25% of Rs 1000000 each year to company $B$. Thus, $B$ will receive Rs 32500 periodically every year from $A$ for 5 years. The following table depicts the amount that will get exchanged between two counter-parties $A$ and $B$.

| Reset Date | Payment Date | 3-month MIBOR on Reset Date | Floating Payment by B to A | Net Flow |
|---|---|---|---|---|
| Jan 10, 2001 | Jan 10, 2002 | 4.564 | 45640 | 13140 |
| Jan 10, 2002 | Jan 10, 2003 | 3.475 | 34750 | 2250 |
| Jan 10, 2003 | Jan 10, 2004 | 3.485 | 34850 | 2350 |
| Jan 10, 2004 | Jan 10, 2005 | 2.22 | 22200 | -10300 |
| Jan 10, 2005 | Jan 10, 2006 | 2.314 | 23140 | -9360 |

The net cash flow figures shown in the last column above are expressed from company $A$'s point of view and indicate that company $A$ must pay company B on each of the last two payment dates. On first three payment dates, since the floating payment received by company $A$ exceeds the fixed payment, company $A$ will receive a net cash inflow on these dates, while in last two payments it is company $B$ that is benefitted . All payments are in rupees.



□

Another very basic interest rate derivative is *repo*. A repo (Repurchase agreement) is a way of borrowing against a collateral. Suppose a financial institution A borrows money from another financial institution B (usually banks or RBI; in context to follow we assume it to be RBI) to meet its short term needs by selling certain securities (like bonds) with an agreement that it will buy back the security at some fixed point in future time (the next day, after a week, etc.) at a predetermined price. It is equivalent to saying that A gets a loan against a collateral (the security) and pays an interest rate to B. The following figure captures this idea.

Repo rate is also called short term lending rate. When the repo rate increases borrowing from Reserve Bank of India (RBI) becomes more expensive. Therefore, we can say that in case, RBI wants to make it more expensive for the banks to borrow money, it increases the repo rate; similarly, if it wants to make it cheaper for banks to borrow money, it reduces the repo rate. If banks are short of funds they can borrow rupees from the RBI at the repo rate, the interest rate with a 1 day maturity. On the other hand, the reverse repo rate is the interest rate that

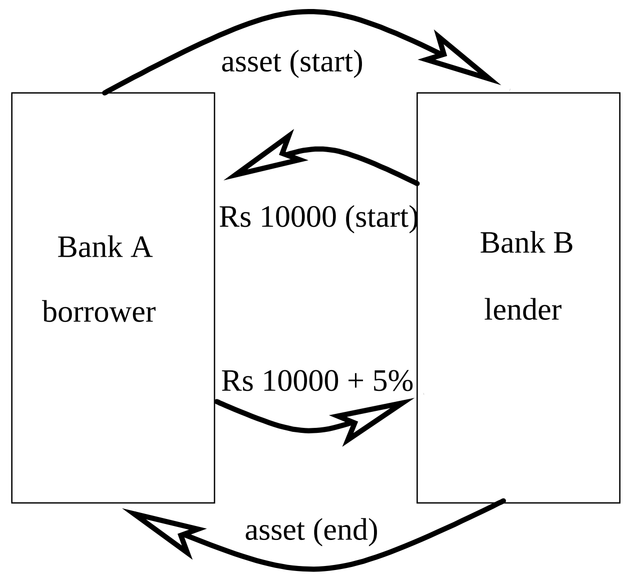



**Fig. 11.5.** Repo rate



banks receive if they deposit money with the RBI. This reverse repo rate is always lower than the repo rate. The RBI uses the reverse repo rate tool when it feels there is too much money floating in the banking system. Increases or decreases in the repo and reverse repo rate have an effect on the interest rate on banking products such as loans, mortgages and savings. An increase in the reverse repo rate means that the RBI will borrow money from the banks at a higher rate of interest. As a result, banks would prefer to keep their money with the RBI. To conclude the discussion on short-term interest rate, we would like to bring to the readers note a very common news line when the repo rate and/or reversed repo rate are changed: "The RBI today hiked short-term lending and borrowing rates sharply by 50 basis points". What we mean by one basis point is 1/100 of 1%, so, 50 basis point means 0.5% change in the existing short-term lending and borrowing interest rates.

## 11.5 Binomial Lattice Approach for Term Structure

We all agree that the lattice models (particularly the binomial lattice models studied in previous chapters) may not be good enough to capture the prevalent scenario exactly but they indeed provide a basic framework for understanding and building more sophisticated continuous-time models. A very natural question therefore, "can our most reliable friend, the binomial lattice, be again called to help us to model the short rate?" Luckily, the answer is YES.

We first partition the time horizon for which we intend to design short term interest rate into finite number of periods, like per day or per week etc. The lattice is drawn in a right angled triangle form; it is assumed that at time $t$, the two branches from any node are either in "up" state or in "flat" state. An index $i$ is used to denote how many ups have been taken to reach the node. Thus, each node of the lattice is indexed by a pair $(t, i)$, where $t$ is the time and $i$ is the node index at time $t$. Then, each node $(t, i)$ is assigned a short rate $r_{ti} \geq 0$. Fig. 11.6 shows a binomial lattice model say for 5-day short rate interest with a basic time span of 1 day. The time $t$ is shown at the bottom of the lattice while index $i$ denotes the ups (or height) of that node.

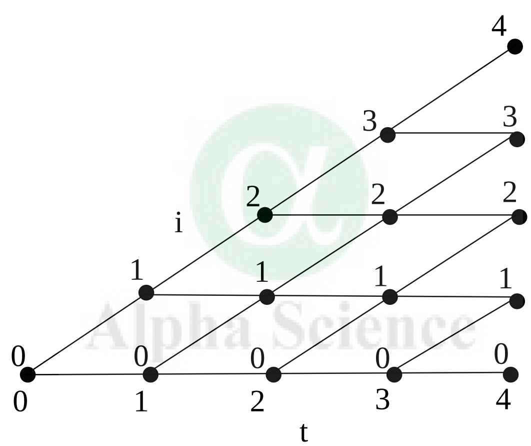



**Fig. 11.6.** Binomial lattice model for term structure

This lattice forms the basis for pricing interest rate securities by using the martingale pricing. For example, if $S_{ti}$ is the value of a non-dividend/coupon paying security at time $t$ and state $i$, then we insist that

$$S_{ti} = \frac{1}{1 + r_{ti}}(p_u S_{t+1,i+1} + p_d S_{t+1,i}), \tag{11.5}$$

where $p_d = 1 - p_u$ and $p_u$ is the risk neutral probability measure (RNPM). Since the probability $p_u$ is assigned rather than computed, here, it is a convention to take it $\frac{1}{2}$. Observe that such a model is arbitrage-free by construction.

If the security pays a coupon/dividend $C_{ti}$ at node $(t, i)$, then formula (11.5) should be taken as follows.

$$S_{ti} = \frac{1}{1 + r_{ti}}(p_u S_{t+1,i+1} + p_d S_{t+1,i}) + C_{ti}.$$

**Example 11.5.1** *Suppose the current short term rate is* 7% *per annum, and the up factor is* 1.2 *while the flat factor is* 0.8. *Construct a short term interest lattice for* 6 *years from now.*

**Solution** The lattice is described in Table 11.1.

| | | | | | |
|---|---|---|---|---|---|
| | | | | | 0.1742 |
| | | | | 0.1452 | 0.0116 |
| | | | 0.1210 | 0.0968 | 0.0774 |
| | | 0.1008 | 0.0806 | 0.0645 | 0.0516 |
| | 0.084 | 0.0672 | 0.0538 | 0.043 | 0.0344 |
| 0.07 | 0.056 | 0.0448 | 0.0358 | 0.0287 | 0.0229 |
| t=0 | t=1 | t=2 | t=3 | t=4 | t=5 |

**Table 11.1.** Short term interest rate lattice

□



**Example 11.5.2** *Assuming the data of Example 11.5.1, find the price of a ZCB maturing* 4*-years from now.*

**Solution** The term structure lattice is described in Table 11.1. To compute the bond price of a 5-year ZCB, we assign a face value Rs 1 to this bond at $t = 4$. Then we work backwards and use formula (11.5) with $r_{ti}$ read from Table 11.1.

| | | | | |
|---|---|---|---|---|
| | | | | 1.0000 |
| | | | 0.8921 | 1.0000 |
| | | 0.8255 | 0.9254 | 1.0000 |
| | 0.7858 | 0.8782 | 0.9490 | 1.0000 |
| 0.7642 | 0.8496 | 0.9161 | 0.9654 | 1.0000 |
| t=0 | t=1 | t=2 | t=3 | t=4 |

**Table 11.2.** 4 year ZCB price

For instance, the top entry in second column , i.e. node $(1, 1)$, is computed using $r_{11} = 0.084$ from Table 11.1 and formula (11.5), to get, $\frac{1}{1 + 0.084}\Big(\frac{1}{2}(0.8255) + \frac{1}{2}(0.8782)\Big) = 0.7858$.

□

The option contracts in which the underlying asset is a bond are called *bond options*. We have to realize that the price of a call bond option on bond increases and the price of a put bond option on bond decreases as the short-term interest rate rises (through the impact of short term interest rate on underlying bond price). There is no significant difference between the stock options and the bond options except that the underlying asset is the bond than the stock. Another characteristic difference to appreciate is that bonds are long term investment and because of a very low risk of default, they are more for an investment rather than speculation. This is why the bond options are traded on OTC (over the counter) basis, that is, traded between two private parties and are not listed on an exchange, unlike stock options which are highly speculative and generally traded in exchange markets. Despite this difference, the mathematics of bond option price is similar to the one we are already familiar with stock option pricing.

For instance, let us look at the put-call parity for European bond options. Consider a portfolio where we purchase one ZCB, take a short position (sell) on one European call bond option, and take a long position (buy) on one European put bond option, both bond options have same time to maturity $T$ and same strike price $K$.

At $t = 0$, the portfolio worth is $V(0) = B(0,0) - C(0,0) + P(0,0)$, where $B(0,0)$ denotes the current price of a ZCB, and $C(0,0)$, $P(0,0)$ are prices of a call option and a put option on the underlying ZCB at $t = 0$ respectively. At $T$, the value of this portfolio is

$$V(T) = \begin{cases} B(0,T) + K - B(0,T)\,, & \text{if } B(0,T) < K \\ B(0,T) - (B(0,T) - K)\,, & \text{if } B(0,T) \geq K. \end{cases}$$

Thus, no matter what is the state, the portfolio is worth $K$ at time of expiration. With no arbitrage in force, the payoff from the portfolio is risk-free, and we can discount its value at the spot rate $Y(0,T)$, YTM of a $T$-ZCB, to get $B(0,0) - C(0,0) + P(0,0) = Ke^{-Y(0,T)T}$, which is similar to the put-call parity for stock options.

Let us investigate pricing of bond option through some simple examples.

**Example 11.5.3** *Compute the price of a European call option on the ZCB of Example 11.5.2 that expires in* 3 *years and has strike price Rs* 93.

**Solution** At time of expiry $t = 3$, $K = 93$, hence $C(i,3) = \text{Max}\{100\,B(i,3) - K, 0\}$, $i = 0,1,2,3$. The same is depicted in the last column in Table 11.3. Thereafter the call price is computed iterating backward and using (11.5).

| | | | |
|---|---|---|---|
| | | | 0 |
| | | 0 | 0 |
| | 0.4102 | 0.8894 | 1.8983 |
| 0.9643 | 1.6534 | 2.6025 | 3.54 |
| t=0 | t=1 | t=2 | t=3 |

**Table 11.3.** 3-year European call bond option price

The European call price is Rs $0.9643$ on a ZCB governed by short term interest rates given in Table 11.1.

□

**Example 11.5.4** *Compute the price of a European put option on the ZCB of Example 11.5.2 that expires in* $3$ *years and has strike price Rs* $93$*.*

**Solution** We shall work out the put price using Table 11.2 and formula (11.5), working backward. The Table 11.4 depicts the complete working of the same.

| | | | |
|---|---|---|---|
| | | | 3.7908 |
| | | 1.9318 | 0.4622 |
| | 0.9909 | 0.2166 | 0 |
| 0.511 | 0.1025 | 0 | 0 |
| t=0 | t=1 | t=2 | t=3 |

**Table 11.4.** 3-year European put bond option price

□

By now we have realized that the term structure dynamics is characterized by the evolution of the short term interest rates. In the sections to follow we shall be describing two basic models which capture the dynamics of short term interest rates and consequently the bond pricing.

## 11.6 Vasicek Model

Vasiceks [143] pioneering work in 1977 is the first account of a bond pricing model that incorporates stochastic interest rate. The short interest rate or instantaneous rate (simply called short rate from now onwards) dynamics is modeled as a diffusion process with constant parameters. Remember we have already discussed that the *short-term interest rate* is the shortest maturity yield or perhaps the overnight rate offered by the market, that is, $f(t,t)$ which is same as $Y(t)$.

Vasiceks model is a special version of Ornstein-Uhlenbeck (O-U) process, with constant volatility. In fact the O-U process (named after Leonard Ornstein and George Eugene Uhlenbeck) is a stochastic process that describes the velocity of a massive Brownian particle under the influence of friction. Over time, the process tends to drift towards its long-term mean. Such a process is thus called *mean-reverting* .

We shall be denoting the short rate stochastic process by $\{r(t),\ t \geq 0\}$ to keep unanimity with the existing literature in this context. The short rate is modeled through a stochastic differential equation (SDE) involving three parameters. The model assumes that $r(t)$ follows the following SDE.

$$dr(t) = (\alpha - \beta r(t))dt + \sigma dW(t), \tag{11.6}$$

where $\alpha,\ \beta,\ \sigma$ are positive constants and $W(t)$ is a Wiener process.

This model is very tractable, and there are explicit solutions for a number of derivatives based on it. In order to find the distribution of $r(t)$ in (11.6) we apply the Ito Lemma. Consider the function $g(t,r) = e^{\beta t}r$. Apply Ito Lemma, we get

$$\begin{aligned} dg &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial r}dr + \frac{1}{2}\frac{\partial^2 g}{\partial^2 r}(dr)^2 \\ &= \beta e^{\beta t} r(t)dt + e^{\beta t}dr(t). \end{aligned}$$

Since the right hand side is independent of $g$, on stochastic integration equivalent, we get



$$\begin{aligned} g(t,r) - g(r,0) &= \int_0^t \beta e^{\beta s} r(s)ds + \int_0^t e^{\beta s}((\alpha - \beta r(s))dt + \sigma dW(s)) \\ &= \alpha \int_0^t e^{\beta s}ds + \sigma \int_0^t e^{\beta s}dW(s) \\ &= \frac{\alpha}{\beta}(e^{\beta t} - 1) + \sigma \int_0^t e^{\beta s}dW(s), \end{aligned}$$

where the name of the integration variable has been changed from $t$ to $s$ to avoid confusions.

Remembering that $g(t,r) = e^{\beta t}r$, the above relation yields

$$E(g(t,r)) = E(e^{\beta t}r) = r(0) + \frac{\alpha}{\beta}(e^{\beta t} - 1), \tag{11.7}$$

where in computing (11.7) we have use the fact that $\int_0^t e^{\beta s}dW(s)$, being an Ito integral with non-random integrand, so a martingale, consequently its expected value is zero.

Moreover, as we have

$$r(t) = r(0)e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}) + \sigma e^{-\beta t} \int_0^t e^{\beta s} dW(s), \tag{11.8}$$

its variance is given by

$$Var(r(t)) = \sigma^2 e^{-2\beta t} \int_0^t e^{2\beta s} ds.$$

The immediate above relation needs a bit of explanation. Recall that from the definition of the Brownian motion the increments $\triangle W_i = W(t_{i+1}) - W(t_i)$ have variance $t_{i+1} - t_i$. So, for $I = \int_0^t f(s)dW(s) \simeq \sum_i f(t_i)\triangle W_i$, the variance is, $Var(I) = \sum_i f^2(t_i)Var(\triangle W_i) = \sum_i f^2(t_i)(t_{i+1} - t_i)$, implying, $Var(I) = \int_0^t f^2(s)ds$. We have used this fact in computing $Var(r(t))$ from (11.8). Next on simplifying the $Var(r(t))$ relation, we get

$$Var(r(t)) = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}). \tag{11.9}$$

Moreover, as the increments of the Brownian motion are independent and normally distributed, so from (11.9), $r(t)$ is also normally distributed.

The above discussion can be summarized to explicitly state the distribution of $r(t)$ as follows.

$$r(t) \sim N\Big(r(0)e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}),\ \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\Big).$$

Observe that when $t \to \infty$, the $E(r(t)) \to \dfrac{\alpha}{\beta}$ while $Var(r(t)) \to \dfrac{\sigma^2}{2\beta}$. Hence, in this case, $r(t) \sim N\Big(\dfrac{\alpha}{\beta}, \dfrac{\sigma^2}{2\beta}\Big)$. Observe that the variance of the short rate converges to a finite value in contrast to the case of Brownian motion.

We now provide an interpretation of the term $\dfrac{\alpha}{\beta}$. Recall the Vasicek model SDE

$$\begin{aligned} dr(t) &= (\alpha - \beta r(t))dt + \sigma dW(t) \\ &= \beta\Big(\frac{\alpha}{\beta} - r(t)\Big)dt + \sigma dW(t). \end{aligned}$$

(i) If $r(t) = \dfrac{\alpha}{\beta}$, then the *drift* (term associated with $dt$) is zero.

(ii)If $r(t) > \dfrac{\alpha}{\beta}$, then the drift is negative. The process will try to pull up $r(t)$.

(iii)If $r(t) < \dfrac{\alpha}{\beta}$, then the drift is positive. The process will try to pull down $r(t)$.

Hence the drift is always directed to $\dfrac{\alpha}{\beta}$, that may thus be interpreted as a long run mean of the short rate $r(t)$. The parameter $\beta$ represents then the *strength* of this mean reversion.

Let us see what happens to the distribution of $r(t)$ when the volatility $\sigma = 0$. In that case the Vasicek model SDE (11.6) reduces to

$$dr(t) = (\alpha - \beta r(t))dt,$$

that is,

$$\frac{dr}{\frac{\alpha}{\beta} - r} = \beta dt.$$

The solution is

$$r(t) = \frac{\alpha}{\beta} + \left(r(0) - \frac{\alpha}{\beta}\right)e^{-\beta t}.$$

Again observe that when $t \to \infty$, $r(t) \to \dfrac{\alpha}{\beta}$ from below when $r(0) < \dfrac{\alpha}{\beta}$ and from above when $r(0) > \dfrac{\alpha}{\beta}$. Thereby depicting the mean reverting nature of the short interest rate $\{r(t),\ t \geq 0\}$ process.

An obvious limit of this model is that $r(t)$, having a normal distribution, can always assume negative values with positive probability. Although practically, on calibration of the Vasieck model, it is found that the probability of $r(t)$ taking on negative values is negligible, yet it could not be completely ruled out. The Vasicek model is not realistic enough because the short interest rate $r(t) < 0$ with a positive probability is ambiguous.

## 11.7 Cox-Ingersoll-Ross Model

The Cox-Ingersoll-Ross (CIR) model [34] is another well-known mathematical model which describes interest rate movements driven by only one source of market risk. It was introduced in 1985 as an extension of the Vasicek model. The dynamic of the short interest rate process, $\{r(t), t \geq 0\}$, is described by the following SDE.

$$dr(t) = (\alpha - \beta r(t))dt + \sigma \sqrt{r(t)}dW(t), \tag{11.10}$$

where $\alpha, \beta, \sigma$ are positive constants, $W(t)$ is a Wiener process which models the random market risk factor. The parameter $\sigma$ determines the volatility of the interest rate.

**Example 11.7.1** *Assume a particular short interest rate follows the following CIR model*

$$dr(t) = 0.22(0.06 - r(t))dt + 0.45 \sqrt{r(t)}dW(t).$$

*At some particular time* $t$, $r(t) = 0.05$, *and then* $r(t)$ *suddenly becomes* $0.02$. *What is the resulting change in the volatility?*

**Solution** The volatility in the CIR model is $\sigma \sqrt{r(t)}$. For $\sigma = 0.45$ and $r(t) = 0.05$, the volatility is $0.45 \sqrt{0.05}$, while for $r = 0.02$, the volatility is $0.45 \sqrt{0.02}$. The change in volatility is thus $0.45 \sqrt{0.02} - 0.443 \sqrt{0.05} = -0.09426$.

□

Unlike the Vasicek model, the CIR model is not Gaussian and is therefore considerably more difficult to analyze. Furthermore, it does not have the closed form solution like the one (11.8) we have for the Vasicek model. But still we can find the distribution of $r(t)$. For this text we restrict ourselves to find only the first two moments of $r(t)$ by applying the Ito Lemma.

Let $g(t, r) = e^{\beta t} r$. From Ito Lemma, we have

$$\begin{aligned} dg &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial r}dr + \frac{1}{2}\frac{\partial^2 g}{\partial^2 r}(dr)^2 \\ &= \beta e^{\beta t} r(t)dt + e^{\beta t}dr \\ &= \beta e^{\beta t}dt + e^{\beta t}((\alpha - \beta r(t))dt + \sigma \sqrt{r(t)}dW(t)) \\ &= \alpha e^{\beta t}dt + \sigma \sqrt{r(t)}dW(t). \end{aligned}$$

An equivalent integration form is

$$\begin{aligned} e^{\beta t} r(t) &= r(0) + \alpha \int_0^t e^{\beta s} ds + \sigma \int_0^t e^{\beta s} \sqrt{r(s)}dW(s) \\ &= r(0) + \frac{\alpha}{\beta}(e^{\beta t} - 1) + \sigma \int_0^t e^{\beta s} \sqrt{r(s)}dW(s). \end{aligned} \tag{11.11}$$

Though (11.11) looks very similar to the one we obtain in (11.8) for the Vasicek model, yet there is a noticeable difference. Observe that the term $\sqrt{r(s)}$ appears in the right hand side Ito integral in (11.11). Thus we are not able to obtain the closed form expression for $r(t)$. We need to restore to some numerical techniques for computing $r(t)$ value at given $t$.

Going back to our discussion, we compute the first moment, that is expectation of $r(t)$, from (11.11). Recall that the expectation of an Ito integral is zero, we get

$$E(e^{\beta t}r(t)) = r(0) + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right), \tag{11.12}$$

equivalently

$$E(r(t)) = r(0)e^{-\beta t} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right).$$

Note that the $E(r(t))$ in the CIR model is the same as the $E(r(t))$ in the Vasicek model.

Next we wish to compute the variance $Var(r(t))$. For this, let $h(t,r) = (g(t,r))^2 = e^{2\beta t}r^2$. Applying the Ito Lemma on $h(t,r)$, we have



$$\begin{aligned} dh &= \frac{\partial h}{\partial t}dt + \frac{\partial h}{\partial r}dr + \frac{1}{2}\frac{\partial^2 h}{\partial^2 r}(dr)^2 \\ &= 2\beta e^{2\beta t}r^2(t)dt + 2e^{2\beta t}r(t)dr + e^{2\beta t}(dr)^2 \\ &= 2\beta e^{2\beta t}r^2(t)dt + 2e^{2\beta t}r(t)((\alpha - \beta r(t))dt + \sigma\sqrt{r(t)}dW(t)) + e^{2\beta t}\sigma^2 r(t)dt, \end{aligned}$$

where the last term in the above equality follows on account of $dW(t)dW(t) = dt,\ dW(t)dt = 0,\ dtdt = 0$. Simplifying the above relation, we get

$$\begin{aligned} dh &= (2\alpha + \sigma^2)e^{2\beta t}r(t)dt + 2\sigma e^{2\beta t}(r(t))^{3/2}dW(t) \\ &= (2\alpha + \sigma^2)e^{\beta t}g(t,r)dt + 2\sigma e^{\frac{\beta t}{2}}(g(t,r))^{3/2}dW(t), \end{aligned}$$

which on equivalent integration form yields

$$h(t,r) = r^2(0) + (2\alpha + \sigma^2)\int_0^t e^{\beta s}g(s,r(s))ds + 2\sigma\int_0^t e^{\frac{\beta s}{2}}(g(s,r(s)))^{3/2}dW(s).$$

Taking expectation, and using the fact that the expectation of the Ito integral is zero, we get

$$\begin{aligned}E(h(t,r)) &= r^2(0) + (2\alpha+\sigma^2)\int_0^t e^{\beta s}E(g(s,r))ds \\ &= r^2(0) + (2\alpha+\sigma^2)\int_0^t e^{\beta s}(r(0)+\frac{\alpha}{\beta}(e^{\beta s}-1))ds \\ &= r^2(0) + \frac{2\alpha+\sigma^2}{\beta}(r(0)-\frac{\alpha}{\beta})(e^{\beta t}-1) + \frac{(2\alpha+\sigma^2)\alpha}{2\beta^2}(e^{2\beta t}-1).\end{aligned}$$

We have used (11.12) for $E(g(s,r))$ in the second last equality. Therefore

$$E(r^2(t)) = e^{-2\beta t}r^2(0) + \frac{2\alpha+\sigma^2}{\beta}(r(0)-\frac{\alpha}{\beta})(e^{-\beta t}-e^{-2\beta t}) + \frac{(2\alpha+\sigma^2)\alpha}{2\beta^2}(1-e^{-2\beta t}).$$

To complete the discussion

$$\begin{aligned}Var(r(t)) &= E(r^2(t)) - (E(r(t)))^2 \\ &= \frac{\sigma^2}{\beta}r(0)(e^{-\beta t}-e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta^2}(1-2e^{-\beta t}+e^{-2\beta t}).\end{aligned}$$

Observe that when $t \to \infty$, then $E(r(t)) \to \dfrac{\alpha}{\beta}$ and $Var(r(t)) \to \dfrac{\alpha\sigma^2}{2\beta^2}$.

The CIR model has a mean reverting property just like Vasicek model. Here, $r(t) \geq 0$ almost surely, and as soon as $r(t) = 0$ the model SDE (11.10) reduces to a pure positive drift $\alpha dt$ with zero volatility. This forces the interest rate back into the positive zone. It is more instructive to compare the CIR model to the Vasicek model. The following matters stand out in particular.

(i) In the Vasicek model, interest rates can be negative. In the CIR model, negative interest rates are impossible.

(ii) In the Vasicek model, the volatility (the term associated with $dW(t)$ in the SDE) of the short-term interest rate is a constant. In the CIR model, the volatility of the short-term interest rate increases as the short-term interest rate increases.

(iii) As in the Vasicek model, the short-term interest rate in the CIR model exhibits mean reversion with the same mean $\dfrac{\alpha}{\beta}$.

## 11.8 Partial Differential Equation

The Vasicek model and the CIR model discussed above are examples of simplest fixed income market models that begin with the SDE of the type

$$dr(t) = \delta(t, r(t))dt + \gamma(t, r(t))dW(t). \tag{11.13}$$

Since $r(t)$ is driven by only single SDE, these models are refer to *one factor short rate models*.

Let the short rate $r(t)$ be driven by one factor model under risk neutral measure $\widetilde{P}$ as follows.

$$dr(t) = \delta(t, r(t))dt + \gamma(t, r(t))d\widetilde{W}(t), \tag{11.14}$$

where $\widetilde{W}(t)$ is the Brownian motion under $\widetilde{P}$.

Since $r(t)$ is a solution of SDE (11.14), it must be a Markov process. This implies that the bond price at time $t$ of a ZCB with maturity time $T$ and face value 1 can be written as some function of $r(t)$ and $t$. That is,

$$B(t, T) = \widetilde{E}(e^{-\int_t^T r(u)du}/\mathcal{F}_t) = g(t, r),$$

for some function $g(t, r)$ of two variables $t$ and $r$, and filtration $\{\mathcal{F}_t, t \geq 0\}$ to which the process $\{r(t),\ t \geq 0\}$ is adapted.

In trading on these bonds, one wants to be sure that there are no arbitrage opportunity. According to the first fundamental theorem of asset pricing (see, Chapter 3), for no arbitrage, a risk neutral probability measure $\widetilde{P}$ must exist such that the discounted bond prices $D(t)B(t, T)$ are martingale under $\widetilde{P}$, where $\{D(t),\ t \geq 0\}$ is a discount process given by $e^{-\int_0^t r(s)ds}$. We apply the following three standard steps of the Feynman-Kac theorem (Chapter 9) to determine the partial differential equation (PDE) describing the dynamics of ZCB price.

**Step 1.** $D(t)B(t, T) = D(t)g(t, r)$ is a martingale.

**Step 2.** Observe that

$$\begin{aligned} d(D(t)B(t, T)) &= -r(t)D(t)g(t, r)dt + D(t)dg(t, r) \\ &= D(t)(-r(t)g + g_t + \delta(t, r(t))g_r + \frac{1}{2}\gamma^2(t, r(t))g_{rr})dt \\ &\quad +D(t)\gamma(t, r(t))g_r d\widetilde{W}(t). \end{aligned}$$

**Step 3.** Set $dt$ coefficient equal to zero.

Thus we get that the bond price $B(t, T) = g(t, r)$ is a solution of the following PDE.

$$g_t(t, r) + \delta(t, r)g_r(t, r) + \frac{1}{2}\gamma^2(t, r)g_{rr}(t, r) = rg(t, r), \tag{11.15}$$

with the terminal condition $B(T, T) = g(T, r) = 1$, for all $r$.

We present below two simplest scenarios when the bond price can be computed explicitly.

**Example 11.8.1** *Find the value of a ZCB when the short rate is governed by the SDE* $dr = 0$.

**Solution** As discussed above, the value of the ZCB $B(t, T) = g(t, r)$, is the solution of (11.15). Since $dr = 0$ implies $r$ is a constant, (11.15) reduces to $g_t = rg$. A solution is given by

$$ln(g(t, r)) = c + rt,$$

where $c$ is a constant. Using the boundary condition $g(T, r) = 1$ yields $c = -rT$. Thus, $B(t, T) = e^{-r(T-t)}$, which is in agreement with our knowledge about bond value when interest rate is a constant.

□

The next example is a particular case of Ho-Lee model in which the short term interest rate can be described by the SDE

$$dr = \theta(t)dt + \sigma d\widetilde{W},$$

for some appropriate function $\theta(t)$. Comparing the above expression with (11.14), we have $\delta(t, r(t)) = \theta(t)$ and $\gamma(t, r(t)) = \sigma$.

**Example 11.8.2** *Find the value of a ZCB when the short rate is governed by the SDE*

$$dr = adt + \sigma d\widetilde{W}, \quad a \text{ is a constant.}$$



**Solution** The given SDE is similar to the Ho-Lee model SDE with $\theta(t) = a$. Thus, $\delta(t, r(t)) = a$ and $\gamma(t, r(t)) = \sigma$. The bond value of a ZCB is $B(t, T) = g(t, r)$ is a solution of (11.15); which, in the considered case, becomes

$$g_t + ag_r + \frac{1}{2}\sigma^2 g_{rr} = rg.$$

Assume that $g(t, r) = e^{-r(T-t)}A(t, T)$ is a solution of the above PDE. Then, by means of classical calculus, we have

$$(rA + A_t) - aA(T - t) + \frac{1}{2}\sigma^2 A(T - t)^2 = rA,$$

that is,

$$dlnA = ((T - t)a - \frac{1}{2}(T - t)^2\sigma^2)dt.$$

Integrating and thereafter using the boundary condition that $B(T, T) = g(T, r) = A(T, T) = 1$, we get the constant of integration zero. Consequently,

$$lnA(t,T) = -\frac{(T-t)^2}{2}a + \frac{1}{6}(T-t)^3\sigma^2.$$

We thus have an explicit formula for $B(t,T) = e^{-r(T-t)}A(t,T)$, where $A(t,T)$ can be computed using an immediate above expression.

□

We now go back to the interest rate models studied in sections 11.6 and 11.7.

Recall the Vasicek model. We want to find the bond price of a ZCB with face value 1 and maturity time $T$ when the dynamics of the market spot rate $r(t)$ is governed by the Vasicek model SDE (11.6). For this we need to solve the PDE (11.15) under Vasicek model. Comparing (11.6) with (11.13), we have, $\delta(t, r(t)) = \alpha - \beta r(t)$ and $\gamma(t, r(t)) = \sigma$. We assume that a solution of (11.15) is of the form

$$g(t,r) = e^{-rC(t,T)+A(t,T)}, \ t \in [0,T],$$

where $A(t,T)$ and $C(t,T)$ are non-random functions to be determined; although we skip the details as to how we guess this form of solution. Then,

$$g_t = (A_t - rC_t)g\,, \quad g_r = -Cg\,, \quad g_{rr} = C^2 g.$$

Note that, for notational convenience, we have avoided writing the variables in all functions. Thus, from (11.15), we get



$$(A_t - rC_t - (\alpha - \beta r)C + \frac{1}{2}\sigma^2 C^2 - r)g = 0.$$

For this to hold for all values of $r$, we must have

$$C_t = \beta C - 1$$
$$A_t = \alpha C - \frac{1}{2}\sigma^2 C^2.$$

Also, the terminal condition $g(T,r) = 1$ in (11.15) reduces to $A(T,T) = 0,\ C(T,T) = 0$. It is easy to solve the first equation to get

$$C(t,T) = \frac{1}{\beta} + ke^{\beta t}.$$

Using the terminal condition $C(T,T) = 0$, we get $k = -\dfrac{e^{-\beta T}}{\beta}$. Thus,

$$C(t,T) = \frac{1}{\beta}(1 - e^{-\beta(T-t)}).$$

Substituting the function $C(t,T)$ in the second equation and thereafter solving the resultant first order PDE with terminal condition $A(T,T) = 0$, we can get the function $A(t,T)$. After some work (which we leave for the readers to complete) we can see that

$$A(t,T) = \frac{(\frac{1}{2}\sigma^2 - \alpha\beta)((T-t) - C)}{\beta^2} - \frac{\sigma^2 C^2}{4\beta}.$$

Once the bond price $B(t,T)$ is known then the bond yield $Y(t,T)$ can be determined by using

$$\begin{aligned} Y(t,T) &= -\frac{1}{T-t} ln(B(t,T)) \\ &= \frac{1}{T-t}(r(t)C(t,T) - A(t,T)). \end{aligned}$$

Next, suppose we wish to find the bond price $B(t,T)$ when the spot rate $r(t)$ follows the dynamics of the CIR model. Recall the CIR model SDE (11.10). Comparing it with (11.13), we have, $\delta(t, r(t)) = \alpha - \beta r(t)$ and $\gamma(t, r(t)) = \sigma\sqrt{r(t)}$.

Again we assume that the solution of the PDE (11.15) is of the form

$$g(t,r) = e^{-rC(t,T) + A(t,T)},\ t \in [0,T],$$

where $A(t,T)$ and $C(t,T)$ are non-random functions to be determined. In CIR model case, the PDE (11.15) becomes

$$(A_t - \alpha C + (-C_t + \beta C + \frac{1}{2}\sigma^2 C^2 - 1)r)g = 0.$$

The above equation to hold for all $r$ is possible only when the following hold.

$$\begin{aligned} C_t &= \frac{1}{2}\sigma^2 C^2 + \beta C - 1 \\ A_t &= \alpha C. \end{aligned}$$

The first of the two equations is a Riccati equation and can easily be solved by a standard technique to get $C(t,T)$. Once through with it, we can use the $C(t,T)$ in the second equation to solve it for $A(t,T)$. In this solution procedure we also need to use the terminal conditions $A(T,T) = 0$ and $C(T,T) = 0$. Moreover, the bond yield is given by $Y(t,T) = \frac{1}{T-t}(r(t)C(t,T) - A(t,T))$.

For both the Vasicek model and the CIR model, the ZCB yield is an affine function of short rate $r(t)$. Such models are therefore also called the *affine yield models*.

## 11.9 Summary and Additional Notes

- Though we have not covered all types of bonds in the chapter but some other important ones we would like to take note in brief. One such bond is called a *callable bond* which is a fixed rate bond where the issuer has the right but not the obligation to repay the face value of the security at a pre-agreed value prior to the final original maturity of the security. Callable bonds can be viewed as a combination of noncallable bonds and an option to call a bond. The writer of the call option is the holder of the bond and the buyer of the call is the stockholder of the issuing corporation. The call option gives the issuer the right to call the bond at a fixed strike price any time before bond maturity, after an initial protection period. There is a good literature available on the pricing of such bonds. We cite here few of them [16, 54] for readers to explore. Opposite of these bonds are *puttable bond*. The other types of bonds are floating rate bonds, index linked bonds, foreign bonds or sovereign bonds, perpetual bonds, to name a few.
- We all can note that most of the term structure operations and the dynamic we've carried out used the prices of zero coupon bonds as inputs. However, in reality, vast majority of traded bonds do have coupons. We have seen in Section 11.2 that the pricing of the coupon bonds can be done through the pricing of ZCBs. It is the converse that can interest us that how can we infer ZCBs prices (or yields) from coupon bonds prices (or yields). The method generally used to achieve this is called *bootstrapping*. Some textbooks, like [120], explain this method. We also refer to a good paper [35] to know more about the method and its technicalities.
- We have talked about the plain vanilla interest rate swaps in Section 11.5. However, there are other types of swaps which could be of some interest to readers, like currency swaps and commodity swaps. These swaps help the companies to hedge against the risk of unanticipated interest rates in the market. For instance, if a buyer company is engaged in a commodity (say oil) swap over a swap tenor on a predetermined notional principal, then, despite the changes in oil price and forward interest rate on it, the company has to pay fixed rate to the other floating rate counter-party as per the terms of the swap, over the period of the swap. Currency swaps also has a fairly large market. For more on swaps and their economics, we refer the readers to [46, 87, 127, 133].
- Researchers around the world have made efforts on determining the dynamics of interest rates [60, 66, 72, 142]. But no single model appears to be valid in all markets across the world. Besides the two first generation models that we had

discussed in Sections 11.6-11.8, there are several other term structure models studied in the literature under different set ups. Prominent among them are the binomial lattice model of Black-Derman-Toy [13] and a continuous-time Heath-Jarrow-Morton (HJM) model [61]. The HJM model evolves the yield curve in terms of the forward rates. This model is significant so much so that every term structure model driven by the Brownian motion is the HJM model including the Vasicek model and the CIR model. For details on how the model dynamics and its evaluation can be described we urge the interested readers to go through Chapter 10 in [121]. The continuous-time models of Black-Dermon-Toy and the Black-Karasinski [14] instead assume that $lnr(t)$ is normally distributed. The theory of one-factor models has also been extended to two-factor model and multi-factor models which involve several factors, like short term rate, long term rate, etc. The main difficulties in the latter models lies in their calibration which requires lot of data to determine all the parameters in the model under consideration. There are several other models, like Heston model [62], which assume the stochastic volatility, besides stochastic short-term returns, governed by two SDEs. Certain studies are available on www which try to create appropriate model for a specific market of a specific country. The literature in the context is extremely vast to be quoted here but we encourage readers to take a look at some excellent texts and web pages on interest rate models.

- Another significant point to bring forth is the calibration of the interest rate models. Calibration of interest rate models under the risk neutral measure typically entails the availability of some derivatives such as swaps, caps or swaptions. The primary tool used to estimate the parameters in a system of stochastic differential equations are Bayesian method, Cholesky decomposition, principal component analysis, Monte Carlo simulation, various regression techniques, interpolations and splines, generating function, and machine learning. We refer to few texts [69, 142, 126] in this regard.

## 11.10 Exercises

**Exercise 11.1** *Find the spot rate for a 1-year ZCB trading at Rs 92 with a face value of Rs 100.*

**Exercise 11.2** *A 1-year ZCB with face value Rs 100 is currently selling for Rs 85. What is the interest rate after 6 months if the investment for 6 months in zero coupon bonds gives a continuously compounding annual return of* 12%.
*(Hint: Find* $y(0.5)$ *using* $B(0,1)\,e^{r} = Fe^{-y(0.5)/12}$*).*

**Exercise 11.3** *A coupon bond with a face value Rs 100 makes coupon payments of Rs 1.50 every three months. What is the coupon rate?*

**Exercise 11.4** *What asset is consider to be riskier, a 3-month ZCB bond or a 30-year coupon bond maturing in 3 months?*

**Exercise 11.5** *If the 6-month spot rate is* 3% *and the 1-year spot rate is* 5%*, then find the price of a 1-year bond with a* 8% *annual coupon rate payable semiannually and a face value Rs 100.*
*(Hint: Use* $B(0,1) = Fce^{-r_1} + (Fc + F)e^{-2r_2}$ *and note* $c = 0.04$*).*

**Exercise 11.6** *You intend to purchase a 10-year, Rs 100 face value bond that pays Rs 5 coupon every 6 months. If the required continuous compounding return is* 5% *per annum, how much should you be willing to pay for the bond ?*
*(Hint: Example 11.2.6 with coupon value).*

**Exercise 11.7** *Compute the price of a 5-year coupon bond with an* 6% *annual coupon payable semiannually, the annual bond yield is* 8%*, and the face value of bond is Rs 100.*
*(Hint: Example 11.3.4 with* $n = 10$*,* $c = 0.03$*,* $Y_m = 0.04$*).*

**Exercise 11.8** *What is the yield to maturity on a 10-year ZCB with a face value of Rs 100 that is selling for Rs 65? (Hint: Use* $B(0,T) = Fe^{-TY_m}$ *to compute YTM* $Y_m$*).*



**Exercise 11.9** *A bond with 20 years remaining to maturity is selling for Rs 120 has a* 9% *annual coupon paid semiannually. If the face value of the bond is Rs 100, calculate the bond yield to maturity.*

**Exercise 11.10** *Calculate the YTM on a 1-year coupon bond with annual coupon Rs 10 and the face value Rs 100. The bond is selling for Rs 95.*
*(Hint: Example 11.3.5 with coupon rate replace by coupon value).*

**Exercise 11.11** *Which among the following security has the higher YTM?*

*(a)A 1-year ZCB with a face value of Rs 100 that is selling at Rs 90.*
*(b)A 10-year bond with a* 5% *coupon and a face value of Rs 100 that is selling at par.*

*(Note: A bond selling at par means that it is selling at full face value).*

**Exercise 11.12** *Compute the forward rate when the 6-month spot rate is* 3%*, and the 1-year spot rate is* 5%*.*

**Exercise 11.13** *A 3-month ZCB is selling for Rs 95, a 6-month ZCB is selling for Rs 92, a 9-month ZCB is selling for Rs 88 and a 1-year ZCB selling for Rs 85. All four bonds have a face value of Rs 100. Compute the spot rates for the four periods. Also, compute the second, third and fourth periods forward rates. (Hint: Use Definition 11.3.1 with* $F = 100$ *and then (11.3)).*

**Exercise 11.14** *For ZCBs, each bond with face value Rs 1, the following bond prices per annum are observed.*

| B(0,1) | B(0,2) | B(0,3) | B(0,4) | B(0,5) |
|---|---|---|---|---|
| 0.9654 | 0.9173 | 0.8735 | 0.8115 | 0.7855 |

*For each maturity year, compute the bond yields and the 1-year implied forward rate.*

**Exercise 11.15** *The ZCB continuously compounded annual yields are observed as follows.*

| Y(0,1) | Y(0,2) | Y(0,3) | Y(0,4) | Y(0,5) |
|---|---|---|---|---|
| 0.03 | 0.035 | 0.04 | 0.045 | 0.05 |

*For each maturity year, compute the ZCB prices and and the 1-year implied forward rate. What are the simple rate annual yields on the bonds? (Hint: Simple rate annual yield* $y(0,T)$ *should be such that* $1 + y(0,T) = e^{Y(0,T)}$*).*

**Exercise 11.16** *Suppose a loan of Rs 100000 will be taken 3 months from now (date of borrow). It is expected that the loan will be repaid 1 year from now (repayment date). An FRA is designed with guaranteed* $6\%$ *annual simple interest rate for 1 year on Rs 100000. The actual interest rate is* $4\%$ *on the date of borrow. Determine the settlement of the FRA if the settlement occurs on the date the loan is (i) borrowed, (ii) repaid. (Hint: If the FRA is settled at the time the money is borrowed, payments will be less than when the same is settled on the date of repayment because the borrower has time to earn interest on the FRA settlement.)*

**Exercise 11.17** *Suppose that the risk neutral probabilities are equal to* $0.5$ *in every state. If the short-term rates (in percentage per month) are described by the following table, then find the prices of a 3-month ZCB with face value Rs 100 (take a one month step).*

| | | |
|---|---|---|
| | | 0.0101 |
| | 0.0098 | 0.0091 |
| 0.0095 | 0.0088 | 0.0081 |

**Exercise 11.18** *A discrete-time model is used to model both the price of a non-dividend paying stock and the short-term interest rate. The stock is selling for Rs 100 and the annual simple interest rate is* 5%. *After 1 year, only two states, upstate and downstate, are observed in the economy. The stock prices are Rs 110 (upstate) and Rs 95 (downstate) and the annual short interest rates are* 6% *(upstate) and* 4% *(downstate). Find the price of 2-year ZCB with face value Rs 100.*
*(Hint: Compute risk neutral probability measure* $p^*$ *using 1 period binomial lattice model for stock. Find* $B(0,2)$ *using* $B(0,2) = \frac{1}{1+r}(p^* B_u(1,2) + (1-p^*)B_d(1,2))$.*)*

**Exercise 11.19** *Compute the spot rate binomial lattice model for 6 months with upward and downward parameters 1.25 and 0.9 respectively. Using this lattice, compute the price of a 4-months ZCB with face value 100. A European call option is written on the above bond. If the strike price of the call is Rs 85 and expiration time 3 months then find the price of the call.*

**Exercise 11.20** *An American put option is written on the 3-year ZCB with face value Rs 100 governed by the following short rate lattice.*

| | | | |
|---|---|---|---|
| | | | 0.08 |
| | | 0.064 | 0.059 |
| | 0.052 | 0.047 | 0.041 |
| 0.042 | 0.038 | 0.032 | 0.026 |



*If the strike price of put is Rs 110 and expiration 2 years, then calculate the put price.*

**Exercise 11.21** *Suppose the following market data for ZCBs with a maturity payoff of Rs 1 is given.*

| Maturity (years) | Price | Volatility ($\sigma$) |
|---|---|---|
| 1 | 0.945 | 0% |
| 2 | 0.885 | 10% |

*Calibrated the data on a 2-period interest rate binomial lattice with upstate annual interest* $re^{2\sigma}$ *and downstate annual interest* $r$. *Calculate* $r$.

$$\left( Note : B(0,2) = B(0,1)\Big(\frac{1}{2(1+r_u)} + \frac{1}{2(1+r_d)}\Big) \right).$$

**Exercise 11.22** *Given a current short rate of* 10%, *upward and downward parameters of* $u = 1.15$ *and* $d = 0.95$. *Generate a 3-period binomial lattice of short-term rate. Using this lattice, determine the value of a 2-period coupon bond with face* *value Rs 100. The coupons are paid at every period at* 4% *on the bond face value. (Hint: Take* $B(0, 3) = 104$ *for all four nodes at third period. Move backward and at each preceding node add a coupon of* 4 *to the node value to get the actual value (bond price) of that node).*

**Exercise 11.23** *Suppose that the short rate process* $\{r(t), t \geq 0\}$ *follows the following SDE*

$$dr = \theta dt + \sqrt{\beta} dW,$$

*where* $\theta > 0$ *and* $\beta > 0$ *are constants. Is the process mean reverting? Justify.*

**Exercise 11.24** *Suppose the short rate process* $\{r(t), t \geq 0\}$ *follows the Vasicek interest rate model with* $\alpha = 0.01$, $\beta = 0.1$, $\sigma = 0.02$, *and the short rate is* 10%. *Find the 2-year ZCB price of face value Rs 1. What is the bond yield?*

**Exercise 11.25** *Suppose that the short rate is currently* 3% *and its volatility measure is* 0.9% *per annum. What happens to the volatility measure when the short rate increases to* 5.5% *in (i) Vasicek's model, (ii) the CIR model?*

# 12
# Optimal Trading Strategies

## 12.1 Introduction

The concept of portfolio optimization and diversification has played a key role in the development and understanding of financial markets. The major breakthrough came in 1952 with the publication of Harry Markowitz's theory of portfolio selection [90]. The theory, popularly referred to as *mean-variance portfolio theory*, provided an answer to investors fundamental question: How should an investor allocate funds among the possible investment choices? Thus, the major interest for the investor is to have a balance between the total risk of the portfolio and its expected return. Investors generally set their priority in terms of minimizing total risk and maximizing return of the portfolio. On the lines of Markowitz's, Sharpe [119] proposed *capital asset pricing model* (CAPM) which introduces the notions of systematic risk and specific risk. We have already studied Markowitz's theory and CAPM in Chapter 5. An interesting question which every investor faces apart from holding the shares is about how to acquire them. This is where optimal trading strategy enters into the picture. The goal of optimal trading strategies is to formulate a mathematical approach which tries to answer some of those questions that arise during the phase of implementation of investments.

With the established positions, fund managers or the institutional investors need to re-balance their portfolios frequently, either to include new stock picks, sell stocks that are out of favor, or to improve the risk/reward characteristics of the portfolio. This generates huge orders that must be executed in a fixed time horizon. The execution costs associated with such orders can be substantial. Numerous studies have shown that these costs typically comprise the largest quantity of the *fund tracking error*. This is hardly a problem for an individual investor as trading volumes are normally small and he/she hardly worry about the execution costs in acquiring or selling out of a portfolio. On the other hand, the quantities traded by institutional investors and portfolio managers comprise a large fraction of the average daily volume of many stocks. They need to estimate the execution costs associated with the transactions to keep a track of their actual returns. Optimal trading strategies give us a way to estimate the transaction costs and how to acquire these portfolios with minimal cost and risk.

Financial transaction cost consists of fixed and variable components. In the next section we will first explain each one of them and then a brief introduction to types of orders that are executed in the market. Once the framework of transaction cost and market structure is ready, we will move forward with the execution algorithms that helps in minimizing the execution cost.

## 12.2 Introduction to Market Structure

We first explain the components of the transaction cost involved in trading the large size/block orders and then taking forward from there to types of orders traded by the traders in the market.

Execution cost has nine different transaction cost components namely - *commission, fees, taxes, spreads, investment delay, price appreciation, market impact, timing risk and opportunity cost*. We may refer to Fig.12.1 in this regard. Some of the components are fixed costs like commission, fees and taxes and they are independent of the implementation strategy. The variable cost components like *price appreciation* (natural price movement of the stock), *market impact* (movement in stock prices caused by particular orders or trades), *timing risk* (uncertainties surrounding orders) and *opportunity cost* (foregone profit of not being able to implement investment decisions) are of higher interest to us while formulating our strategies. These arise due to buying or selling pressure exerted into the market by the prospective sellers or buyers. The price of the security is decided by the law of supply and demand in the market. Now, demanding more of a good or service will lead to price appreciation and extra premium has to be paid on top of the purchase price to attract additional suppliers of the good into the market. Same goes for the supplier too.

Market impact is defined as the movement in the price of the stock caused by a particular trade. It is the difference between the stock's price trajectory with the order and what the price trajectory would have been had the order not been submitted into the market. Since we cannot simultaneously measure these two occurrences, market impact is sometime referred as the "Heisenberg uncertainty principle of finance" [77]. Because of this limitation it is measured as the difference between the stock price at the beginning of trading and the order's

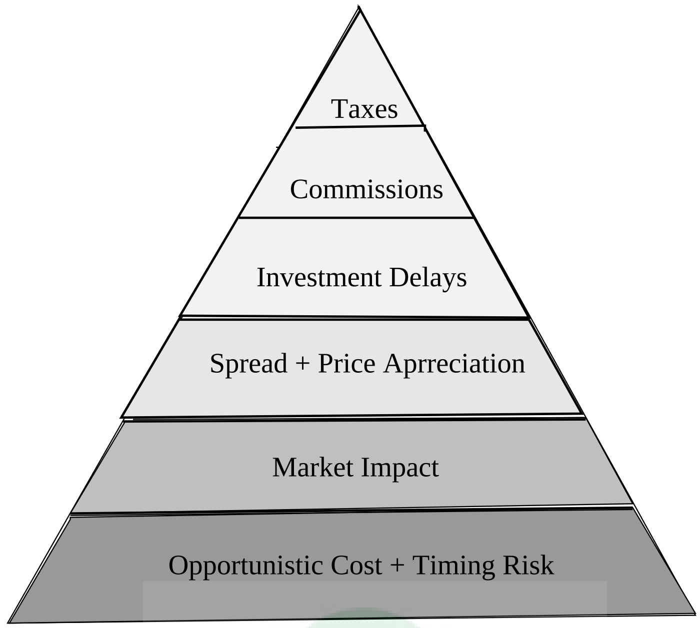



**Fig. 12.1.** Transaction Cost Components

average execution price. It causes investors to pay premiums to complete the buy orders and provides discounts to complete the sell orders. Market impact is caused by two primary reasons. These are (i) *Supply Demand Imbalance (liquidity needs)* and (ii) *Information Leakage*.

In efficient market, where the price of every security fully reflects all available information and hence is equal to its true investment value, prices adjust continuously to ensure that buying demand equals selling demand. As investors seek to buy shares, they are required to raise their price to attract additional sellers into the market. Further, due to the immediacy needs, investors often have orders that are larger than the quantity of shares available at the best market quote. To achieve immediate execution *(market order)*, it is often necessary to eat into the *limit order book*, making each successive transaction more expensive. This shift in price of shares is temporary and market soon returns to equilibrium position where demand equals supply. This type of market impact is called *temporary market impact*.

Every time an order is released to market, it conveys information regarding the investment and trading intentions of investors which in turn change the beliefs of other investors for a long time. This causes the market to believe that the future prices will be different than originally expected or there is a change in the stock's intrinsic value. This brings a quick price adjustment causing a jump or drop in the price that remains permanent. This type of market impact is called *permanent market impact*.

A portfolio manager while minimizing market impact may increase timing risk (market risk) of his transaction. To explain clearly, take a simple example of buying a stock whose price in the market is currently $Rs$ 100. If we need to buy a large quantity of this stock, we can place a market order to immediately buy the stock. Then we might drive the price of the stock up so that the average cost of buying is say $Rs$ 102 or $Rs$ 103. An obvious preventative step to reduce temporary as well as permanent market impact is to trade more slowly, i.e. break the large trade into a number of small trades and trade over a longer interval. In this case, market risk exists where the price of the stock might move in the opposite direction (increases) due to trades by other market participants in the same interval. Thus, trading strategy of an investor is the decision of how fast does one want to trade and it depends on the balance one wishes to strike between the execution cost (due to market impact) and the market risk.

Thus, trading for institutional investors is not about making profits. It is about rapid execution with minimal execution costs. Therefore, the aim of the '*optimal trading strategies*' is to formulate a mathematical approach to address the questions and issues that arise during the implementation phase of the investment cycle- like: How do we estimate trading cost? How long should execution take? How should the order be sliced? How do we choose an alternate strategy like to trade aggressively or passively? etc. To begin with, in the following, we describe the types of the order available to the traders while trading large/small orders in the market.

Orders are trade instructions. They specify what traders want to trade, whether to buy or sell, how much, when and how to trade, and, most important, on what terms. Thus, orders are the fundamental building blocks of the trading strategies. To trade effectively, a trader must specify exactly what he/she wants. An order submission strategy is the most important determinant of success of a trader. Therefore, the proper order used at the right time can make the difference between a good trade, a costly trade and no trade at all [59].

Traders indicate their willingness to buy or sell by making bids or offers respectively. Traders quote their bids and offers when they arrange their own trades. Otherwise they use orders to convey their bids and offers to the brokers or automated trading systems that arrange their trades.

The highest bid price in a market is the best bid. The lowest offer price is the best offer (or, equivalently best ask). Traders also call them the market bid and the market offer because they are the best prices available in the market. The prices at which orders fill are trade prices. The difference between the best bid and the best ask is the *bid/ask spread*.

A market is said to be liquid when traders can trade without significant adverse affect on execution price. An order offers liquidity or equivalently supplies liquidity if it gives an opportunity to the other traders to trade large size quickly at a low cost. Both buyers and sellers can offer liquidity. Buyers offer liquidity when their bids give other traders opportunity to sell, sellers offers liquidity when their offers give other traders opportunity to buy.

Markets and traders treat orders differently, depending on whether they are agency orders or proprietary orders. Agency orders are orders that brokers represent as agents for their clients. Proprietary trading (also "prop trading") occurs when a firm trades stocks, bonds, currencies, commodities, their derivatives, or other financial instruments, with the firm's own money as opposed to its customers' money, so as to make a profit for itself.

In the subsequent section we will define various types of orders that traders send to the market once they decide upon how much to execute, their time limit, and the side of the trade (i.e. buy or sell).

## 12.3 Various Types of Orders in the Market



A *market order* is an instruction to trade securities at the price currently available in the market. Market orders usually fill quickly, but sometimes at inferior prices. Traders who want to be certain about their order execution, are sometimes known as impatient traders. The execution of a market order depends on its *size* and on the liquidity currently available in the market. Small size market orders usually fill immediately with little or no effect on the prices.

The bid/ask spread (also known as bid/offer or buy/sell spread) for securities (such as stocks, futures contracts, options, or currency pairs) is the difference between the prices quoted (either by a single market maker or in a limit order book) for an immediate sale (ask) and an immediate purchase (bid). Generally market order traders pay the bid/ask spread. The bid/ask spread is the price impatient traders pay for immediacy. Impatient traders buy at the ask price and sell at the bid price. The spread is the compensation dealers and limit order traders receive for offering immediacy. The trader initiating the transaction is said to demand liquidity, and the other party (counter party) to the transaction supplies liquidity. Such traders are called *liquidity demanders* and *liquidity suppliers* respectively.

Liquidity demanders place market orders and liquidity suppliers place limit orders. For a round trip (a purchase and sale together) the liquidity demander pays the spread and the liquidity supplier earns the spread. The size of the bid-offer spread in a security is one measure of the liquidity of the market and of the size of the transaction cost.

Large market orders are more difficult to execute than smaller ones. Traders willing to take the other side of a large trade are often hard to find market order. The reason could be due to possibility of informed trading. However to attract buyers (or sellers) impatient trader often move prices. Large buyers increase bid prices of their order to encourage sellers to sell to them and vice versa. The premiums that large buyers pay and the discounts that large sellers offer are price concessions.

Let us explain the market impact using an example. Let us assume that the currently available trades in the market are 500 shares at level first with ask price $Rs$ 100, and 300 shares at level second with ask price $Rs$ 101. Now trader wants to buy 700 shares. Therefore he/she has to look towards the ask side of the market i.e. if sellers are available or not. Hence he will buy 500 shares at $Rs$ 100 and rest 200 shares at $Rs$ 101, therefore the impact price at which he will buy is $(500 \times 100 + 200 \times 101)/700 = 100.2857$. Similarly, if in the market 300 shares are available at first level with bid price $Rs$ 99, further if trader buys 200 shares at $Rs$ 100 and sell 200 shares at $Rs$ 99 then he has paid $Rs$ 1 as bid ask spread for the respective trade.

The price, at which market orders trade, depend on current market conditions. Since market conditions can change quickly, traders who use market orders are at risk of trading at worse prices than what they expected. This risk in literature is termed as *execution price uncertainty*. Execution price uncertainty is due to quote changes that may occur between the submission of an order and its execution, and to the unpredictable price concessions that may be required to fill large orders. Thus, those traders who are concerned about the execution price risk may prefer to submit limit orders.

A *limit order* is an instruction to trade at best price available, but only if it is no worse than the limit price specified by the trader. For buy orders, the trade price must be at or below the limit price. For sell orders, the price must be at or above limit price.

In continuously trading markets, a broker (or a exchange) will attempt to trade a newly submitted limit order as soon as it arrives. If no trader is immediately willing to take the opposite side at an acceptable price, the order will not be traded. Instead, it will stand as an offer to trade until someone is willing to trade at its limit price, until it expires, or until the trader who submitted it cancels it. Standing limit orders are placed in a file called a limit order book.

The probability that a limit order will trade depends on its limit price. If the limit price of a buy order is too low, the order will not trade. Likewise, if a sell limit price is too high, the order will not trade. Buy limit orders with high prices and sell limit orders with low prices are aggressively priced. Aggressively prices limit orders are easiest limit orders to fill.

Traders classify limit orders with limit prices at which they are placed relative to the market. The market is the range of prices bounded above by the best offer (lowest price) and below by the best bid (highest price). A *marketable limit order* is an order that the broker can execute immediately when a trader submits it. The limit price of a marketable limit buy order is at or above the best offer. The broker therefore can manage to buy immediately from the seller quoting the best offer. Moreover, marketable limit orders are like market orders, except that they limit the price concessions that brokers can make to fill them. Marketable limit orders with very high limit buy prices or very low sell prices are essentially market orders. Traders use marketable limit orders instead of market orders to limit execution price uncertainty and to limit what they will pay for liquidity.

Limit buy orders that stand at the best bid, and limit sell orders that stand at the best offer, are at the market. The traders who submit these orders make the market. To summarize, marketable limit orders are the most aggressively prices limit orders. Traders who submit standing limit orders offer liquidity to the other traders. Their limit orders give others the opportunity to trade when they want to trade. In particular, sell limit orders are call options that give other traders an opportunity to buy when they want to buy. Buy limit orders, likewise are put options that give other traders opportunity to sell when they want to sell. The option strike price is limit price.

The other type of orders are *stop orders*. A stop instruction stops an order from executing until price reaches a stop price specified by the trader. Traders attach stop instructions to their orders when they want to buy only after price rises to the stop price or sell only after price falls to the stop price. Orders with the stop instructions are called stop orders.

Traders most commonly use stop orders to stop their losses when prices move against their positions. For example, suppose that a trader buys 100 sugar future contracts at *Rs* 100 each. To limit the potential loss on this position, he/she may issue a market sell order for 100 contracts with a stop price of *Rs* 90. If the sugar drops to or below *Rs* 90, his/her broker will immediately try to sell 100 contracts at the best price then available in the market. Traders often calls such orders *stop* *loss orders*. The price at which a stop order executes may not be the stop price. In the above example, if sugar falls quickly from $Rs$ 97 to $Rs$ 88, his/her broker may be able to sell the 100 contracts at $Rs$ 87.5.

When a trader attach a stop instruction to a limit order, they must specify two prices. The stop price indicates when the limit order becomes active, and the limit price indicates the terms upon which a trade may be arranged. The combined order is stop limit order. With stop limit order, trader do not need to monitor the market, and thus are free to attend to other business.

Stop orders accelerate price changes. Prices often change because traders on one side of the market demand more liquidity than what is available. When these price changes activate the stop orders, it contributes to the one-sided demands for liquidity. Stop orders accelerate price changes by adding buying pressure when prices are rising and selling pressure when prices are falling. They demand liquidity when it is least available. Traders claim that stop orders add momentum to the market. Traders who pursue momentum trading strategies buy when prices are rising and sell when price are falling. They basically take the advantage of stop orders to create momentum in the market. On the other hand, contrarian traders employ the opposite trading strategy. They buy when prices are falling and sell when prices are rising. They therefore stabilize prices when they trade.

*A trailing stop order* is entered with a stop parameter that creates a moving or trailing activation price, hence the name. This parameter is entered as a percentage change or actual specific amount of rise (or fall) in the security price. Trailing stop sell orders are used to maximize and protect profit as a stock's price rises and limit losses when its price falls. Trailing stop buy orders are used to maximize profit when a stock's price is falling and limit losses when it is rising.

For example, a trader has bought stock ABC at $Rs$ 10 and immediately places a trailing stop sell order to sell ABC with a $Rs$ 1 trailing stop. This sets the stop price to $Rs$ 9. After placing the order, ABC doesn't exceed Rs10.00 and falls to a low of $Rs$ 9.01. The trailing stop order is not executed because ABC has not fallen $Rs$ 1 from $Rs$ 10. Later, the stock rises to a high of $Rs$ 15 which resets the stop price to $Rs$ 14. It then falls to $Rs$ 14 ($Rs$ 1 from its high of $Rs$ 15) and the trailing stop sell order is entered as a market order.

Another type of order is *good till cancelled order*. A good-till-cancelled (GTC) order is an order to buy or sell a security at a specific or limit price that lasts until the order is completed or cancelled. A GTC order will not be executed until the limit price has been reached, regardless of how many days or weeks it might take. Investors often use GTC orders to set a limit price that is far away from the current market price. Some brokerage firms may limit the time a GTC order can remain in effect and may charge more for executing this type of order. An uptick is when the last (non-zero) price change is positive, and a downtick is when the last (non-zero) price change is negative. Any tick sensitive instruction can be entered at the trader's option, for example buy on downtick, although these orders are rare.

So far in this section, we have discussed various types of orders that traders execute. In making the decision of how much to trade depends upon the market conditions like price and volume available in the market. Thus traders generally use technical indicators to calculate the price and volume change in the previous trading interval using historical data to forecast for the next trading interval. Now a days traders prefer to use algorithmic trading strategies that learn the market conditions and based on the favorable trading rules/circumstances schedule the trade for the current trading interval. The traders while trading large blocks of trade face market impact on their execution price.

In the following sections, we describe some models which provide quantitative trading strategies that are helpful to the traders to minimize the market impact. These models are refered as *execution models*.

## 12.4 Execution Models

Optimal execution characterizes the minimal cost realization of an investment decision in a financial market. An investment decision concerns the change in portfolio position of an investor through the acquisition or liquidation of listed assets. For an investor wishing to buy (sell) a given volume at a particular price there must exist a counterparty (or number of counterparties) willing to sell (buy) the same volume. In a listed financial market the buyers and sellers post orders for the assets that they wish to acquire or liquidate. For an investor wishing to buy (sell) at a particular price there must be enough counter liquidity at the specified price. If there is insufficient counter liquidity then the investor must choose to either transact at a worse price where there is sufficient counter liquidity or over a period of time in the hope that market dynamics attract more counter liquidity at an acceptable price. The performance of the trade execution, on completion of the given transaction, may then be measured and benchmarked. Performance can be measured using a simple process of taking the initial price at the start of the transaction, referred to as the arrival price, multiplied by the total volume that was exchanged, and comparing this against the weighted average price received multiplied by the volume exchanged over the duration of the transaction. The difference between the theoretical benchmark price and the actual price received is the *implementation shortfall* (Perold [107]).

The first quantitative description of estimating and minimizing market impact has been given by Bertsimas and Lo [11] in 1998. Here, given a fixed block $\bar{S}$ of shares to be executed (acquired/liquidated) within a fixed time interval [0,T], and given price dynamics that captures price impact, the objective is to find an optimal sequence of trades (as a function of the state variables) that will minimize the expected cost of executing $\bar{S}$ shares within time $T$. For this, Bertsimas and Lo [11] considered various price processes and employed dynamic programming technique to develop optimal trading strategies for a risk neutral investor. We will discuss their models and also limitations of these models.

Another major contribution in this area is due to Almgren and Chriss [2], who observed that market impact and timing risk associated with the execution of trades are conflicting concepts and hence it is not possible to simultaneously minimize both. They proposed a mean variance based objective function for their model and obtained the static strategies where the optimal trading path was determined in advance of trading. Further, a model is called *static* if there is no serial correlation among prices, or no change in the market perception of the price of the stock. In general, we expect the optimal strategy to be *dynamic* because optimal trading path cannot be determined priori of trading, as one needs to update the optimal strategy based on observed state variables at each time point. We shall not discuss the details of Almgren and Chriss' model here and shall refer to [2] for further study in this regard. However, taking motivation from [2], we shall incorporate variance terms in the orginal models of Bertsimas and Lo [11] and present a static approximation technique for the same (Khemchandani et al. [76]).

The problem statement for obtaining optimal strategy that minimizes the execution cost can be formulated mathematically as follows: Consider an investor who wishes to buy a large block $\bar{S}$ of shares of some stock over a fixed time interval $[0, T]$. We divide the interval $[0, T]$ into $N$ subintervals of length $\tau = T/N$ and define the discrete times $t_k = k\tau$, for $k = 0, 1, \ldots, N$. Thus $t_0 = 0$ and $t_N = T$. Here it is understood that the trading begins from the first period. Thus at $t_1$, $S_1$ shares are purchased at price $P_1$. In general, at $t_k$, $S_k$ shares are purchased at price $P_k$, $k = 1, \ldots, N$. Let $W_k$ be the number of shares that remain to be purchased at time $t_k$. The investor's objective is to minimize expected cost of buying $\bar{S}$ shares for a given level of risk, say $V^*$. If we use variance of cost of buying $\bar{S}$ shares as a measure of timing risk [2], [64], then investor's problem is represented as follows.

$$\begin{aligned}
&\underset{\{S_k\}}{\text{Min}} \qquad E_1\Big[\textstyle\sum_{k=1}^{N} P_k S_k\Big]\\
&\text{subject to}\\
&\qquad\qquad Var_1\Big[\textstyle\sum_{k=1}^{N} P_k S_k\Big] \le V^*\\
&\qquad\qquad\qquad \textstyle\sum_{k=1}^{N} S_k = \bar{S}\\
&\qquad\qquad\qquad\quad S_k \ge 0 \quad (k = 1, 2, \ldots, N).
\end{aligned}$$

Here $E_1$ and $Var_1$ respectively represent the expected cost and risk which trader will face at the starting of the interval assuming that he/she has traded $\{S_1, S_2, \ldots, S_N\}$ shares in the intervals $\{T_1, T_2, \ldots, T_N\}$ respectively. The constraint $S_k \ge 0$ implies shares purchased while $S_k \le 0$ implies shares sold. The problem to sell $\bar{S}$ shares in time $[0, T]$ is a symmetric problem. We need to maximize the objective function, i.e. maximize the revenue generated by liquidating $\bar{S}$ shares.

We can reformulate the above problem as

$$\begin{aligned}
&\underset{\{S_k\}}{\text{Min}} \qquad E_1\Big[\sum_{k=1}^{N} P_k S_k\Big] + \lambda Var_1\Big[\sum_{k=1}^{N} P_k S_k\Big]\\
&\text{subject to}\\
&\qquad\qquad \textstyle\sum_{k=1}^{N} S_k = \bar{S}\\
&\qquad\qquad\quad S_k \ge 0 \quad (k = 1, 2, \ldots, N),
\end{aligned} \tag{12.1}$$



where $\lambda \ge 0$ is a parameter. The parameter $\lambda$ has a financial interpretation, and it is a measure of risk-aversion, that is, how much we penalize execution risk relative to expected cost. The only way traders can incur a lower trading cost is through a strategy with a higher quantity of risk and vice versa. For a risk neutral investor $\lambda = 0$, while for a risk averse investor $\lambda > 0$.

Depending upon the law of motion for price process $\{P(t), t \ge 0\}$, optimal strategies can be static or dynamic. Static strategy is one where optimal trading path is determined in advance of trading. Dynamic strategy is one where optimal trading path cannot be determined priori of trading and one needs to update the optimal strategy based upon the market information at each time period. Bertsimas and Lo [11] and Huberman et al. [64] have used stochastic dynamic programming to solve the above problem for various law of motion for price process. Some of their models are discussed in more detail in the following sections.

## 12.5 Static Optimal Strategies for Risk Neutral Investor: A Dynamic Programming Approach

To provide a concrete illustration of the dynamic programming approach, we assume that the investor is risk neutral ($\lambda = 0$) and the price in the absence of our trade is given by arithmetic random walk. The impact cost is a linear function of trade size so that a purchase of $S_k$ shares may be executed at the prevailing price $P_{k-1}$ plus an impact premium of $\theta S_k, \theta > 0$. Therefore, the law of motion for $P_k$ may be expressed as

$$P_k = P_{k-1} + \theta S_k + \epsilon_k, \quad \theta > 0, \tag{12.2}$$

where $\epsilon_k' s$ are assumed to be a zero-mean independent and identically distributed (i.i.d) random shocks, which provide randomness to the underlying price process

$$E[\epsilon_k | S_k, P_{k-1}] = 0, \quad Var[\epsilon_k | S_k, P_{k-1}] = \sigma^2.$$

This model is same as the basic model of Bertsimas and Lo [11] who derived the optimal solution of (12.1) for risk neutral investor under price process (12.2), using dynamic programming. The basic requirements for any dynamic programming problem are the state of the environment at time $t_k$, the control variables, the randomness, the objective function, and the law of motion. In our context, the state at time $t_k$, for $k = 1, \ldots, N$ consists of the price $P_{k-1}$ realized at the previous period, and $W_k$, the number of shares remain to be purchased at time $t_k$. The state variables give all the information the investor requires in each period $t_k$ to make decision regarding the control. The control variable at time $t_k$ is the number of shares $S_k$ purchased. The objective function is given by

$$\underset{\{S_k\}}{\text{Min}} \quad E_1\Big[\sum_{k=1}^{N} P_k S_k\Big],$$

and the law of motion for state variable $W_k$ is represented as

$$W_k = W_{k-1} - S_{k-1},\ W_1 = \bar{S},\ W_{N+1} = 0. \tag{12.3}$$

Here we ignore the non-negativity constraint $S_k \geq 0$ which is discussed later in this section.

The dynamic programming algorithm is based on 'the principle of optimality' which states that, "An optimal policy has the property that, whatever the initial state and control are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision". Therefore for every $t_k$, $k = 1, \ldots, N$, the sequence $\{S_k^*, S_{k+1}^*, \ldots, S_N^*\}$ must still be optimal for the remaining program $V_k(P_{k-1}, W_k) = E_t[\sum_{i=k}^N P_i S_i]$. This property is summarized by the Bellman equation, which relates the optimal value of the objective function in period $t_k$ to its optimal value in period $t_{k+1}$

$$V_k(P_{k-1}, W_k) = \underset{\{S_k\}}{\text{Min}} \;\; E_k[P_k S_k + V_{k+1}(P_k, W_{k+1})]. \tag{12.4}$$

By starting at the end (time $t_N$) and applying the Bellman equation (12.4) and the law of motion for $P_k$ (12.2) and $W_k$ (12.3) recursively, the optimal value function can be derived as a function of the state variables. The Bellman equation (12.4) for time $t_N$ is given by

$$V_N(P_{N-1}, W_N) = \underset{\{S_N\}}{\text{Min}} \;\; E_N[P_N S_N] = (P_{N-1} + \theta W_N) W_N. \tag{12.5}$$

Since this is the last period and $W_{N+1} = 0$ by (12.3), there is no choice but to execute the entire remaining order $W_N$, hence the optimal trade size $S_N^* = W_N$. In the next to last period $t_{N-1}$, the Bellman equation is

$$V_{N-1}(P_{N-2}, W_{N-1}) = \underset{\{S_{N-1}\}}{\text{Min}} \;\; E_{N-1}[P_{N-1} S_{N-1} + V_N(P_{N-1}, W_N)] \tag{12.6}$$

Substituting the law of motion for $P_{N-1}$ from (12.2) and $W_N$ from (12.3), we get

$$\begin{aligned} V_{N-1}(P_{N-2}, W_{N-1}) = \underset{\{S_{N-1}\}}{\text{Min}} \;\; & E_{N-1}[(P_{N-2} + \theta S_{N-1} + \varepsilon_{N-1}) S_{N-1} \\ & + V_N(P_{N-2} + \theta S_{N-1} + \varepsilon_{N-1}, W_{N-1} - S_{N-1})] \end{aligned}$$

Using the fact that $E_{N-1}[\varepsilon_{N-1}] = 0$, and using the right-hand side of (12.5), the above equation can be expressed as explicit function of $S_{N-1}$ as

$$V_{N-1}(P_{N-2}, W_{N-1}) = \underset{\{S_{N-1}\}}{\text{Min}} \;\; [\theta S_{N-1}^2 - \theta S_{N-1} W_{N-1} + \theta W_{N-1}^2 + P_{N-2} W_{N-1}].$$

This can be minimized by taking its derivative with respect to $S_{N-1}$, and solving for its zero, yielding

$$S_{N-1}^* = W_{N-1}/2,$$

$$V_{N-1}(P_{N-2}, W_{N-1}) = W_{N-1}\Big(P_{N-2} + \frac{3}{4}\theta W_{N-1}\Big).$$

Continuing through backward recursion, we get

$$S^*_{N-k} = W_{N-k}/(k+1), \tag{12.7}$$

$$V_{N-k}(P_{N-k-1}, W_{N-k}) = W_{N-k}\Big(P_{N-k-1} + \frac{k+2}{2(k+1)}\theta W_{N-k}\Big). \tag{12.8}$$

We encourage the readers to work out (12.7) and (12.8). When we finally reach the beginning of the program (time $t_1$), we get

$$S^*_1 = W_1/N,$$

$$V_1(P_0, W_1) = W_1\Big(P_0 + \frac{N+1}{2N}\theta W_1\Big).$$

Substituting the initial condition $W_1 = \bar{S}$ from (12.3) into above equation gives

$$S^*_1 = \bar{S}/N, \tag{12.9}$$

$$V_1(P_0, W_1) = E_1\Big[\sum_{k=1}^{N} P_k S^*_k\Big] = P_0\bar{S} + \frac{\theta\bar{S}^2}{2}\Big(1 + \frac{1}{T}\Big).$$

Using (12.9), (12.7) and (12.3), we find that

$$S^*_1 = S^*_2 = \ldots = S^*_N = \bar{S}/N.$$

The optimal execution strategy is simply to divide the total order $\bar{S}$ into $N$ equal parts and trade them at regular intervals. This strategy is called a "naive" strategy. This simple trading strategy comes from the fact that the price impact $\theta S_k$ does not depend on either the prevailing price $P_{k-1}$ or the size of the unexecuted order $W_k$. Hence, the price impact function is same in each period and independent from one period to the next.

## 12.6 Static Optimal Strategies for A Risk Averse Investor: A Dynamic Programming Approach

Bertsimas and Lo [11] minimized only expected cost of execution and ignored the risk associated with execution strategy. Almgren and Chriss [2], and Huberman and Stanzl [64] defined risk as the variance of cost of execution and used a mean-variance objective function. But, when we use mean-variance objective function, it no longer remains additive-separable. Thus, the dynamic programming technique cannot be used to compute the optimal strategy for a risk averse investor. For static models, Huberman and Stanzl [64] have shown that an equivalent dynamic program exists and one can find the optimal strategy for risk averse investors. We used the same result to derive the optimal strategy for risk averse investor for basic model of Bertsimas and Lo.

For a risk averse investor, the objective function is given by

$$\underset{\{S_k\}}{\text{Min}} \;\; E_1\Big[\sum_{k=1}^{N} P_k S_k\Big] + \lambda Var_1\Big[\sum_{k=1}^{N} P_k S_k\Big].$$

The law of motion for state variables $P_k$ and $W_k$ are as before in (12.2) and (12.3) respectively. The Bellman equation for a risk averse investor is given by

$$V_k(P_{k-1}, W_k) = \underset{\{S_k\}}{\text{Min}} \;\; \Big\{E_k[P_k S_k + V_{k+1}(P_k, W_{k+1})] + \lambda Var_k[P_k S_k + V_{k+1}(P_k, W_{k+1})]\Big\}. \tag{12.10}$$

As before, the Bellman equation at time $t_N$ is given by

$$\begin{aligned} V_N(P_{N-1}, W_N) &= \underset{\{S_N\}}{\text{Min}} \;\; \Big\{E_N[P_N S_N] + \lambda Var_N[P_N S_N]\Big\} \\ &= (P_{N-1} + \theta W_N)W_N + \lambda W_N^2 \sigma^2. \end{aligned}$$

In the next to last period $t_{N-1}$, the Bellman equation is

$$\begin{aligned} V_{N-1}(P_{N-2}, W_{N-1}) = \underset{\{S_{N-1}\}}{\text{Min}} \;\; &\Big\{E_{N-1}[P_{N-1} S_{N-1} + V_N(P_{N-1}, W_N)] \\ &+ \lambda Var_{N-1}[P_{N-1} S_{N-1} + V_N(P_{N-1}, W_N)]\Big\}. \end{aligned}$$

Substituting the law of motion for $P_{N-1}$ from (12.2) and $W_N$ from (12.3) and simplifying, we get a quadratic function explicit in $S_{N-1}$ which can be minimized by solving for zero of its derivative as before. The best execution strategy and optimal value function can be obtained by recursively solving. Here, they are given by

$$S^*_{T-k} = b_k W_{T-k},$$

$$V_{T-k} = P_{T-k-1} W_{T-k} + (a_k + \lambda \sigma^2) W_{T-k}^2,$$

for $k = 0, 1, \ldots, T-1$, where

$$a_k = \theta\Big(1 - \frac{\theta}{4(a_{k-1} + \lambda\sigma^2)}\Big), \quad a_0 = \theta, \tag{12.11}$$

$$b_k = 1 - \frac{\theta}{2(a_{k-1} + \lambda\sigma^2)}, \quad b_0 = 1. \tag{12.12}$$

The optimal strategies for risk neutral, risk averse and risk seeking investors are compared using a numerical example. The results are shown in Fig 12.2 and Fig 12.3. Risk averse investor follows an aggressive strategy trading heavily in the initial phases. Risk seeking investor on the other hand follows passive strategy, waiting for the price to fall and trades at the end. Risk-neutral investor follows the naive strategy.

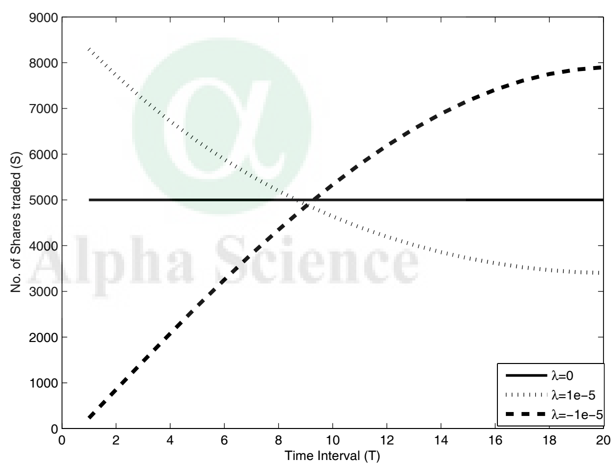



**Fig. 12.2.** Comparison of strategies for risk neutral, risk averse and risk seeking investors for Model 1. S represents the number of shares acquired in each period. The black bold curve corresponds to a risk neutral investor ($\lambda = 0$). The black dash curve corresponds to risk averse investor ($\lambda = 1 \times 10^{-5}$), while the black dotted curve corresponds to risk seeking investor ($\lambda = -1 \times 10^{-5}$).

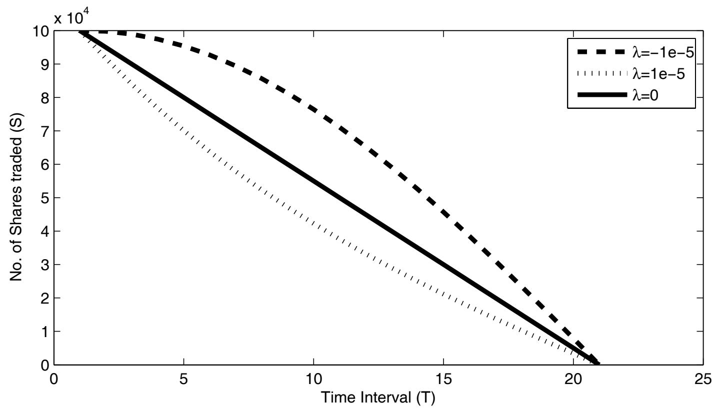



**Fig. 12.3.** Comparison of strategies for risk neutral, risk averse and risk seeking investors for Model 1. We represents the number of unexecuted shares. The black bold curve corresponds to a risk neutral investor ($\lambda = 0$). The black dashed curve corresponds to risk averse investor ($\lambda = 1 \times 10^{-5}$), while the black dotted curve corresponds to risk seeking investor ($\lambda = -1 \times 10^{-5}$).

## 12.7 Dynamic Models

The strategies obtained so far were static where the optimal solution $S^*$ was determined in advance of trading. It was based on the observation that static optimization procedures lead to globally optimal trading strategies in absence of market information or serial correlation among price process. If we introduce market information variable, the optimal strategy will no longer be static. The dynamic programming approach for risk neutral investor, for various dynamic models, given by Bertsimas and Lo is discussed in the sequel.



**Linear Price Impact with Market Information Model**

If we assume that price $P_k$ at time $t_k$ is a function of market information which changes over time, then the number of shares traded at time $t_k$ also depends on the value of market information at time $t_k$. Therefore, the optimal strategy obtained is dynamic where one can not determine the strategy in advance of trading. The following model illustrates the concept of dynamic strategy.

Let $X_k$ be a serially-correlated state variable which affects the execution price $P_k$ linearly. The price impact function is again assumed to be linear in $S_k$. Therefore, the law of motion for execution price $P_k$ is represented as:

$$P_k = P_{k-1} + \theta S_k + \gamma X_k + \epsilon_k, \ \theta > 0, \tag{12.13}$$

$$X_k = \rho X_{k-1} + \eta_k, \ \ \rho \in (-1, 1), \tag{12.14}$$

where $\epsilon_k$ and $\eta_k$ are independent *white noise* processes with mean 0 and variances $\sigma_\varepsilon^2$ and $\sigma_\eta^2$, respectively. The presence of $X_k$ in the law of motion for $P_k$ captures the potential impact of changing market conditions or of private information about the security, and $\gamma$ represents its importance for the price process $P_k$. For example, $X_k$ might be the return on the S&P 500 Index, a common component in the prices of most equities. Using dynamic programming technique [11] the best-execution strategy can be obtained and is given by

$$S^*_{N-k} = W_{N-k}/(k+1) + \delta_k X_{N-k} \tag{12.15}$$

where $\delta_k$ is the coefficient of $X_{N-k}$ which can be computed recursively, see [11] for details.

In contrast to the case of a linear price-impact function with no information, the best execution strategy (12.15) varies over time as a linear function of the remaining shares $W_{N-k}$ and the information variable $X_{N-k}$. The first term of (12.15) is simply the naive strategy of dividing the remaining shares $W_{N-k}$ at time $N-k$ evenly over the remaining $k+1$ periods. The second term of the (12.15) is an adjustment that arises from the presence of serially correlated information $X_{N-k}$. Note that the number of shares traded at time $t_k$ is a function of $X_k$, and therefore the optimal strategy obtained is dynamic in nature.

If $\rho = 0$, the term $\delta_k$ vanishes then the naive strategy is an optimal solution. If $\rho > 0$, then $\delta_k$ is positive, implying that the positive realizations of $X_{N-k}$ increases the number of shares purchased at time $t_{N-k}$. Similarly, if $\rho < 0$, positive realizations of $X_{N-k}$ decreases the number of shares purchased at time $t_{N-k}$ as $\delta_k$ is negative.

Though the strategy obtained in (12.15) is dynamic in nature, the above model suffers from a number of drawbacks that are mentioned below.

- Prices $P_k$ are assumed to follow arithmetic random walk in (12.13) implying a positive probability for negative prices.
- Price impact and information have only permanent effect on prices whereas several recent empirical studies suggest that some combination of permanent and temporary impact exists in the market.
- The percentage price impact - as a percentage of the execution price - is a decreasing function of the price level, which is counterfactual.

To overcome above limitations, Bertsimas and Lo [11] suggested the linear percentage temporary price impact model.

**Linear-percentage Temporary Price Impact Model**
In this model, we assume that the execution price at time $t_k$ comprises of two components, the no-impact price $\tilde{P}_k$ and the price impact $\Delta_k$:

$$P_k = \tilde{P}_k + \Delta_k\,.$$

The no impact price may be viewed as the price which would prevail in the market in the absence of any market impact. To ensure that prices do not go negative we assume geometric Brownian motion for the price dynamics $\tilde{P}_k$ and is defined as

$$\tilde{P}_k = \tilde{P}_{k-1}\exp(Z_k)$$

where $Z_k$ are i.i.d. normal random variables with mean $u_z$ and variance $\sigma_z^2$.

The price impact $\Delta_k$ captures the effect of trade size $S_k$ on the transaction price, hence

$$\Delta_k = (\theta S_k + \gamma X_k)\tilde{P}_k,$$

$$X_k = \rho X_{k-1} + \eta_k,$$

where $\eta_k$ is white noise with mean 0 and variance $\sigma_\eta^2$. We set $X_k$ to be an $AR(1)$ (auto-regressive with lag 1) process. The parameters $\theta$ and $\gamma$ measure the sensitivity of price impact to trade size and market conditions. The closed form solution can be obtained using dynamic programming [11] as

$$S^*_{N-k} = \delta_k^w W_{N-k} + \delta_k^x X_{N-k} + \delta_k^1, \tag{12.16}$$

where $\delta_k^w$, $\delta_k^x$ and $\delta_k^1$ are fixed coefficients. We are skipping the details here but interested readers are encourage to see [11].

The linear percentage temporary price impact model resolves a number of problems discussed with the aforementioned models.

- First, $\tilde{P}_k$ is guaranteed to be non-negative and $P_k$ is also guaranteed to be non negative under mild restriction on $\Delta_k$.
- Second, by separating the transaction price $P_k$ into a no-impact price component $\tilde{P}_k$ and the impact component $\Delta_k$, the price impact of a trade is temporary, moving the current transaction price but having no effect on future prices.
- Third, the percentage price impact increases linearly with the trade size.

- Fourth, the linear percentage temporary price impact law of motion implies a natural decomposition of execution costs, decoupling market microstructure effects from price dynamics. The objective function can be separated into two terms:

$$\underset{\{S_k\}}{\text{Min}} \ E_1\Big[\sum_{k=1}^{N} P_k S_k\Big] = \underset{\{S_k\}}{\text{Min}} \ \Big\{E_1\Big[\sum_{k=1}^{N} \tilde{P}_k S_k\Big] + E_1\Big[\sum_{k=1}^{N} \Delta_k S_k\Big]\Big\}.$$

The first term is the no-impact cost of execution and second term is the total impact cost. This decomposition is precisely the one proposed by Perold [107] in his definition of implementation shortfall, but now applied to executing $\bar{S}$.

## 12.8 Limitations of Dynamic Programming

The dynamic programming technique provides a general approach to find the optimal strategy, nevertheless, it has several important limitations.

**(*i*) Inability to impose constraints**

The most important limitation of dynamic programming approach to obtain optimal strategy is the inability to impose constraints in the dynamic models. In most practical applications, there will be constraints on the kind of execution strategies that institutional investors can follow. For example, if a block of shares is to be purchased within time $[0, T]$, it is very difficult to justify selling the stock during this time period even if such sales are warranted by the best-execution strategy. For a buy program, imposing a simple constraint like $S_k \geq 0, (k = 1, \dots, N)$ increases the feasible intervals for calculating $S^*_{N-k}$ exponentially in $k$.

We know that for $k = 0$ the optimal control is $S^*_T = W_T$. At the next stage, $k = 1$, the optimal control $S^*_{T-1}$ is computed by minimizing a quadratic function of $S_{T-1}$ subject to the constraints $0 \leq S_{T-1} \leq W_{T-1}$. The solution is given by

$$S^*_{T-1} = \begin{cases} 0, & \text{if } a_1 W_{T-1} + b_1 X_{T-1} < 0, \\ a_1 W_{T-1} + b_1 X_{T-1}, & \text{if } 0 < a_1 W_{T-1} + b_1 X_{T-1} < W_{T-1}, \\ W_{T-1}, & \text{if } a_1 W_{T-1} + b_1 X_{T-1} > W_{T-1}. \end{cases}$$

This partitions the range of $W_{T-1}$ into three intervals, over each interval there is a different optimal control $S^*_{T-1}$. For example, if $N = 20$, there are $3^{20} = 3,486,784,401$ intervals at the last stage of the dynamic program [11]. The discussion pertaining to the complexity of imposing constraint to execution problem with linear price impact with information model for $N = 3$, have been summarized in [55].

Further, the closed form solution obtained above does not impose non-negativity constraint and it is possible that the solution given by optimal strategy sells the the stock and buys it again in the time interval $[0, T]$. This limitation of dynamic programming can be overcome by quadratic programming approach discussed in Section 12.6. Other types of constraint like shrinking portfolio constraint, participation rate constraint, tax-motivated constraint are also possible [77], but difficult to work out with dynamic programming approach.

### (*ii*) Risk averse investors

For a dynamic programming problem, the objective function must be additive-separable. When we use mean-variance objective function, it no longer remains additive-separable. As discussed above, one can find the optimal strategy for risk averse investors for static models but not for dynamic models.

### (*iii*) Curse of dimensionality

We have assumed that the price follows an arithmetic random walk or geometric Brownian motion, while the market information is an autoregressive process of order one. These assumptions may not hold true in real markets where more general law of motions are observed. It may not be possible to obtain close form solutions for general dynamics of price and market information. Therefore, some numerical techniques have been suggested to solve such problems [11]. The most common technique is to discretize the state and control space and use grid search to find the optimal solution at each stage. As the number of levels for state and control space increases, the computational requirement increases dramatically due to which it becomes practically infeasible to solve the problem in the time frame available to a trader. This problem is called *curse of dimensionality*.

Let us consider an example to understand the computational requirement of discretization. Suppose that $100,000$ shares, currently trading at $Rs\ 50$, must be executed over the next 20 periods. Assuming that price will be within $Rs\ 45$ to $Rs\ 55$, with a minimum price variation of $0.125$, there are 80 possible price states. We can assume that shares are traded in a lot size of $100$. Hence, there are $1,000$ lots possible. suppose that information variable can also be discretized into $10$ states. Then at each step we will need to compute the optimal value function for $(1000 * 80 * 10)$ values of the state and control variables. With $20$ periods this implies a total of 16 million evaluations of the optimal value-function. Assuming that computation of an optimal value function takes $10^{-6}$ seconds, 160 seconds are required to find the optimal strategy which is huge as compared to the time available to investor to make rational decision. If more refined discretization is needed, the computational complexity increases dramatically. For example, if we allow the trade size to vary in increments of 1 share instead of 100 shares, and if the information variable takes on 100 discrete values instead of 10, the computational demands increases by a factor of 1,000, i.e, 16,000 seconds or roughly 4 hours.

The following table summarizes the situations or types of models which can/can not be handelled by the dynamic programming approach.

| | Static models | Dynamic models |
|---|---|---|
| Imposing constraint | Yes | No |
| Risk neutral investor | Yes | Yes |
| Risk averse investor | Yes | No |

**Table 12.1.** Models which dynamic programming can solve

## 12.9 Quadratic Programming Approach for Optimal Execution: Static Models

The prominent limitations of dynamic programming technique are the inability to impose constraints and inclusion of risk in the objective function. These limitations can be overcome by using an approximation method where we determine the static solution and make it dynamic artificially by recomputing the static strategy whenever there is a change in the market information. For a linear price impact function, expectation and variance of implementation shortfall computed at any time instant is a quadratic functions of control variables. Therefore the static optimization problem becomes a quadratic programming problem which can be solved using any standard optimizer like interior point algorithm. We can easily impose any kind of constraints on this static optimization problem.

For a static model, this approximation method always gives optimal solutions that are in line with the solution obtained with Dynamic Programming. For dynamic models, we show that this approximation method is very close to the optimal strategy. In fact, for the unconstrained problem of risk-neutral investor, the approximation method gives exact optimal solutions.

We will discuss the approximation method in more detail in the following sections and try to compare the results with those of optimal strategies computed using dynamic programming.

We begin the discussion first with static models.

**Arithmetic Random Walk with Linear Price Impact Model: A Static model**

Let $S_k$ be the number of shares acquired in period $t_k$ at price $P_k$. The price process is same as that is mentioned in Section 12.4

$$P_k = P_{k-1} + \theta S_k + \epsilon_k, \;\; \theta > 0, \;\; E[\epsilon_k | S_k, P_{k-1}] = 0, \tag{12.17}$$

where $\epsilon_k$ is assumed to be a zero-mean (i.i.d.) random process and $P_0$ is initial price of stock. To illustrate the approach, we divide $T$ into 2 intervals. Later, we will generalize for $N$ periods. For a risk neutral investor, the problem for two time periods is

$$\begin{aligned} \underset{\{S_1,S_2\}}{\text{Min}} \quad & E_1[P_1S_1 + P_2S_2] \\ \text{subject to} \quad & \\ & S_1 + S_2 = \bar{S} \\ & S_1, S_2 \geq 0. \end{aligned} \tag{12.18}$$

Using (12.17), the price $P_1$ and $P_2$ can be computed as

$$\begin{aligned} P_1 &= P_0 + \theta S_1 + \epsilon_1 \\ P_2 &= P_1 + \theta S_2 + \epsilon_2 \\ &= (P_0 + \theta S_1 + \epsilon_1) + \theta S_2 + \epsilon_2 \end{aligned}$$

Therefore,

$$\begin{aligned} E_1[P_1S_1 + P_2S_2] &= E_1[(P_0 + \theta S_1 + \epsilon_1)S_1 + ((P_0 + \theta S_1 + \epsilon_1) + \theta S_2 + \epsilon_2)S_2] \\ &= P_0S_1 + \theta S_1^2 + P_0S_2 + \theta S_1S_2 + \theta S_2^2 \\ &= P_0(S_1 + S_2) + \theta(S_1^2 + S_1S_2 + S_2^2) \end{aligned}$$

Thus, the problem (12.18) is equivalent to

$$\begin{aligned} \underset{S}{\text{Min}} \quad & C^TS + S^TQS \\ \text{subject to} \quad & \\ & e^TS = \bar{S} \\ & S \geq 0, \end{aligned} \tag{12.19}$$

where

$$C = \begin{pmatrix} P_0 \\ P_0 \end{pmatrix}, \; Q = \begin{pmatrix} \theta & \theta/2 \\ \theta/2 & \theta \end{pmatrix}, \; S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}.$$

Generalizing above problem for $N$ periods, we have the following quadratic programming problem,

$$\begin{aligned} \underset{S}{\text{Min}} \qquad & C^T S + S^T Q S \\ \text{subject to} \qquad & \\ & e^T S = \bar{S} \\ & S \geq 0, \end{aligned} \tag{12.20}$$

where

$$C = \begin{pmatrix} P_0 \\ P_0 \\ \vdots \\ P_0 \end{pmatrix}, \; Q = \begin{pmatrix} \theta & \theta/2 \dots \theta/2 \\ \theta/2 & \theta \quad \dots \theta/2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \theta/2 & \theta/2 \dots \quad \theta \end{pmatrix}, \; S = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{pmatrix}.$$

We encourage the readers to see that the optimal solution of above QPP will be naive strategy, that is,



$$S^* = \begin{pmatrix} \bar{S}/N \\ \bar{S}/N \\ \vdots \\ \bar{S}/N \end{pmatrix}$$

which is same as one obtained using dynamic programming.

For a risk averse investor, the optimization problem is

$$\begin{aligned} \underset{S_k}{\text{Min}} \qquad & E\Big[\sum_{k=1}^{N} P_k S_k\Big] + \lambda Var\Big[\sum_{k=1}^{N} P_k S_k\Big] \\ \text{subject to} \qquad & \\ & \textstyle\sum_{k=1}^{N} S_k = \bar{S} \\ & S_1, \dots, S_N \geq 0\,. \end{aligned} \tag{12.21}$$

For two period problem,

$$Var[P_1S_1 + P_2S_2] = S_1^2 Var[P_1] + S_2^2 Var[P_2] + 2S_1S_2 Cov[P_1, P_2]$$

where

$$\begin{aligned} Var[P_1] &= Var[P_0 + \theta S_1 + \epsilon_1] \\ &= Var[\epsilon_1] \\ &= \sigma_\epsilon^2. \end{aligned}$$

$$\begin{aligned} Var[P_2] &= Var[P_1 + \theta S_2 + \epsilon_2] \\ &= Var[P_0 + \theta S_1 + \epsilon_1 + \theta S_2 + \epsilon_2] \\ &= Var[\epsilon_1] + Var[\epsilon_2] + 2Cov[\epsilon_1, \epsilon_2] \\ &= 2\sigma_\epsilon^2. \end{aligned}$$

$$\begin{aligned} Cov[P_1, P_2] &= Cov[P_1, P_1 + \theta S_2 + \epsilon_2] \\ &= Cov[P_1, P_1] + Cov[P_1, \theta S_2] + Cov[P_1, \epsilon_2] \\ &= Var[P_1] + 0 + 0 \\ &= \sigma_\epsilon^2. \end{aligned}$$

Therefore, $Var[P_1S_1 + P_2S_2] = S^T \Sigma S$ where



$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & 2\sigma_\epsilon^2 \end{pmatrix}, \ S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}.$$

In general for $N$ periods, we have

$$Var[P_k] = k\sigma_\epsilon^2, \quad k = 1, \ldots, N,$$

$$Cov[P_i, P_j] = Var[P_i] = i\sigma_\epsilon^2, \quad i < j.$$

Therefore, we get $Var[\sum_{k=1}^N P_k S_k] = S^T \Sigma S$ where

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 & \ldots & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & 2\sigma_\epsilon^2 & \ldots & 2\sigma_\epsilon^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\epsilon^2 & 2\sigma_\epsilon^2 & \ldots & N\sigma_\epsilon^2 \end{pmatrix}, \ S = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{pmatrix}.$$

Thus, for a risk averse investor with risk aversion parameter $\lambda$, we have the following quadratic programming problem

$$\begin{aligned} \min_{S} \quad & C^T S + S^T Q S + \lambda S^T \Sigma S \\ \text{subject to} \quad & \\ & e^T S = \bar{S} \\ & S \geq 0\,. \end{aligned}$$

Using a numerical example, the results of QP approach and DP are compared. The following parameter values are used

$$\theta = 5 \times 10^{-5}, \quad N = 20, \quad \sigma_\epsilon = 0.125, \quad \lambda = 0.00005.$$

The price process is simulated assuming $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The results are shown in Table 12.2. The solution for two techniques are exactly same as expected.

## 12.10 Quadratic Programming Approach for Optimal Execution: Dynamic Models

Almgren and Chriss [2] proved that optimal strategies are piecewise static i.e. trades are statical up to an event like change in market information, and then reacts explicitly to the outcome of the event. Therefore, we follow a static strategy till a new information arrives, and then recompute the strategy using a new information. In particular, if we assume that market information is updated at each period, then we need to recompute the strategy at each period using the updated information. We can make the static optimization method artificially dynamic by the following method.

- Solve $QPP_1$ which depends on initial price and market information and get the optimal strategy, $\{S_1^{(1)}, S_2^{(1)}, \ldots, S_N^{(1)}\}$.
- Implement $S_1^{(1)}$ in the first period, and reformulate a new QPP, say $QPP_2$ which depends on latest price and market information available at the end of first period. Let $\{S_2^{(2)}, S_3^{(2)}, \ldots, S_N^{(2)}\}$ be the optimal solution of $QPP_2$.
- In the second period, implement $S_2^{(2)}$ and reformulate QPP using updated information and so on.

In general, follow an optimal strategy till information is updated, and then reformulate the quadratic program. This method is illustrated in the following

| Period | P1 | S1 | W1 | P2 | S2 | W2 |
|---|---|---|---|---|---|---|
| 1 | 50.75666 | 16215 | 100000 | 50.75666 | 16215 | 100000 |
| 2 | 51.22827 | 13596 | 83785 | 51.22827 | 13596 | 83785 |
| 3 | 51.81408 | 11403 | 70189 | 51.81408 | 11403 | 70189 |
| 4 | 52.32833 | 9565 | 58786 | 52.32833 | 9565 | 58786 |
| 5 | 52.5864 | 8028 | 49221 | 52.5864 | 8028 | 49221 |
| 6 | 53.07228 | 6740 | 41193 | 53.07228 | 6740 | 41193 |
| 7 | 53.50411 | 5664 | 34453 | 53.50411 | 5664 | 34453 |
| 8 | 53.7376 | 4764 | 28789 | 53.7376 | 4764 | 28789 |
| 9 | 53.97918 | 4013 | 24025 | 53.97918 | 4013 | 24025 |
| 10 | 54.1704 | 3388 | 20012 | 54.1704 | 3388 | 20012 |
| 11 | 54.29048 | 2868 | 16624 | 54.29048 | 2868 | 16624 |
| 12 | 54.50313 | 2439 | 13756 | 54.50313 | 2439 | 13756 |
| 13 | 54.53384 | 2085 | 11317 | 54.53384 | 2085 | 11317 |
| 14 | 54.89655 | 1796 | 9232 | 54.89655 | 1796 | 9232 |
| 15 | 54.9577 | 1564 | 7436 | 54.9577 | 1564 | 7436 |
| 16 | 55.04097 | 1381 | 5872 | 55.04097 | 1381 | 5872 |
| 17 | 55.23632 | 1240 | 4491 | 55.23632 | 1240 | 4491 |
| 18 | 55.30066 | 1138 | 3251 | 55.30066 | 1138 | 3251 |
| 19 | 55.34233 | 1073 | 2113 | 55.34233 | 1073 | 2113 |
| 20 | 55.29029 | 1040 | 1040 | 55.29029 | 1040 | 1040 |

**Table 12.2.** Arithmetic random walk with linear permanent market impact model. $\{P1, S1, W1\}$ is the solution using dynamic programming. $\{P2, S2, W2\}$ is the solution using quadratic programming. The two strategies are exactly same.

section and the results are compared with those of the dynamic programming approach.

### (i) Linear Price Impact with Information Model

As discussed earlier let $X_k$ be the information variable which is modeled using $AR(1)$ process. The law of motion for execution price $P_k$ is given by

$$P_k = P_{k-1} + \theta S_k + \gamma X_k + \epsilon_k,\ \ \theta > 0 \tag{12.22}$$

$$X_k = \rho X_{k-1} + \eta_k,\ \ \rho \in (-1, 1) \tag{12.23}$$

where $\epsilon_k$ and $\eta_k$ are independent white noise process with mean 0 and variances $\sigma_\epsilon^2$ and $\sigma_\eta^2$, respectively. At time $t_1$, initial price $P_0$ and market information $X_1$ are known. Therefore, for a risk neutral investor, the optimization problem is defined as

$$
\begin{aligned}
&\underset{S_k}{\text{Min}} \qquad E_1\Big[\sum_{k=1}^{N} P_k S_k\Big] \\
&\text{subject to} \\
&\qquad \textstyle\sum_{k=1}^{N} S_k = \bar{S} \\
&\qquad S_1, \ldots, S_N \geq 0.
\end{aligned}
\tag{12.24}
$$

We can express $E_1[\sum_{k=1}^{N} P_k S_k]$ as a quadratic function (see [55]) of $S_1, \ldots, S_N$ as

$$
E_1\Big[\sum_{k=1}^{N} P_k S_k\Big] = C^T S + S^T Q S\,,
$$

where $C$ is $N \times 1$ matrix, $S$ is $N \times 1$ matrix and $Q$ is $N \times N$ matrix given by

$$
C = \begin{pmatrix} P_0 + \gamma X_1 \\ P_0 + \gamma X_1(1+\rho) \\ \vdots \\ P_0 + \gamma X_1(1 + \rho + \ldots + \rho^{N-1}) \end{pmatrix}, \quad S = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{pmatrix},
$$



$$
Q = \begin{pmatrix} \theta & \theta/2 \ldots \theta/2 \\ \theta/2 & \theta \ \ldots \theta/2 \\ \vdots & \vdots \ \ddots \ \vdots \\ \theta/2 & \theta/2 \ldots \ \theta \end{pmatrix}.
$$

We urge the readers to work out the above details. Therefore, a risk neutral investor solves the following quadratic programming problem $(QPP_1)$ at time $t_1$

$$
\begin{aligned}
&\underset{S}{\text{Min}} \qquad C^T S + S^T Q S \\
&\text{subject to} \\
&\qquad e^T S = \bar{S} \\
&\qquad S \geq 0.
\end{aligned}
\tag{12.25}
$$

Let $\{S_1^{(1)}, S_2^{(1)}, \ldots, S_N^{(1)}\}$ be optimal strategy for optimization problem (12.25). The investor acquires $S_1^{(1)}$ shares at time $t_1$. At the beginning of time $t_2$ the price and information parameters are updated. Let $P_1$ and $X_2$ be price and information available at time $t_1$. Moreover, the number of shares to be acquired at time $t_1$ are $W_2 = \bar{S} - S_1^{(1)}$. So, the investor reformulates the quadratic program $(QPP_2)$ as follows

$$\begin{aligned} \underset{S}{\text{Min}} \quad & C^T S + S^T Q S \\ \text{subject to} \quad & \\ & e^T S = W_2 \\ & S \geq 0, \end{aligned} \tag{12.26}$$

where $C$ is $(N-1)\times 1$ matrix, $Q$ is $(N-1)\times(N-1)$ matrix and $S$ is $(N-1)\times 1$ matrix given by

$$C = \begin{pmatrix} P_0 + \gamma X_2 \\ P_0 + \gamma X_2(1+\rho) \\ \vdots \\ P_0 + \gamma X_2(1+\rho+\ldots+\rho^{N-2}) \end{pmatrix}, \quad S = \begin{pmatrix} S_2 \\ S_3 \\ \vdots \\ S_N \end{pmatrix},$$

$$Q = \begin{pmatrix} \theta & \theta/2 \ldots \theta/2 \\ \theta/2 & \theta \quad \ldots \theta/2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \theta/2 & \theta/2 \ldots \quad \theta \end{pmatrix}.$$

Let $\{S_2^{(2)}, S_3^{(2)}, \ldots, S_N^{(2)}\}$ be the optimal solution of (12.26). Now, the investor acquires $S_2^{(2)}$ shares at time $t_2$ and update the remaining number of shares to be purchased at time $t_3$ to $W_3 = W_2 - S_2^{(2)}$. Using latest price and information available at time $t_2$, investor reformulates the quadratic program. Therefore, at each step, a new optimization problem is solved. At the last period, the investor acquires the remaining number of shares.

For $N = 5$, the results of all quadratic programs are shown in Table 12.3. The following parameter values are used for calculation

$$\theta = 5\times 10^{-5}, \quad \rho = 0.50, \quad \gamma = 5.0, \quad \sigma_\epsilon^2 = (0.125)^2, \quad \sigma_\eta^2 = 0.001 \tag{12.27}$$

$\epsilon_k$ and $\eta_k$ are assumed to be drawn from normal distribution.

If we follow a similar method for a static model like the basic model of Bertsimas and Lo, the number of shares to be traded in time period $t_2$ obtained from $QPP_1$

| Time | $QPP_1$ | $QPP_2$ | $QPP_3$ | $QPP_4$ | $QPP_5$ |
|---|---|---|---|---|---|
| $t_1$ | **22307** | | | | |
| $t_2$ | 20424 | **22421** | | | |
| $t_3$ | 19482 | 19600 | **19550** | | |
| $t_4$ | 19011 | 18189 | 18199 | **18457** | |
| $t_5$ | 18776 | 17483 | 17523 | 17264 | **17264** |

**Table 12.3.** Static strategy at each time instant for $N = 5$ for dynamic model. The $i^{th}$ column denotes the solution of ($QPP_i$). The number of shares to be traded in period $t_2$ obtained from ($QPP_1$) is not equal to that obtained from ($QPP_2$). Since, we assume that market information is updated at each period, the diagonal gives the final strategy.

will be same as that obtained from $QPP_2$. This can be seen in Table 12.4. The following parameter values are used for calculation

$$\theta = 5 \times 10^{-5}, \quad N = 5, \quad \lambda = 1 \times 10^{-5}, \quad \sigma_\epsilon = 0.125.$$

| Time | $QPP_1$ | $QPP_2$ | $QPP_3$ | $QPP_4$ | $QPP_5$ |
|---|---|---|---|---|---|
| $t_1$ | **20740** | 0 | 0 | 0 | 0 |
| $t_2$ | **20244** | **20244** | 0 | **0** | 0 |
| $t_3$ | **19876** | 19876 | **19876** | 0 | 0 |
| $t_4$ | **19631** | 19631 | 19631 | **19631** | **0** |
| $t_5$ | **19509** | 19509 | 19509 | 19509 | **19509** |

**Table 12.4.** Static strategy for $N = 5$ for a static model. The $i^{th}$ column denotes the solution of ($QPP_1$). The number of shares to be traded in each period $t_k$ are same for all QPP's.

The results of dynamic programming approach and static approximation approach for a risk neutral investor are compared for various cases. The parameters used are given in (12.27).

In what follows we use the abbreviations *DP* and *QP* for the dynamic programming and the quadratic programming respectively.

**Case 1 Unconstrained DP and Unconstrained QP**

The strategies given by two techniques are exactly the same as depicted in Fig. 12.4.

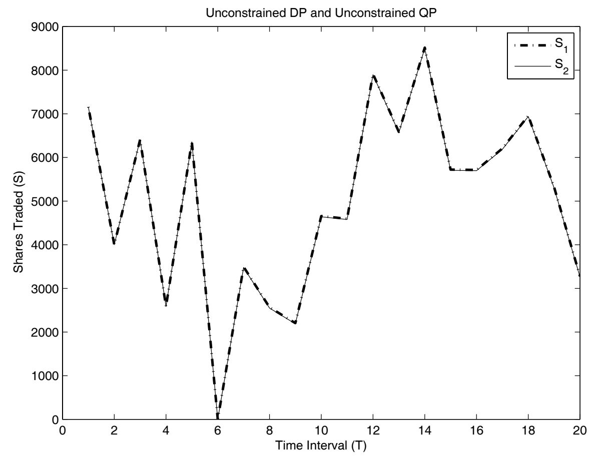



**Fig. 12.4.** Best execution strategy without constraints for $N = 20$ using DP and QP. The black dotted curve represents DP strategy while the black bold curve represents QP strategy. The two curve coincides.

## Case 2 Unconstrained DP and Constrained QP

In Fig 12.5, the optimal strategy using DP is never negative and the two strategies are exactly same. In Fig. 12.6, the two strategies follow the same trend. The strategy from DP at time $t_3, t_{22}, t_{23}, t_{24}$ are negative while it is zero from QP approach. These figures clearly indicate that static approximation procedure gives very similar strategy to dynamic programming.



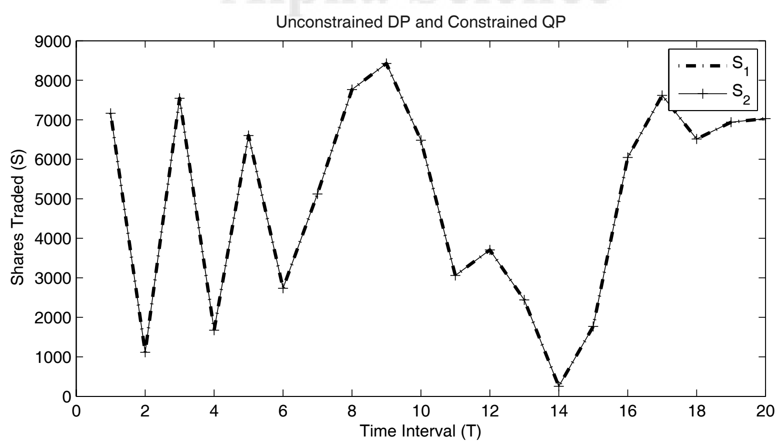



**Fig. 12.5.** Best execution strategy for $N = 20$ using DP and QP. The black dotted curve represents DP strategy without constraints while the black bold curve represents QP strategy with non-negativity constraints. $S_1$ is never negative and the two curve coincides.

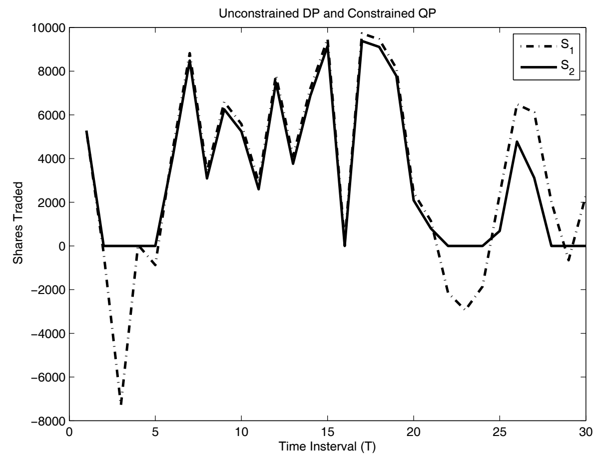



**Fig. 12.6.** Best execution strategy for $N = 30$ using DP and QP. The black dotted curve represents DP strategy with no constraints while the black bold curve represents QP strategy with non-negativity constraint. The effect of imposing non-negativity constraint is clearly visible. The strategies follow same trend, but QP strategy becomes zero as DP strategy goes negative.

### Case 3 Constrained DP and Constrained QP

Constrained DP is solved for $N = 3$ with the following parameters,

$$\bar{S} = 10000, \quad \theta = 5 \times 10^{-5}, \quad \rho = 0.50, \quad \gamma = 5.0, \quad \sigma_\epsilon^2 = (0.125)^2, \quad \sigma_\eta^2 = 0.013$$

and the results are shown in Table 12.5. The static approximation procedure gives strategy similar to one given by dynamic programming without non-negativity constraint, but the optimal strategy with non-negativity constraints is different. The optimal strategy by dynamic programming with non-negativity constraints for time $t_1$ takes into account that there is a finite probability that strategy at time $t_2$ goes negative. Therefore, optimal solution trades more at time $t_1$ as compared optimal solution without constraint.

For a risk averse investor, the quadratic programming problem at time period $t_1$ will be

$$\begin{aligned} \underset{S}{\text{Min}} \quad & C^T S + S^T Q S + \lambda S^T \Sigma S \\ \text{subject to} \quad & \\ & e^T S = \bar{S} \\ & S \geq 0 \end{aligned}$$

| Time | DP without constraint | QP with constraint | DP with constraint |
|---|---|---|---|
| $t_1$ | 2162 | 2162 | 3643 |
| $t_2$ | 861 | 861 | 120 |
| $t_3$ | 6977 | 6977 | 6237 |
| Total | 10000 | 10000 | 10000 |

**Table 12.5.** Comparison of best execution strategy for unconstrained DP, constrained QP and constrained DP.

where the matrices $C$ and $S$ are defined in ($QPP_1$) above, and $\Sigma$ is variance-covariance matrix of price $P_k$ are discussed in [55]. As before, we reformulate the quadratic program at each time period to get a dynamic strategy. Fig. 12.7 shows the best execution strategy with non-negativity constraints obtained using static approximation procedure. As $\lambda$ is increased, the strategy becomes more and more aggressive, that is, trade more in the beginning to reduce the risk. For lower values of $\lambda$, the strategy follows a similar trend as market information variable $X_k$ (black bold line with □).

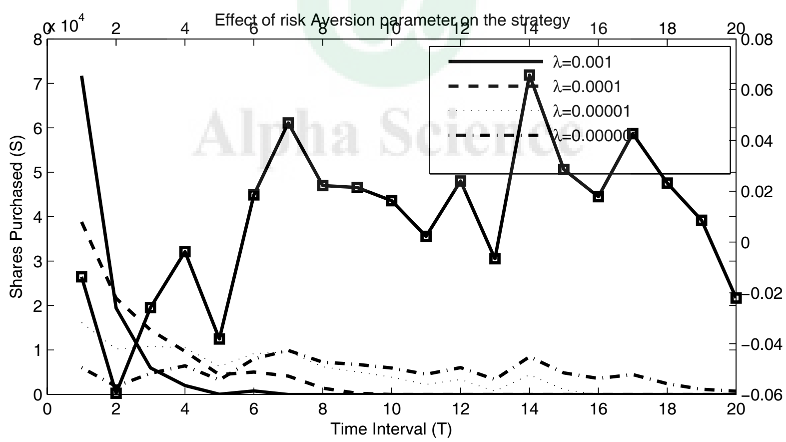



**Fig. 12.7.** Effect of $\lambda$ on best execution strategy

## (ii) Linear-Percentage Temporary Price Impact Model

In this model, the no-impact price $\tilde{P}_k$ is modelled using the geometric Brownian motion and market information is modeled using $AR(1)$ process:

$$\tilde{P}_k = \tilde{P}_{k-1} \exp(Z_k)\,, \tag{12.28}$$

$$X_k = \rho X_{k-1} + \eta_k\,, \tag{12.29}$$

where $Z_k$ are i.i.d. normal random variables with mean $u_z$ and variance $\sigma_z^2$ and $\eta_k$ is white noise with mean 0 and variance $\sigma_\eta^2$. The execution price $P_k$ at time $t_k$ is comprised of two components, the no-impact price $\tilde{P}_k$ and the price impact $\Delta_k$

$$P_k = \tilde{P}_k + \Delta_k\,, \tag{12.30}$$

$$\Delta_k = (\theta S_k + \gamma X_k)\tilde{P}_k. \tag{12.31}$$

At time $t_1$, initial price $\tilde{P}_0$ and market information $X_1$ are known. Thus, a risk neutral investor solves the following quadratic programming problem with $Q$ and $C$ defined in [55] at time $t_1$

$$\begin{aligned} \underset{S}{\text{Min}} \quad & C^T S + S^T Q S \\ \text{subject to} \quad & \\ & e^T S = \bar{S} \\ & S \geq 0\,, \end{aligned} \tag{12.32}$$



where

$$C = \begin{pmatrix} q\tilde{P}_0(1+\gamma X_1) \\ q^2\tilde{P}_0(1+\gamma\rho X_1) \\ \vdots \\ q^N\tilde{P}_0(1+\rho^{N-1}\gamma X_1) \end{pmatrix}, \quad S = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{pmatrix},$$

$$Q = \begin{pmatrix} \theta q\tilde{P}_0 & 0 & \dots & 0 \\ 0 & \theta q^2\tilde{P}_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta q^N\tilde{P}_0 \end{pmatrix}.$$

In this model **Q** is a diagonal matrix. This is due to the fact that we consider only temporary market impact. The number of shares traded at time $t_1$ does not affect the future price. Therefore, the execution price $P_2$ for shares traded at time

$t_2$ is independent of number of shares traded at time $t_1$. As before, market information is updated at each time period, so we need to reformulate the quadratic program at each period.

The strategy obtained using DP approach and QP approach can be compared using numerical example with following parameters that were used by Bertsimas and Lo [11].

$$\mu_z = 0, \quad \sigma_z = 0.02/\sqrt{13}, \quad \theta = 5 \times 10^{-7}, \quad \gamma = 0.01,$$

$$\rho = 0.50, \quad \eta \sim \mathcal{N}(0, 1-\rho^2), \quad \sigma_\eta^2 = 1 - \rho^2.$$

For $N = 5$, the strategies for all the quadratic programs are shown in Table 12.6.

| Time | $QPP_1$ | $QPP_2$ | $QPP_3$ | $QPP_4$ | $QPP_5$ |
|---|---|---|---|---|---|
| $t_1$ | **28703.59** | 0 | 0 | 0 | 0 |
| $t_2$ | 24075.95 | **21608.49** | 0 | 0 | 0 |
| $t_3$ | 16223.69 | 18198.27 | **18053.14** | 0 | 0 |
| $t_4$ | 12415.06 | 15646.57 | 15251.64 | **16267.66** | 0 |
| $t_5$ | 18581.7 | 18581.7 | 15350.2 | 13770.54 | **15367.11** |



**Table 12.6.** Dynamic strategy for $N = 5$ for Model 3. The $i^{th}$ column denotes the solution of ($QPP_i$). The number of shares to be traded in period $t_2$ obtained from ($QPP_1$) is not equal to that obtained from ($QPP_2$). For each period, we use the latest market information, and solve the corresponding QP and follow this strategy till new information comes. Since, we assume that market information is updated at each period, the diagonal gives the number of shares to trade in that period.

In Fig. 12.8, the two strategies obtained using DP and QPP approach are exactly same as the optimal strategy using DP is never negative whereas in Fig. 12.10, the two strategies differ.

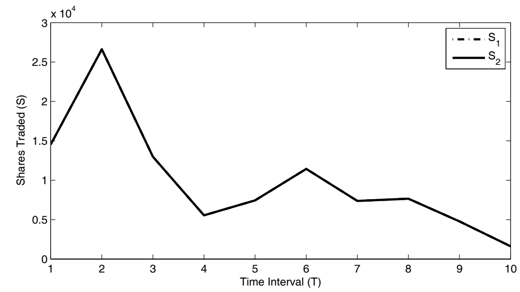



**Fig. 12.8.** Best execution strategy for $N = 10$ for Model 3 (Linear percentage temporary price impact) using DP and QP. The black dotted curve represents DP strategy with no constraints. The black bold curve represents QP strategy with non-negativity constraint. Since $S_1$ is never negative, the two curve coincides.

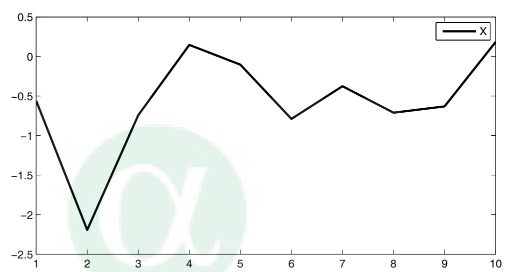



**Fig. 12.9.** Market information, AR(1) process



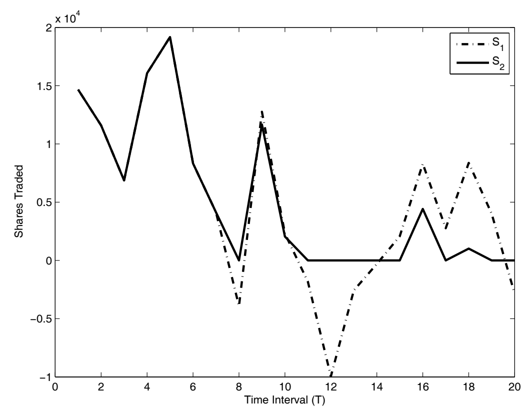



**Fig. 12.10.** Best execution strategy for $N = 10$ for Model 3 (Linear percentage temporary price impact) using DP and QP. The black dotted curve represents DP strategy no constraints. The black bold curve represents QPP strategy with non-negativity constraint.

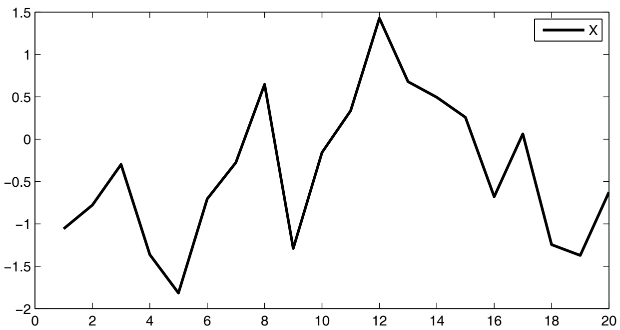



**Fig. 12.11.** Market information, AR(1) process

## 12.11 Summary and Additional Notes

- Section 12.2 introduces the basics of market structure. Here various components of transaction cost are described. Further Section 12.3 presents various types of orders in the market are introduced. Readers may refer to Haris [59] and, Kissell and Glantz [77] for more details in this regard.
- The problem of optimal execution is presented in Section 12.4. The whole discussion in this chapter centers around the study of this problem. There are two major contributions in this area. These are due to Bertsimas and Lo [11] and Almgren and Chriss [2]. We have presented the dynamic programming approach of Bertsimas and Lo [11] in detail, both for risk neutral and risk averse trader. While static models are discussed in Sections 12.5 and 12.6, the dynamic models are presented in Section 12.7. Section 12.8 also discussed the limitations of dynamic programming.
- A quadratic programming approach for optimal execution is presented in Section 12.9 and Section 12.10. The contents of these two sections are based on Gupta and Choudhary [55] and Khemchadani et al [76].
- Algorithmic trading in general, and optimal trading strategies in particular have now become very important in trading scenario. In this chapter we have illustrated the applications of dynamic programming and quadratic programming for determining optimal trading strategies. Various machine learning methodologies have also found applications in this area. One of these is *reinforcement learning*. The readers may refer to Sutton and Barto [132], Kaelbling

and Littman [73] for basics of reinforcement learning. Nevmyvaka et al. [101] provides a good application of reinforcement learning in the area of optimal trading strategies.

- This chapter presents optimal trading strategies in the context of a single asset only. But a more realistic case in that of a portfolio. Bertsimas et al. [12], and Aitsahlia et al. [1] discuss optimal execution for portfolios.
- Similar to the concept of efficient frontier in the Markowitzs' model of portfolio optimization, Kissell and Glantz [77] introduced the concept of *efficient trading frontier* in the context of optimal trading strategies.

## 12.12 Exercises

**Exercise 12.1** *Compute the associated transaction cost of a buy order for 100,000 shares of ABC if the price at the beginning of trading was Rs 50 per share and the average execution price was Rs 52 per share and all shares were executed.*

**Exercise 12.2** *Consider the set of parameters,*

$$\bar{S} = 10000, \quad \theta = 5 \times 10^{-5}, \quad \rho = 0.50, \quad \gamma = 5.0, \quad \sigma_{\epsilon}^2 = (0.125)^2, \quad \sigma_{\eta}^2 = 0.013$$

*Plot the efficient frontier for various values of $\lambda$ lies between 0 and 1 using all the price process mentioned in the chapter.*

# 13
# Credit Risk Management

## 13.1 Introduction

Credit risk is the risk of a trading partner for not meeting its obligations in full on the due date or at any time thereafter. Such an event is called a default, hence another term for *credit risk* is *default risk* . There are three well known types of credit risk: *default risk/counterparty risk*, *credit spread risk* and *downgrade risk.* There are several ways through which credit risk can be taken into account within commercial and retail banking activities. To name some, deposits, mortgage lending and credit cards are part of retail banking, whereas commercial banking includes loans, letters of credit and asset finance.

For example, in a credit card issuance financial service/bank agrees to pay retailers for purchases made by credit card holders in exchange for an unsecured promise to repay the card balance. Thus, any given debt has a short expected life, as when card holders pay off some of their balances and then buy new goods, new exposures roll in to place of old ones.

Financial services/banks need a mechanism to quantify the risk factors relevant for an obligor's ability and willingness to pay. Credit scoring has been used as a tool in modern banking to access the risk factors due to the large number of applications received on a daily basis and the increased regulatory requirements for banks. In this chapter, we shall present some of the basic concepts with emphasis on statistical and data mining algorithms, and their application to *credit scoring.*

Risk management is a core activity conducted by banks, insurance and investment companies, or any financial institution that evaluate risk due to losses. Losses can result from either counter-party default, or from a decline in market value stemming from the credit quality migration of an issuer or counter-party.

There are two primary types of models that attempt to describe default processes in the credit risk literature. These are (i) *structural models* and (ii) *reduced* *form models*. Structural models have been pioneered by Black-Scholes and Merton. The basic idea, common to all structural type models is that a company defaults on its debt if the value of the assets of a company falls below a certain default point. For this reason, these models are also known as *firm-value models*. In these models it has been demonstrated that default can be modeled as an option and, as a result, researchers have been able to apply the same principles used for option pricing to the valuation of risky corporate securities.

The second group of credit models, known as reduced form models, are more recent. These models, most notably the Jarrow-Tunbull and Duffie and Singleton models, do not look inside the firm. Instead, they model directly the likelihood of default or downgrade. Apart from modeling the current probability of default, some researchers attempt to model a '*forward curve*' of default probabilities. This can be used to price instruments of varying maturities.

In recent past, managing credit risk at portfolio level has influenced many researchers. In order to utilize the advantage of the credit portfolio one should know the risk of portfolio and the factors that affect the portfolio risk profile. Thus portfolio managers are interested in knowing the effect of changing the portfolio mix, how would risk-based pricing at the individual contract and the portfolio level be influenced by the level of expected losses and credit risk capital?

Traditionally used tools for assessing and optimizing market risk assume that the portfolio return-loss is normally distributed. With this assumption, the two statistical measures, mean and standard deviation, could be used to balance return and risk. However, to cope with skewed return-loss distributions, *conditional value-at-risk* (CVaR) has been introduced as the risk measure. This measure is also known as *mean excess loss*, *mean shortfall*, or *tail* VaR. By definition, $(1-\beta)$-CVaR is the expected loss exceeding $(1-\beta)$-*value-at-risk* (VaR), i.e. it is the mean value of the worst $\beta \times 100\%$ losses. For instance, at $(1-\beta) = 0.95$, or equivalently $\beta = 0.05$, CVaR is the average of the $5\%$ worst losses.

In this chapter we focus on the quantification of credit risk. We will discuss different approaches to estimate the probability that a company will default. In the subsequent sections we will explain how a bank or other financial institution can estimate its loss given that the default has occurred. Last part of the chapter also covers credit rating migration and default correlation, and application of machine learning techniques in the credit scoring scenario.

## 13.2 Basic Terminology

To begin with we shall discuss some basic definitions that would help us to understand the quantification of credit risk and procedure of calculating the probability that company would default.

A valuation method is used to estimate the attractiveness of an investment opportunity. *Discounted cash flow* (DCF) analysis uses future free cash flow projections and discounts them (most often using the weighted average cost of capital) to arrive at a present value, which is used to evaluate the potential for investment. If the value arrived at through DCF analysis is higher than the current cost of the investment, the opportunity may be a good one.

**Definition 13.2.1 (Discounted Cash Flow)** *Let, for* $i = 1, 2 \ldots n$, $CF_i$ *denote the cash flow for the i-th period (say year) and* $r$ *be the continuing compounded interest rate. Then the* discounted cash flow *(DCF) is defined as*

$$DCF = \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \ldots + \frac{CF_n}{(1+r)^n} \,.$$

**Definition 13.2.2 (Credit Premium)** *The credit premium is the discounted value of cash flows, when there is probability of default. It is given by*

$$CP = \frac{CF_1 q_1}{(1+r)^1} + \frac{CF_2 q_2}{(1+r)^2} + \ldots + \frac{CF_n q_n}{(1+r)^n} \,,$$

*where* $q_i$ *denotes the probability that the counter party is solvent (not refutable) at the* $i^{th}$ *period and* $CF_i$ *and* $r$ *are defined as above.*

The credit/default premium is paid by companies with lower grade bonds or by individuals with poor credit. As an illustration, companies with poor financial state will tend to compensate investors for the additional risk by issuing bonds with high yields. Individuals with poor credit must pay higher interest rates in order to borrow money from the bank.

**Definition 13.2.3 (Credit Spread)** *Credit spread is the yield spread, or difference in yield between different securities, due to different credit quality. The credit spread reflects the additional net yield an investor can earn from a security with more credit risk relative to one with less credit risk. The credit spread of a particular security is often quoted in relation to the yield on a credit risk-free benchmark security or reference rate, typically either U.S. Treasury bonds or LIBOR.*

For instance, the *credit spread* could be the difference between yields on government bonds and those on single A-rated industrial bonds. A company must offer a higher return on their bonds because their credit is worse in comparison to government bonds.

**Definition 13.2.4 (Recovery Rate: Loss Given Default)** *Recovery rate is defined as the proportion of the "claimed amount" received in the the event of a default Loss-given-default is the percentage we expect to lose when default occurs. Obviously R=1-LGD.*

When default occurs, a portion of the value of the portfolio can usually be recovered. Because of this, a recovery rate is always considered when evaluating credit losses. It represents the percentage value which we expect to recover, given default.

**Definition 13.2.5 (Credit Exposure).** *It is the maximum loss that a portfolio experience at the time of default taken with a certain confidence level.*

**Definition 13.2.6 (Default Probability)** *Probability of default is the likelihood that a loan will not be repaid.*



A bank assigns to every customer a default probability (DP); a loss fraction called the loss given default (LGD); describing the fraction of the loans exposure expected to be lost in case of default, and the exposure at default (EAD) expected to be lost in the considered time period. The loss of any obligor is then defined by a loss variable $\tilde{L} = EAD \times LGD \times L$ with $L = 1_D \times P(D) = DP$, where $D$ denotes the event that the obligor defaults in a certain period of time (most often one year), $1_D$ is the characteristic function of $D$ and $P(D)$ denotes the probability of $D$.

The task of assigning a default probability to every customer in a bank credit portfolio is far from being easy. There are essentially two approaches to compute default probabilities: calibration of default probabilities from ratings and calibration of default probabilities from market data. We intend to discuss these in the subsequent sections.

## 13.3 Expected Default Losses on Bonds

The first step in estimating default probabilities from bond prices is to calculate the expected default losses on corporate bonds of different maturities. This involves comparing the price of a corporate bond with the price of a risk-free bond that has the same maturity and pays the same coupon. The usual assumption is that the present value of the cost of defaults equals the excess of the price of the risk free bond over the price of the corporate bond.

**Probability of Default Assuming no Recovery**

Let $y(T)$ be the yield on a $T$ year corporate zero-coupon bond, and $y^*(T)$ be the yield on a $T$ year risk free zero-coupon bond. Let $D(T)$ be the probability that corporation will default between time zero and time $T$.

The value of a $T$ year risk free zero coupon bond with principal 100 is $100e^{-y^*(T)T}$ while the value of similar corporate bond is $100e^{-y(T)T}$. The expected loss from default is therefore $100(e^{-y^*(T)T} - e^{-y(T)T})$.

If we assume that there is no recovery in the event of default, the calculation of $D(T)$ is relatively easy. There is a probability $D(T)$ that the corporate bond will be worth zero at maturity and a probability $1 - D(T)$ that it will be worth Rs 100. The value of the bond is therefore $\{[D(T) \times 0] + [(1 - D(T)) \times 100]e^{-y^*(T)T}\}$. The yield on the bond is $y(T)$, so that

$$100e^{-y(T)T} = 100(1 - D(T))e^{-y^*(T)T}.$$

The above equation gives

$$D(T) = \frac{e^{-y^*(T)T} - e^{-y(T)T}}{e^{-y^*(T)T}}$$

i.e.

$$D(T) = 1 - e^{-[y(T) - y^*(T)]T}.$$



**Probability of Default Assuming Recovery**

Suppose that the claimed amount of the bond is the no-default value of the bond. In the event of a default, the bondholder receives a proportion $R$ of the bond's no-default value. While in no default case, the bondholder receives Rs 100. The bond's no-default value is $100(e^{-y^*(T)T})$ and the probability of default is $D(T)$. The value of bond is therefore

$$\{[D(T) \times 100Re^{-y^*(T)T}] + [(1 - D(T)) \times 100]e^{-y^*(T)T}\}.$$

The yield on the bond is $y(T)$, hence

$$100e^{-y(T)T} = \{[D(T) \times 100Re^{-y^*(T)T}] + [(1 - D(T)) \times 100]e^{-y^*(T)T}\},$$

which on simplification gives

$$D(T) = \frac{1 - e^{-[y(T) - y_*(T)]T}}{1 - R} \, .$$

The above estimation of default probability is based on the following set of assumptions

(i) Amount claimed in the event of default equals the no-default value of the bond.
(ii) Claim made in the event of default equals the bond's face value plus accrued interest.
(iii) Probabilities are calculated on zero coupon bonds.

Over the time, the bonds are liable to move from one rating category to another. This is sometime referred as credit rating migration. Rating agencies generate, from historical data, a rating transition matrix whose entries correspond to percentage probability of a bond moving from one rating to another during a certain period of time. Next we present an approach termed *CreditMetrics* proposed for estimating risks associated with default.

## 13.4 CreditMetrics Analysis

JP Morgan first published and well publicized *CreditMetrics* model in 1997 [99]. CreditMetrics computes the forward distribution of values of the loan portfolio, given that the transition probabilities are estimated from historical data, and that the correlation between the behavior of the rating classes are directly obtained from the stock market data.

CreditMetrics has several aims; two of which are the creation of a benchmark for measuring credit risk and the increase in market liquidity. If the former aim is achieved then it becomes possible to measure risks systematically across instruments and, at the very least, to make relative value judgments. From this follows the second aim. Once instruments, and in particular the risks associated with them, are better understood they would appear less scary to investors and hence would help in promoting liquidity.

The CreditMetrics data set can be freely downloaded from www.jpmorgan.com. This website also contains detailed study of the CreditMetrics methodology. The data set consists of four data types: yield curves, spreads, transition matrices and correlations. The CreditMetrics yield curve data set consists of the risk-free yield to maturity for major instruments. It contains yields for maturities of one, two, three, five, seven, ten and thirty years. For each credit rating, the data set gives the spread above the riskless yield for each maturity. It is generally observed that riskier the bond higher the yield; thus the yield on the BBB bond is higher than that on the AA bond which is in turn higher than the risk-free yield. Thus higher yield for risky bonds is compensation for the possibility of not receiving future coupons or the principal.

In the CreditMetrics framework, the transition matrix has its entries as the probability of a change of credit rating at the end of a given time horizon, for example the probability of a upgrade from AA to AAA might be 5.5%. The time horizon for the CreditMetrics data set is one year. Unless the time horizon is very long, the highest probability is typically for the bond to remain at its initial rating. In the risk-free yield,the spreads and the transition matrix, contain sufficient information for the CreditMetrics method to derive distributions for the possible future values of a single bond. However, when it comes to examine the behavior of a portfolio of risky bonds, we must consider whether there is any relationship between the re-rating or default of one bond to another. In other words, are bonds issued by different companies or governments are some sense correlated? This is where the CreditMetrics correlation data set comes in. This data set gives the correlations between major indices in many countries.

The CreditMetrics methodology is about calculating the possible values of a risky portfolio comprising of bonds at some time in future (the time horizon) and estimating the probability of occurrence of default. There are many credit rating agencies who compile data on individual companies or countries and estimate the likelihood of default. The most famous of these are Standard & Poor's and Moody's. These agencies assign a credit rating or grade to firms as an estimate of their creditworthiness. Standard & Poor's rate businesses as one of AAA, AA, A, BBB, BB, B, CCC or Default. Moody's use Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C. Both these companies also have finer grades within each of these primary categories. The Moody's grades are described in the Table 13.1 below.

Transition matrices, refer to Table 13.2, measure rating movements over time. The diagonal of a transition matrix represents the share of ratings that remain unchanged during the course of the reference period of one year. For instance, 84.67% of issuers rated in the 'A' category at the beginning of 2009 were still rated 'A' by year-end, compared with only 69.34% of issuers in the 'B' category (see Table 13.2). Here NR refers to nondefault rate. The probabilities are based on historical data and therefore real world probabilities.

For the illustration of the concepts we shall be using the transition matrix provided by Standard & Poor's as on 15th March 1996, see Table 13.3.

The credit rating agencies continuously gather data on individual firms and de-

| | The Meaning of Moody's Ratings |
|---|---|
| Aaa | -Bonds of best quality. Smallest degree of risk. |
| | -Interest payments protected by a large or stable margin. |
| Aa | -High quality. Margin of protection lower than Aaa. |
| A | -Many favorable investment attributes. |
| | -Adequate security of principal and interest. |
| | -May be susceptible to impairment in the future. |
| Baa | -Neither highly protected nor poorly secured. |
| | -Adequate security for the present. |
| | -Lacking outstanding investment characteristics. |
| | -Speculative features. |
| Ba | Speculative elements. |
| | -Future not well assured. |
| B | Lack characteristics of a desirable investment. |
| Caa | Poor standing. |
| | -May be in default or danger with respect to principal or interest. |
| Ca | High degree of speculation. Often in default. |
| C | Lowest-rated class. |
| | -Extremely poor chance of ever attaining any real investment standing. |

**Table 13.1.** Moody's Categories

| Initial Rating | AAA | AA | A | BBB | BB | B | CCC/C | D | NR |
|---|---|---|---|---|---|---|---|---|---|
| 2009 | | | | | | | | | |
| AAA | 87.65 | 8.64 | 0 | 0 | 0 | 0 | 0 | 0 | 3.7 |
| AA | 0 | 86.17 | 15.96 | 0.64 | 0.21 | 0 | 0 | 0 | 7.02 |
| A | 0 | 0.36 | 84.67 | 7.74 | 0.43 | 0.29 | 0 | 0.21 | 6.3 |
| BBB | 0 | 0 | 2 | 83.71 | 5.94 | 0.8 | 0.2 | 0.53 | 6.81 |
| BB | 0 | 0 | 0 | 3.09 | 72.95 | 11.48 | 0.6 | 0.7 | 11.18 |
| B | 0 | 0 | 0.16 | 0 | 2.29 | 69.34 | 8.42 | 10.14 | 9.65 |
| CCC/C | 0 | 0 | 0 | 0 | 0 | 6.32 | 27.37 | 48.42 | 17.89 |

**Table 13.2.** One year (2009) transition matrix of percentage probabilities

| Initial Rating | AAA | AA | A | BBB | BB | B | CCC | D |
|---|---|---|---|---|---|---|---|---|
| 1996 | | | | | | | | |
| AAA | 90.81 | 8.33 | 0.68 | 0.06 | 0.12 | 0 | 0 | 0 |
| AA | 0.7 | 90.65 | 7.79 | 0.64 | 0.06 | 0.14 | 0.02 | 0 |
| A | 0.09 | 2.27 | 91.05 | 5.52 | 0.74 | 0.26 | 0.01 | 0.06 |
| BBB | 0.02 | 0.33 | 5.95 | 86.93 | 5.3 | 1.17 | 1.12 | 0.18 |
| BB | 0.03 | 0.14 | 0.67 | 7.73 | 80.53 | 8.84 | 1 | 1.06 |
| B | 0 | 0.11 | 0.24 | 0.43 | 6.48 | 83.46 | 4.07 | 5.2 |
| CCC | 0.22 | 0 | 0.22 | 1.3 | 2.38 | 11.24 | 64.86 | 19.79 |

**Table 13.3.** One year (1996) transition matrix of percentage probabilities

pending on the information, grade/regrade a company according to well-specified criteria. A change of rating is called a *migration* and has an important effect on the price of bonds issued by the company. Migration to a higher rating will increase the value of a bond and decrease its yield, since it is seen as being less likely to default. Clearly there are two stages involved in modeling risky bonds under the credit-rating scenario. First, we must model the migration of the company from one grade to another and second we must price bonds taking into account the migration factor. The main objective of the CreditMetrics framework is to produce a consistent forward distribution of changes in value of the corporate bond for an arbitrary one year horizon.



## 13.5 Credit VaR and Credit CVaR

CreditMetrics/CreditVaR are methodologies based on the estimation of the forward distribution of the changes in value of a portfolio of loan and bond type products at a given time horizon, usually 1 year. The changes in value are related to eventual migrations in credit quality of the obligor, both up and downgrades, as well as default.

Unlike to market-VaR, Credit-VaR posses two new challenging difficulties. First, the portfolio distribution is far from being normal, and second, measuring the portfolio effect due to credit diversification which is much more complex than for market risk. While it is legitimate to assume normality of the portfolio changes due to market risk, it is no longer valid for the case for credit returns which are by nature highly skewed and fat-tailed. Indeed, there is limited upside to be expected from any improvement in credit quality, while there is substantial downside consecutive to downgrading and default. The percentile levels of the distribution cannot be any longer estimated from the mean and variance only. The calculation of VaR for credit risk requires simulating the full distribution of the changes in portfolio value. This, together with lack of historical data to estimate credit correlations, posses significant modeling challenges compared to market risk modeling.

To measure the effect of portfolio diversification we need to estimate the correlations in credit quality changes for all pairs of obligors. But these correlations are not directly observable. CreditMetrics/CreditVaR base their evaluation on the joint probability of asset returns, which itself results from strong simplifying assumptions on the capital structure of the obligor, and on the generating process for equity returns. This is clearly a key feature of CreditMetrics/CreditVaR on which we will elaborate in the sequel.

The first step is to specify a rating system, with rating categories, together with the probabilities of migrating from one credit quality to another over the credit risk horizon. The details have been discussed in the previous section. We next present an example that explain the steps in calculating CreditVaR.

### CreditVaR of the Bond

Let the problem be to find CreditVaR for a senior unsecured BBB rated bond maturing exactly in 5 years and paying an annual coupon of 6%. Following are the steps involved in calculating Credit-VaR.

**Step 1: Specify the transition matrix.**

The bond issuer has currently a BBB rating, 13.4 shows the probabilities estimated by Standard & Poors for a BBB issuer to be, in 1 year from now, in one of the 8 possible states, including default. Obviously, the most probable situation is for the obligor to stay in the same rating category, i.e. BBB, with a probability of 86.93%. The probability of the issuer defaulting within 1 year is only 0.18%, while the probability of being upgraded to AA is also very small, i.e. 0.33%. Such transition matrix is produced by the rating agencies for all initial ratings. Default is an absorbing state, i.e. an issuer who is in default stays in default.

**Step 2: Specify the credit risk horizon.**

The risk horizon is usually 1 year, and is consistent with the transition matrix shown in Table 13.4. But this horizon is purely arbitrary, and is mostly dictated by the availability of the accounting data and financial reports processed by the rating agencies.

**Step 3: Specify the forward pricing model.**

The valuation of a bond is derived from the zero-curve corresponding to the rating of the issuer. Since there are seven possible credit qualities, seven "spread" curves are required to price the bond in all possible states, all obligors within the same rating class being marked-to-market with the same curve. The spot zero curve is used to determine the current spot value of the bond. The forward price of the bond in 1 year from now is derived from the forward zero-curve, 1 year ahead, which is then applied to the residual cash flows from year one to the maturity of the bond. Table 13.4 gives the 1-year forward zero-curves for each credit rating.

| Category | Year 1 | Year 2 | Year 3 | Year 4 |
|---:|---:|---:|---:|---:|
| AAA | 3.60 | 4.17 | 4.73 | 5.12 |
| AA | 3.65 | 4.22 | 4.78 | 5.17 |
| A | 3.72 | 4.32 | 4.93 | 5.32 |
| BBB | 4.1 | 4.67 | 5.25 | 5.63 |
| BB | 5.55 | 6.02 | 6.78 | 7.27 |
| B | 6.05 | 7.02 | 8.03 | 8.52 |
| CCC | 15.05 | 15.02 | 14.03 | 13.52 |

**Table 13.4.** Source: CreditMetrics, JP Morgan. One year forward zero curve for each credit rating (%)



The 1-year forward price of the bond, if the obligor stays BBB, is then

$$V_{BBB} = 6 + \frac{6}{1.0410} + \frac{6}{(1.0467)^2} + \frac{6}{(1.0525)^3} + \frac{106}{(1.0563)^4} = 107.55 .$$

If we replicate the same calculations for each rating category we obtain the values shown in Table 13.5

**Step 4: Derive the forward distribution of the changes in portfolio value.**

The distribution of the changes in the bond value, at the 1-year horizon, due to an eventual change in credit quality is shown in the Table 13.6. The first entry in the last column of Table 13.6 is obtained as $109.37 - 107.55 = 1.82$. The other enterers in this column are obtained in a similar manner. This distribution exhibits long downside tails. The first percentile of the distribution of $\triangle V$, which corresponds to CreditVaR at the 99% confidence level is $83.64 - 107.55 = -23.91$. It is much lower if we compute the first percentile assuming a normal distribution for $\triangle V$. In that case CreditVaR at the 99% confidence level would be only $-7.43$.

| Year-end rating Value | ($) |
|---|---|
| AAA | 109.37 |
| AA | 109.19 |
| A | 108.66 |
| BBB | 107.55 |
| BB | 102.02 |
| B | 98.10 |
| CCC | 83.64 |
| Default | 51.13 |

**Table 13.5.** Source: CreditMetrics, JP Morgan. One year forward values for a BBB bond

Let the mean $m$, and the variance $\sigma^2$, of the distribution for $\triangle V$ are (from the Table 13.5)

$$\begin{aligned} m &= mean(\triangle V) \\ &= \sum_i p_i \triangle V_i \\ &= 0.02\% \times 1.82 + 0.33\% \times 1.64 + \ldots + 0.18\% \times (-56.42) \\ &= -0.46\,. \\ \sigma^2 &= Var(\triangle V) \\ &= \sum_i p_i (\triangle V_i - m)^2 \\ &= 0.02\%(1.82+0.46)^2 + 0.33\%(1.64+0.46)^2 + \ldots + 0.18\%(-56.42+0.46)^2 \\ &= 8.95\,, \end{aligned}$$



and $\sigma = 2.99$. The first percentile of a normal distribution $\mathcal{N}(m, \sigma^2)$ is $(m - 2.33\sigma)$, i.e. -7.43.

**Credit-VaR for a Loan or Bond Portfolio**

First, consider a portfolio composed of two bonds with an initial rating of BB and A, respectively. Given the transition matrix shown in Table 13.3, and assuming no correlation between changes in credit quality, we can derive easily the joint migration probabilities shown in Table 13.7. Each entry is simply the product of the transition probabilities for each obligor. For example, the joint probability that obligor one and obligor two stay in the same rating classes is 80.53% $\times$ 91.05%=73.32% where 80.53% is the probability that obligor one keeps

| Year-end rating | Probability p (%) | Forward price: V (Rs) | $\triangle V(Rs)$ |
|---|---|---|---|
| AAA | 0.02 | 109.37 | 1.82 |
| AA | 0.33 | 109.19 | 1.64 |
| A | 5.95 | 108.66 | 1.11 |
| BBB | 86.93 | 107.55 | 0 |
| BB | 5.30 | 102.02 | -5.53 |
| B | 1.17 | 98.10 | -9.45 |
| CCC | 0.12 | 83.64 | -23.91 |
| Default | 0.18 | 51.13 | -56.42 |

**Table 13.6.** Distribution of the bond values, and changes in value of a BBB bond, in 1 year

his current rating BB, and 91.05% is the probability that obligor two stays in rating class A.

However, this table is not very useful in practice when we need to assess the diversification effect on a large loan or bond portfolio. Indeed, the actual correlations between the changes in credit quality are different from zero. Correlations are expected to be higher for firms within the same industry or in the same region, than for firms in unrelated sectors. In addition, correlations vary with the relative state of the economy in the business cycle. If there is a slowdown in the economy, or a recession, most of the assets of the obligor will decline in value and quality, and the likelihood of multiple defaults increases substantially. The contrary happens when the economy is performing well as default correlations go down. Thus, we cannot expect default and migration probabilities to stay stationary over time. There is clearly a need for a structural model that bridges the changes of default probabilities to fundamental variables whose correlations stay stable over time. Furthermore, for the sake of simplicity, CreditMetrics/CreditVaR have chosen the equity price as a proxy for the asset value of the firm that is not directly observable. This is another strong assumption in CreditMetrics that may affect the accuracy of the method. CreditMetrics estimates the correlations between the equity returns of various obligors, the model then infers the correlations between changes in credit quality directly from the joint distribution of equity returns.

The proposed framework initially developed by Merton [93] is the option pricing approach to the valuation of corporate securities. The firm's assets value, $V_t$, is assumed to follow a standard geometric Brownian motion, i.e.

$$V_t = V_0 \exp\left\{\left(\mu - \sigma^2/2\right) t + \sigma \sqrt{t} Z_t\right\}, \tag{13.1}$$

with $Z_t \sim N(0,1), \mu$ and $\sigma^2$ being respectively the mean and variance of the in-

| | | AAA | AA | A | BBB | BB | B | CCC | Default |
|---|---|---|---|---|---|---|---|---|---|
| | A | 0.09 | 2.27 | 91.05 | 5.52 | 0.74 | 0.26 | 0.01 | 0.06 |
| | BB | | | | | | | | |
| AAA | 0.03 | 0 | 0 | 0.03 | 0 | 0 | 0 | 0 | 0 |
| AA | 0.14 | 0 | 0 | 0.13 | 0.01 | 0 | 0 | 0 | 0 |
| A | 0.67 | 0 | 0.02 | 0.61 | 0.4 | 0 | 0 | 0 | 0 |
| BBB | 7.73 | 0.01 | 0.18 | 7.04 | 0.43 | 0.06 | 0.02 | 0 | 0 |
| BB | 80.53 | 0.07 | 1.83 | 73.32 | 4.45 | 0.6 | 0.2 | 0.01 | 0.05 |
| B | 8.84 | 0.01 | 0.2 | 8.05 | 0.49 | 0.07 | 0.02 | 0 | 0 |
| CCC | 1 | 0 | 0.02 | 0.91 | 0.06 | 0.01 | 0 | 0 | 0 |
| Default | 1.06 | 0 | 0.02 | 0.97 | 0.06 | 0.01 | 0 | 0 | 0 |

**Table 13.7.** Joint migration probabilities (%) with zero correlation for two issuers rated BB and A

stantaneous rate of return on the assets of the firm. $V_t$ is lognormally distributed with expected value $E(V_t) = V_0 \exp\{\mu t\}$ at time $t$. It is further assumed that the firm has a very simple capital structure, financed only by equity $S_t$ and a single zero-coupon debt instrument maturing at time $T$ with face value $F$ and current market value $B_t$. In this framework, default only occurs at maturity of the debt obligation when the value of the assets is less than the promised payment $F$ to the bond holders.

Merton's model is extended by CreditMetrics to include changes in credit quality as illustrated in Fig. 13.1. This generalization consists of slicing the distribution of asset returns into bands in such a way that, if we draw randomly from this distribution, we reproduce exactly the migration frequencies shown in the transition matrix mentioned in Table 13.7. Fig. 13.1 shows the distribution of the normalized assets rates of return, 1 year ahead with mean zero and variance one. The credit rating thresholds correspond to the transition probabilities in Table 13.3 for a BB rated obligor. The right tail of the distribution on the righthand side of $Z_{AAA}$ corresponds to the probability for the obligor of being upgraded from $BB$ to $AAA$, i.e. 0.03%. Then, the area between $Z_{AA}$ and $Z_{AAA}$ corresponds to the probability of being upgraded from $BB$ to $AA$, etc. The left tail of the distribution, on the left-hand side of $Z_{CCC}$, corresponds to the probability of default, i.e. 1.06%. Table 13.8 shows the transition probabilities for two obligors rated $BB$ and $A$, respectively, and the corresponding credit quality thresholds. This generalization of Merton's model is quite easy to implement. It assumes that the log-returns over any period of time are normally distributed with mean 0 and variance 1, and it is the same for all obligors within the same rating category.

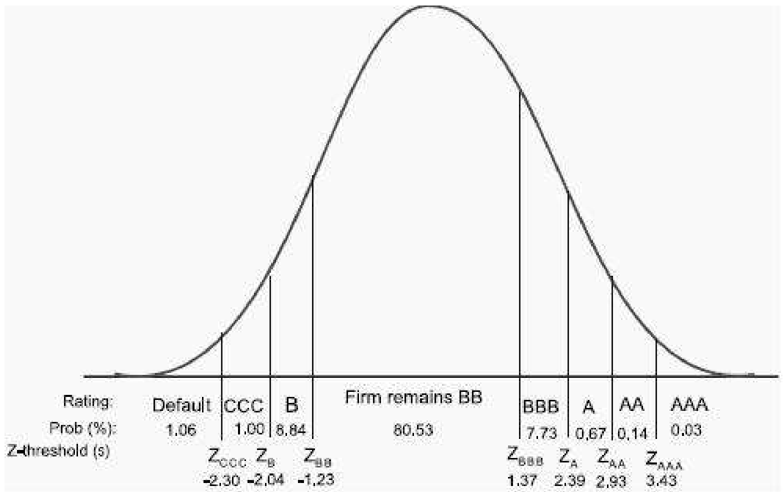



**Fig. 13.1.** Generalization of the Merton model to include rating changes

| | Rated-A obligor | | Rated-BB obligor | |
|---|---|---|---|---|
| Rating in 1 year | Prob. (%) | $Z(\sigma)$ | Prob. (%) | $Z(\sigma)$ |
| AAA | 0.09 | 3.12 | 0.03 | 3.43 |
| AA | 2.27 | 1.98 | 0.14 | 2.93 |
| A | 91.05 | -1.51 | 0.67 | 2.39 |
| BBB | 5.52 | -2.30 | 7.73 | 1.37 |
| BB | 0.74 | -2.72 | 80.53 | -1.23 |
| B | 0.26 | -3.19 | 8.84 | -2.04 |
| CCC | 0.01 | -3.24 | 1.00 | 2.30 |
| Default | 0.06 | | 1.06 | |

**Table 13.8.** Transition probabilities and credit quality thresholds for BB and A rated obligors

If $p_{Def}$ denotes the probability for the $BB$-rated obligor defaulting, then the critical asset value $V_{Def}$ is such that

$$p_{Def} = P[V_t \leq V_{Def}],$$

which can be translated into a normalized threshold $Z_{CCC}$, such that the area in the left tail below $Z_{CCC}$ is $p_{Def}$. Indeed, according to (13.1), default occurs when $Z_t$ satisfies

$$p_{Def} = P\left[\frac{ln(V_{Def}/V_0) - (\mu - \sigma^2/2)\,t}{\sigma\sqrt{t}} \geq Z_t\right] \tag{13.2}$$

$$= P\left[Z_t \leq -\frac{ln(V_0/V_{Def}) + (\mu - \sigma^2/2)\,t}{\sigma\sqrt{t}}\right] \tag{13.3}$$

$$= \Phi(-d_2), \tag{13.4}$$

where the normalized return $r = \dfrac{ln(V_t/V_0) - (\mu - \sigma^2/2)\,t}{\sigma\sqrt{t}}$ is $\mathcal{N}(0,1)$. The $Z_{CCC}$ is simply the threshold point in the standard normal distribution corresponding to a cumulative probability of $p_{Def}$. Then the critical asset value $V_{Def}$, which triggers default, is such that $Z_{CCC} = -d_2$, where

$$d_2 = \frac{ln(V_0/V_{Def}) + (\mu - \sigma^2/2)\,t}{\sigma\sqrt{t}},$$

and is also called *distance-to-default* .

Accordingly $Z_B$ is the threshold point corresponding to a cumulative probability of being either in default or in rating CCC, i.e., $p_{Def} + p_{CCC}$, etc. Further, since asset returns are not directly observable, CreditMetrics/ CreditVaR chose equity returns as a proxy equivalent to assuming that the firm's activities are all equity financed. Now, for the time being, assume that the correlation between asset rates of return is known and is denoted by $\rho$, which is assumed to be equal to 0.20 in our example. The normalized log-returns on both assets follow a joint normal distribution:

$$f(r_{BB}, r_A; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{\frac{-1}{2(1-\rho^2)}[r_{BB}^2 - 2\rho r_{BB} r_A + r_A^2]\right\}.$$

We can then easily compute the probability for both obligers of being in any combination of ratings, like that they remain in the same rating classes $BB$ and $A$, respectively:

$$P(-1.23 < r_{BB} < 1.37, -1.51 < r_A < 1.98) = \int_{-1.23}^{1.37}\int_{-1.51}^{1.98} f(r_{BB}, r_A; \rho)dr_{BB}dr_A = 0.7365.$$

If we implement the same procedure for the other 63 combinations we obtain Table 13.9. We can compare Table 13.9 with Table 13.7, the latter being derived assuming zero correlation, to notice that the joint probabilities are different.

To be more specific, consider two obligors whose probabilities of default are $P1(P_{Def1})$ and $P2(P_{Def2})$, respectively and their asset return correlation is $\rho$. The

| | Second company (A) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| First Company BB | AAA | AA | A | BBB | BB | B | CCC | Def | Total |
| AAA | 0 | 0 | 0.03 | 0 | 0 | 0 | 0 | 0 | 0.03 |
| AA | 0 | 0.01 | 0.13 | 0 | 0 | 0 | 0 | 0 | 0.14 |
| A | 0 | 0.04 | 0.61 | 0.01 | 0 | 0 | 0 | 0 | 0.67 |
| BBB | 0.02 | 0.35 | 7.1 | 0.2 | 0.02 | 0.01 | 0 | 0 | 7.69 |
| BB | 0.07 | 1.79 | 73.65 | 4.24 | 0.56 | 0.18 | 0.01 | 0.04 | 80.53 |
| B | 0 | 0.08 | 7.8 | 0.79 | 0.13 | 0.05 | 0 | 0.01 | 8.87 |
| CCC | 0 | 0.01 | 0.85 | 0.11 | 0.02 | 0.01 | 0 | 0 | 1 |
| Def | 0 | 0.01 | 0.9 | 0.13 | 0.02 | 0.01 | 0 | 0 | 1.07 |
| Total | 0.09 | 2.29 | 91.06 | 5.48 | 0.75 | 0.26 | 0.01 | 0.06 | 100 |

**Table 13.9.** Joint rating probabilities (%) for BB and A rated obligors when $\rho = 0.2$

events of default for obligors 1 and 2 are denoted as DEF1 and DEF2 , respectively, and P(DEF1,DEF2) is the joint probability of default. Then, it can be shown that the default correlation is

$$corr(DEF1, DEF2) = \frac{P(DEF1, DEF2) - P1.P2}{\sqrt{P1(1-P1)P2(1-P2)}} \ .$$

The joint probability of both obligors defaulting is, according to Merton's model,

$$P(DEF1, DEF2) = Pr[V_1 \leq V_{Def1}, V_2 \leq V_{Def2}] \ ,$$

where $V_1$ and $V_2$ denote the asset values for both obligors at time $t$, and $V_{Def1}$ and $V_{Def2}$ are the corresponding critical values which trigger default. It is further equivalent to

$$P(DEF1, DEF2) = Pr[r_1 \leq -d_2^1, r_2 \leq d_2^2] = N_2(-d_2^1, -d_2^2, \rho),$$

where $r_1$ and $r_2$ denote the normalized asset returns for obligors 1 and 2, respectively, and $d_2^1$ and $d_2^2$ are the corresponding distance to default. $N_2(x, y, \rho)$ denotes the cumulative standard bivariate normal distribution and $\rho$ is the correlation coefficient between $x$ and $y$.

CreditMetrics does not itself answer the question regarding the pricing of credit risk and its underlying modeling. Forward zero curves are supplied as an input to the framework that could account for the drawback of CreditMetrics approach. However, forward curves are the result of an estimation process of the term structure of the credit spreads. This leads to an integrating framework which provides an idea about credit portfolio diversification.

The second drawback of CreditMetrics is the assumption of credit homogeneity within a credit rating class. The entire framework relies on the choice of a rating scheme. Each class defines a class for systematic credit risk and two obligors are supposed to have perfectly the same credit behavior if in the same credit rating class. The methodology is good but it does not allow for good evaluation of nonlinear instruments such as credit derivatives, swaps etc.

In this league of work, KMV is quite attractive mainly for the methodology required to estimate consistent correlations between obligors of different categories. KMV's model does not start from any transition matrix between rating categories, therefore making no initial assumption on the common behavior of two counterparties belonging to the same rating class. Using a variant of Merton's model, KMV uses a structural approach to define the "distance-to-default category" (DD) to which the counterparty belongs. Then, relying on a huge default database, KMV is able to relate this DD category to an "expected default frequency" (EDF) for each issuer. Details of KMV method could be read from [96, 147].

CreditPortfolioView is a multi-factor model which is used to simulate the joint conditional distribution of default and migration probabilities for various rating groups in different industries, for each country, conditional on the value of macroeconomic factors like the unemployment rate, the rate of growth in GDP, the level of long-term interest rates, foreign exchange rates, government expenditures and the aggregate savings rate. CreditPortfolioView is based on the casual observation that default probabilities, as well as migration probabilities, are linked to the economy. When the economy worsens both downgrades as well as defaults increase. It is contrary when the economy becomes stronger. In other words, credit cycles follow business cycles closely. Since the state of the economy is, to a large extent, driven by macroeconomic factors, CreditPortfolioView proposes a methodology to link those macroeconomic factors to the default and migration probabilities [96, 147].

Although these models have different calculation techniques and parameters, studies have shown that they in fact represent a remarkable consensus in their underlying frameworks and financial intuition. However, before these models can deliver on their promise, they must not only prove to be conceptually sound and empirically valid; they must also be well integrated with a bank day-to-day credit risk management. Merton-based models, using asset correlation derived from equity data, might be more accurate for publicly traded companies, while actuarial models, using default-rate volatility based on historical experience, might be more accurate for illiquid asset classes or small business portfolios.

In the Table 13.10 we summarize the analytic model data requirements for the aforementioned models.

| Model | Input | Output |
|---|---|---|
| JP Morgans CreditMetrics | Default and migration probabilities<br>(transition matrices)<br>Credit spreads and yield curves<br>Pair-wise correlation<br>Recovery rates<br>Credit exposures | Econo. capital<br>(both expected loss and unexpected loss)<br>Return distribution<br>Loss percentiles |
| CSFPs CreditRisk+ | Default rates<br>Default rate volatility<br>Recovery rates<br>Credit exposures | Econo. capital<br>Loss distribution<br>Loss percentiles |
| KMVs Portfolio Manager | Expected default frequencies<br>Credit spreads<br>Pair-wise correlation<br>Credit exposures | Econo. capital<br>Sharpe ratio<br>Mis-pricing<br>Optim. benefits<br>Return distribution |
| McKinsey & Co.s Credit | Macroeco. variables<br>Default and migration history<br>Credit spreads<br>Recovery rates<br>Credit exposures | Econo. capital<br>return distribution<br>Loss percent |



**Table 13.10.** Input-Output of various credit risk models

## 13.6 Credit Risk Optimization with CVaR Criteria

In earlier section we studied VaR that seems to provide an efficient solution and provide answer to the question: what is the maximum loss with the confidence level $(1-\beta)\times 100\%$ over a given time horizon? Thus, its calculation reveals that the loss will exceed VaR with likelihood $\beta \times 100\%$, but no information is provided on the amount of the excess loss, which may be significantly large. Mathematically, VaR has serious limitations. In case of a finite number of scenarios, it is a nonsmooth, nonconvex, and multi extremum function [91] (with respect to positions), making it difficult to control and optimize. Also, VaR has some other undesirable properties, such as the lack of sub-additivity.

By contrast, CVaR is considered a more consistent measure of risk than VaR. CVaR supplements the information provided by VaR and calculates the quantity of the excess loss. Since CVaR is greater than or equal to VaR, portfolios with a low CVaR also have a low VaR. Under quite general conditions, CVaR is a convex function with respect to positions [140], allowing the construction of efficient optimization algorithms. CVaR has been compared with the widely accepted VaR risk performance measure for which various estimation techniques have been proposed, see for example [39, 109].

Bucay and Rosen [22] applied CreditMetrics methodology to a portfolio of corporate and sovereign bonds issued in emerging markets. They estimated the credit risk of the undertaking portfolio by taking into account both defaults and credit migrations. On the similar lines, we utilize CVaR optimization routine developed in Chapter 6 for the bond portfolio discussed in [3].

The test portfolio has been compiled by a group of financial institutions to access the state-of-the-art of portfolio credit risk models. The portfolio consists of 197 emerging markets bonds, issued by 86 obligors in 29 countries. The date of the analysis was October 13, 1998 and the mark-to-market value of the portfolio was 8.8 billion USD. Most instruments are denominated in US dollars but 11 fixed rate bonds are denominated in seven other currencies; DEM, GBP, ITL, JPY, TRL, XEU and ZAR. Bond maturities ranged from a few months to 98 years and the portfolio duration is approximately five years.

Let $x = (x_1, x_2, \dots, x_n)$ be obligor weights (positions) expressed as multiples of current holdings, $b = (b_1, b_2, \dots, b_n)$ be future values of each instrument with no credit migration (benchmark scenario), and $y = (y_1, y_2, \dots, y_n)$ be the future (scenario-dependent) values with credit migration. The loss due to credit migration for the portfolio is defined as $f(x, y) = (b - y)^T x$. The CVaR optimization problem is formulated as

$$\underset{x \in X \subset \mathbf{R}^n}{\text{Min}} \, \Phi(x) = \Phi(z, \alpha)$$

where $X$ is the feasible set in $\mathbf{R}^n$. For the definition of $\Phi$ please refer to Chapter 6. This set could also contain constraints pertaining mean return constraint, box constraints on the positions of instruments, etc. However to make content simple and easy to read we have not discussed them in optimization routine. Interested reader should read [3, 91]. They have approximated the performance function using scenarios $y_j, j = 1, \dots, J$, which are sampled with the density function $p(y)$. As discussed in Chapter 6, minimization of the CVaR function $\Phi(x)$ could be reduced (approximately) to the following linear programming problem

$$\min_{x,\alpha,z} \quad \alpha + \nu \sum_{j=1}^{J} z_j$$

subject to

$$\begin{aligned} f(x, y_j) - \alpha \le z_j \quad & (j = 1, 2, \ldots, J) \\ e^T x = 1 \quad & \\ z_j \ge 0 \quad & (j = 1, 2, \ldots, J) \\ l_i \le x_i \le u_i \quad & (i = 1, 2, \ldots, n), \end{aligned} \tag{13.5}$$

where $\nu = \dfrac{1}{(1-\beta)J}$, where $\beta$ is the confidence level.

Also, if $(x^*; \alpha^*; z^*)$ is an optimal solution of the optimization problem (13.5), then $x^*$ is an approximation of the optimal solution of the CVaR optimization problem, the function $\Phi(z^*; \alpha^*)$ equals approximately the optimal CVaR, and $\alpha^*$ is an approximation of VaR at the optimal point. Thus, by solving problem (13.5) we can simultaneously find approximations of the optimal CVaR and the corresponding VaR.

Further, in [3, 91], two sets of additional constraints, mentioned below, in the optimization of CVaR are considered to evaluate the performance against benchmark case. These set of constraints are

(i) no short positions allowed, and the positions can be at most doubled in size.
(ii) positions, both long and short, can be at most doubled in size.

The first constraint simply means that $l_i = 0$ and $u_i = 2$, i.e. $0 \le x_i \le 2$ and the second constraint implies that $-2 \le x_i \le 2$. Further, authors [3] have considered that the re-balanced portfolio should maintain the future expected value, in absence of any credit migration, i.e. $\sum_{i=1}^{n} b_i x_i = \sum_{i=1}^{n} b_i$. After adding these set of constraints the result of the optimization, in the case of no short positions (no short), and in the case of both long and short positions (long and short), are presented in Table 13.10.

Table 13.10 shows that the two risk measures, VaR and CVaR, are significantly improved after the optimization. When no short positions are allowed, VaR and CVaR reduce by about 60%. For example, at $(1-\beta) = 0.99$, we lowered CVaR to 559 million from the original 1320 million USD. By allowing both short and long positions slightly improves reductions, but not significantly. Thus, it has been observed that risk measures could be reduced by about 60% with the multiple obligor optimization.

| Case | $\beta$ | VaR | VaR (%) | CVaR | CVaR (%) |
|---|---|---|---|---|---|
| Original | 0.900 | 340 | - | 621 | - |
| | 0.950 | 518 | - | 824 | - |
| | 0.990 | 1,026 | - | 1,320 | - |
| | 0.999 | 1,782 | - | 1,998 | - |
| No Short | 0.900 | 163 | 52 | 279 | 55 |
| | 0.950 | 239 | 54 | 359 | 56 |
| | 0.990 | 451 | 56 | 559 | 58 |
| | 0.999 | 699 | 61 | 761 | 62 |
| Long | 0.900 | 149 | 56 | 264 | 58 |
| and | 0.950 | 226 | 56 | 344 | 58 |
| Short | 0.990 | 433 | 58 | 542 | 59 |
| | 0.999 | 680 | 62 | 744 | 63 |

**Table 13.11.** VaR, CVaR (in Millions of USD) and corresponding VaR and CVaR reductions (in %) for the Multiple Obligor Optimization.

Table 13.11 shows that the two risk measures, *expected loss and standard deviation*, also are dramatically improved when we minimized CVaR. For example, in case of both long and short positions, the expected loss and standard deviation are reduced about 50%. The corresponding position weights for the original twelve largest risk contributors are presented in Table 13.12.



| Case | $\beta$ | Expected Loss | Expected Loss (%) | Standard Deviation | Standard Deviation (%) |
|---|---|---|---|---|---|
| Original | - | 95 | - | 232 | - |
| No Short | 0.90 | 50 | 47 | 107 | 54 |
| | 0.95 | 51 | 46 | 109 | 53 |
| | 0.99 | 60 | 37 | 120 | 48 |
| | 0.999 | 63 | 34 | 126 | 46 |
| Long | 0.90 | 42 | 56 | 105 | 55 |
| and | 0.95 | 44 | 54 | 107 | 54 |
| Short | 0.99 | 53 | 44 | 118 | 49 |
| | 0.999 | 58 | 39 | 124 | 47 |

**Table 13.12.** Expected Loss (in Millions of US-D), Standard Deviation (in Millions of US-D) and Corresponding Reductions (in %) for the Multiple Obligor Optimization

| Obligor | Original | No Short | Short and Long |
|---|---|---|---|
| Brazil | 1 | 0.08 | 0.18 |
| Russia | 1 | 0 | 0.09 |
| Venezuela | 1 | 0 | -0.41 |
| Argentina | 1 | 0.35 | 0.47 |
| Peru | 1 | 0 | -0.38 |
| Colombia | 1 | 0.89 | 1.00 |
| Morocco | 1 | 0.02 | 0.12 |
| RussiaIan | 1 | 0 | -2.00 |
| MoscowTel | 1 | 1.52 | 1.99 |
| Romania | 1 | 0.45 | 1.33 |
| Mexico | 1 | 0.94 | 0.90 |
| Philippines | 1 | 1.04 | 0.94 |

**Table 13.13.** Positions for the Multiple Obligor Optimization (Minimization of CVaR with $(1-\beta) =$ 0:99).

This idea behind using the CVaR optimization framework is to simultaneously optimize two closely related risk measures: CVaR and VaR. Although with the CVaR as a performance function, the optimization leads to reduction of all other aforementioned risk measures e.g. CVaR, VaR, the expected loss, and the standard deviation. From a bank perspective, this approach looks quite attractive. The bank should have reserves to cover expected loss and capital to cover unexpected loss.

## 13.7 Credit Scoring and Internal Rating

So far we have considered the CreditMetrics, credit VaR and CVaR to analyze and estimate risks associated with credit default by risk departments of the banks and financial firms. However, some companies run a credit risk department where the job is to assess the financial health of their customers, and extend credit accordingly. The credit scoring model forms an essential part of the aforementioned framework in order to grant credit to clients. Thus, credit scoring can be defined as a technique that helps credit providers decide whether to grant credit to customers by segregating 'good' and 'bad' customers in terms of their creditworthiness.

The aim of a credit scoring model is to build a single aggregated risk indicator for a set of risk factors. Hence it is a tool that aid decision whether to grant credit to a new applicant. The other type of decision is how to deal with existing customers i.e if an existing customer wants to increase his credit limit should the firm agree to that? If the customer starts to fall behind in repayments what actions should the firm take? Techniques that help with these decisions are called *behavioural scoring*.

In order to grant loan to customers following are the steps considered in the lending process:

(i) Solicitation: Either the firm solicits applications for loans, for instance via advertising, or a client approach to firm with a new request.

(ii) Information gathering: Information about the applicant via various means including interviews, visits, review of financial data or accounts, possible use of credit reference agencies, ratings agencies or other available data has been gathered to rate the client.

(iii) Recommendation: On the basis of information gathered and consideration of why the client wants the money an internal rating is assigned and a lending decision is recommended. This is then reviewed, perhaps by a loan committee in the case of a corporate loan or branch staff in the case of a retail exposure.

(iv) Closing administration: Any collateral is perfected, documentation is finalized and signed and funds are made available.

(v) Monitoring: The performance of the obligator and their condition are re-assessed periodically. One aim of internal ratings is to allow many applications from different kinds of corporates in different countries to be assessed on an equitable basis. It also allows the bank to set break-even spreads for internal ratings classes which encourage lending to good counterparties.

The analogue of internal ratings for retail exposures is *credit scoring*. Credit scoring can be formally defined as a statistical (or quantitative) method that is used to predict the probability that a loan applicant or existing borrower will default or become delinquent [95]. This helps to determine whether credit should be granted to a borrower [100]. Credit scoring can also be defined as a systematic method for evaluating credit risk that provides a consistent analysis of the factors that have been determined to cause or affect the level of risk [135]. Data for a system should have discriminating power that could be based on certain parameters like gross income, age, number of years spent in the current position, gender etc. This can then be used both reactively to decide on whether to accede to a mortgage application, for instance and proactively to solicit application for a particular type of credit card.

Credit scoring has many benefits that accrue not only to the lenders but also to the borrowers. For example, credit scores help to reduce discrimination as scoring models provide an objective analysis of a customers credit worthiness.

With the help of the credit scores, financial institutions are able to quantify the risks associated with granting credit to a particular applicant in a shorter time. Further, credit scores can help financial institutions determine the interest rate that they should charge their customers and to price portfolios [6]. Higher-risk customers are charged a higher interest rate and vice versa. Based on the customers credit scores, the financial institutions are also able to determine the credit limits to be set for the customers [106, 118]. This further help financial institutions to manage their accounts more effectively and profitably.

Credit scores are also used as a basis to adjust premiums. Generally, customers with bad credit scores have a higher chance of failing insurance claims compared with customers with good credit scores. Therefore, the former are charged a higher premium. Further, the credit information is used to assess a customers accountability and performance under the conditions of an insurance policy.

The methods generally used for credit scoring are based on statistical pattern recognition techniques. Historically, discriminant analysis and linear regression were the most widely used techniques for building scorecards. Later the *logit* and *probit* models were suggested in Martin [88] and Ohlson [105]. All these models belong to the class of generalized linear models (GLM) and could also be interpreted using a latent (score) variable. Their core decision element is a linear score function (graphically represented as a hyperplane in a multidimensional space) separating successful and failing companies. The company score is computed as a value of that function. In the case of the probit and logit models the score is via a link function which is directly transformed into a probability of default (PD). The major disadvantage of these popular approaches is the enforced linearity of the score and, in the case of logit and probit models, the prespecified form of the link function (logit and Gaussian) between PDs and the linear combination of predictors.

Some of the techniques that have been previously used, but rather infrequently, to construct credit scoring models include genetic algorithm, $k-$nearest neighbor, linear programming, and expert systems Thomas et al. [135]. In recent years, new techniques have been increasingly used to construct credit scoring models. In particular, the decision tree approach has become a popular technique for developing credit scoring models because the resulting decision trees are easily interpretable and visualized. Further, neural networks are also commonly used. All the methods and techniques mentioned above can be considered as an important data mining techniques for predictive modeling.

## 13.8 Construction of Credit Scoring Models

The methodology for constructing credit scoring models generally involves the following process. First, a sample of previous customers is selected and classified as good or bad depending on their repayment performance over a given period (for simplicity, only a dichotomy is used here). Next, data are compiled from loan applications, personal and/or business credit records, and various sources if available (e.g. credit bureau reports). Finally, statistical (or other quantitative) analysis is performed on the data to derive a credit scoring model. The model comprises weights to apply to the different variables (or attributes) in the data and a cut-off point. The sum of the weights applied to the variables for an individual customer constitutes the credit score. The cut-off point determines whether this customer should be classified as "good" or "bad". The probability associated with this classification can also be generated. Different models can be constructed for different segments of the data (e.g. for different products).

The concept of incurred cost associated with the probabilities of repayment of loans will be used to illustrate some of the methods. For simplicity, assume that the population of loans consist of two groups or classes G and B that denote loans that (after being granted) will turn out to be good or bad in the future, respectively. Good loans are repaid in full and on time. Bad loans are subject to the default.

Usually the class/group sizes are very different, so that for the probability that a randomly chosen customer belongs to group $G$, denoted as $p_G$, one has $p_G > p_B$. Let $x$ be a vector of independent variables (also called the measurement vector) used in the process of deciding whether an applicant belongs to group $G$ or $B$. Let the probability that an applicant with measurement vector $x$ belongs to group $G$ be $p(G/x)$, and that of $B$ be $p(B/x)$. Let the probability $p(x/G)$ indicate that a good applicant has measurement vector $x$. Similarly, for bad applicants the probability is $p(x/B)$. The task is to estimate probabilities $p(\cdot/x)$ from the set of given data about applicants which turn out to be good or bad and to find a rule for how to partition the space $X$ of all measurement vectors into the two groups $A_G$ and $A_B$ based on these probabilities, so that in $A_G$ would be the measurement vectors of applicants who turned out to be good and vice versa.

It is usually not possible to find perfect classification as it may happen that the same vector is given by two applicants where one is good and the other is bad. Therefore it is necessary to find a rule that minimizes the cost of a bank providing credit connected with the misclassification of applicants. Let $c_G$ denote the costs connected with misclassifying a good applicant as bad and $c_B$ the costs connected with classifying a bad applicant as good. Usually $c_B > c_G$, because costs incurred due to misclassifying a bad customer are financially more damaging than cost associated with the former kind of error. If applicants with $x$ are assigned to class $B$, the expected cost is $c_B p(B/x)$ and the expected loss for the whole sample is $c_B \sum_{x \in A_B} p(B/x)p(x) + c_G \sum_{x \in A_G} p(G/x)p(x)$, where $p(x)$ is a probability that the measurement vector is equal to $x$. This is minimized when, into group $G$, such applicants are assigned who have their group of measurement vectors $A_G = \{x : c_B p(B|x) \leq c_G p(G/x)\}$ which is equivalent to $A_G = \{x : p(G/x) \geq \dfrac{c_B}{c_G + c_B}\}$. Without loss of generality, the misclassification costs can be normalized to $c_G + c_B = 1$. In this case, the rule for classification is to assign an applicant with $x$ to class $G$ if $p(G/x) > c_B$ and otherwise to class $B$. An important task is to specify the cost of lending errors and to accurately specify the optimal cutoff-score for credit scoring, as banks have to choose the optimal trade-off between profitability and risk. Credit policies that are too restrictive may ensure minimal costs in terms of defaulted loans, but the opportunity costs of rejected loans may exceed potential bad debt costs and thus profit is not maxmized. On the other hand, policies that are too liberal may result in high losses from bad debt.

In the sequel, we would discuss three algorithms which are very popular in the domain of credit scoring. These are logistic regression, Fisher's linear discriminant analysis and support vector machines. Our presentation here is motivated by Thomas et al. [135].



## 13.9 Logistic Regression

The traditional *linear regression* approach tries to find the best linear combination $w_0 + w_1 X_1 + w_2 X_2 + \ldots + w_n X_n$ of characteristics which explains the probability of default. Here $(X_1, X_2, \ldots, X_n)$ is the set of $n$ random variables that describe the information available on an applicant for credit from the application form. Let $p_i$ $(i = 1, 2, \ldots, n)$ be the probability that the $t^{th}$ applicant in the sample has defaulted. Then our aim in linear regression is to find $\overline{w}_0, \overline{w}_1, \ldots, \overline{w}_n$ such that the linear combination $\overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n}$ best approximates the probability of default $p_i$, i.e., $p_i = \overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n}$ for all $i$. For the $i^{th}$ applicant, the observed data is $x_{i,1}, x_{i,2}, \ldots, x_{i,n}$ for the random variables $X_1, X_2, \ldots, X_n$ respectively.

The regression approach to linear discrimination is not very satisfactory. This is because the linear combination $\overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n}$ could take any value in $(-\infty, \infty$, but the probability of default $p_i$ takes values between 0 and 1 only. Therefore it would be better if we could modify the usual linear regression approach so that the linear combination $\overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n}$ best approximates some function of $p_i$ which could take wider range of values. One such function is $ln\left(\frac{p_i}{1-p_i}\right)$ which results in logistic regression. Thus in logistic regression we aim to find $\overline{w}_0, \overline{w}_1, \ldots, \overline{w}_n$ so that the linear combination $\overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n}$ best approximates the function $ln\left(\frac{p_i}{1-p_i}\right)$ for all $i$. Since $\left(\frac{p_i}{1-p_i}\right)$ takes values between 0 and $\infty$, $ln\left(\frac{p_i}{1-p_i}\right)$ takes values between $-\infty$ and $\infty$. Thus for all $i$, we take

$$ln\left(\frac{p_i}{1-p_i}\right) = \overline{w}_0 + \overline{w}_1 x_{i,1} + \ldots + \overline{w}_n x_{i,n} \ . \tag{13.6}$$

Taking exponentials on both sides of (13.6) we have $p_i = \frac{e^{Z^{(i)}}}{1+e^{Z^{(i)}}} = \frac{1}{(1+e^{-Z^{(i)}})}$, where $Z^{(i)} = \frac{e^{w^T x^{(i)}}}{(1+e^{w^T x^{(i)}})}$, $w = (w_0, w_1, w_2, \ldots, w_n)^T$. Therefore the logistic regression approach has the assumption that $p_i$ is of this specific form. The function $f(u) = \frac{1}{(1+e^{-u})}$ is called the *logistic function.*

If we assume that the distribution of the characteristics values of the 'good' and of the 'bad' is multivariate normal then we can show that the assumption of logistic regression is satisfied. In this case we can obtain a close form expression for $ln\left(\frac{p_i}{1-p_i}\right)$ in terms of the vectors $\mu_G$, $\mu_B$ and the matrix $\sum$. Here $\mu_G$ is the vector of means among the 'goods', $\mu_B$ is the vector of means among the 'bads' and $\sum$ is the variance-covariance matrix for the random vector $(X_1, X_2, \ldots, X_n)$.

In classical regression, it is simple to calculate the vector of coefficients $\overline{w}$ by employing the method of least squares. But the determination of the vector $\overline{w}$ in logistic regression is not that simple. We need to use the maximum likelihood approach to get estimates for these coefficients. This involves an application of the Newton-Raphson method to solve the equation that arise.

In conclusion the usual linear regression and logistic regression are very similar in approach but very different conceptually. The linear regression tries to fit the probability $p$ of defaulting by a linear combination of the attributes while logistic regression tries to fit $ln\left(\frac{p}{1-p}\right)$ by a linear combination of attributes. The advantage of logistic regression is that it does not require linearity of relationship between the independent variables and the dependent variable, and it does not require the assumption of normality. But it assumes that the independent variables are linearly related to *logit* (*i.e.* $ln\left(\frac{p}{1-p}\right)$) of the dependent variable.

## 13.10 Fisher's Linear Discriminant Analysis

Discriminant analysis which was studied by Fisher as early as 1936 is an alternative to logistic regression. It assumes that the explanatory variables follow a multivariate normal distribution and have a common variance-covariance matrix. This method is used to classifying observations in two classes. In the context of credit scoring, the two groups are those classified by the lender as 'goods' and 'bads' and the characteristics are the applicant's form details and the credit bureau information.

Let $Y = w_1X_1 + w_2X_2 + \ldots + w_nX_n$ be any linear combination of characteristics $X = (X_1, X_2, \ldots, X_n)$. The most natural measure of separation seems to be the difference between the mean values of two different groups of 'good' and 'bads' in the sample. Thus it makes sense to find $\overline{w}_i$ with $\sum_i^n \overline{w}_i = 1$ such that the difference between $E(Y|G)$ and $E(Y|B)$ is maximum. But it is also important to see how closely each of the two groups cluster together when we discuss their separation. Therefore the aim should be to minimize the distance within each group and maximize the distance between different groups using some suitable discriminant function.



Fisher suggested that, under the assumptions as stated above, an appropriate measure of separation is

$$M = \frac{\text{square of distance between sample means of two groups}}{(\text{sample variance of each group})^{1/2}}.$$

Here we have division by the square root of the variance to make the measure scale independent. Thus if $Y$ is changed to $CY$, the measure $M$ will not change.

Let $m_G$ and $m_B$ be the vector of sample means of 'goods' and 'bads' and $S$ be the common sample variance. For $Y = w_1X_1 + w_2X_2 + \ldots + w_nX_n$, the separating measure $M$ is given by

$$M = \frac{w^T(m_G - m_B)w}{(w^TSw)^{1/2}} \; . \tag{13.7}$$

The above equation follows because $E(Y|G) = w^Tm_G$, $E(Y|B) = w^Tm_B$, and $Var(Y) = (w^TSw)$. To maximize $M$ we compute $\nabla_w M$ and equate it to zero. This gives

$$(m_G - m_B)w.(w^T Sw) = w^T(m_G - m_B)w.Sw,$$

i.e. $(m_G - m_B)w = M.Sw$, giving

$$Mw = S^{-1}(m_G - m_B)w$$

Therefore solving generalize eigenvalue problem leads to eigen vector $w$ directly proportional to $S^{-1}(m_G - m_B)$. This relation can be used to find weights $\overline{w}_i$ which maximize $M$.

We then compute the score $s(x) = \sum_i \overline{w}_i x_i$, where $x_i$ are explanatory variables. We use the score $s(x)$ to discriminate between the two groups. The normality assumption only becomes important if significance tests are to be undertaken. The advantages of the linear discriminant analysis (LDA) method are that it is simple, it can be very easily estimated and it actually works very well. The theoretical disadvantage of linear discriminant analysis is that it requires normally distributed data but the credit data are often non-normal (and categorized). But if one is not too much concerned about significance tests then it can be used for most of data.

## 13.11 Support Vector Machines

Support vector machine is a classifier technique, first proposed by Vapnik [141]. This method involves three elements. A score formula which is a linear combination of features selected for the classification problem, an objective function which considers both training and test samples to optimize the classification of new data, an optimizing algorithm for determining the optimal parameters of training sample objective function.

Credit applicants are assigned to good or bad risk classes according to their records of defaulting. Each applicant is described by a high dimensional input vector of situational charateristics and with an associated class label. A statistical model which maps the inputs to the labels can decide whether a new credit applicant should be accepted or rejected by predicting the class label given the new inputs. Support vector machines (SVM), from the statistical learning theory, can build such models from the data, requiring but extremely weak prior assumptions about the model structure. Furthermore, SVM divide a set of labeled credit applications into subsets of “typical” and “critical” patterns. The correct class label of a typical pattern is usually very easy to predict, even with the linear classification methods. Such patterns do not contain much information about the classification boundary. The critical patterns (the support vectors) contain the less trivial training samples.

The advantages of the method are that, in the nonparametric case, SVM requires no data structure assumptions such as normal distribution and continuity. SVM can perform a nonlinear mapping from an original input space into a high dimensional feature space and this method is capable of handling both continuous and categorical predictions. The weaknesses of this method are that, it is difficult to interpret unless the features interpretable and the standard formulations do not contain any specification of business constraints.

Given a training set of instance-label pairs $(x_i, y_i), i = 1, 2, \ldots, m$ where $x_i \in R^n$ and $y_i \in \{+1, -1\}$, SVM finds an optimal separating hyperplane with the maximum margin by solving the following optimization problem

$$(LP) \quad \underset{w,b}{\text{Min}} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{m} \xi_i$$

subject to

$$y_i(w^T x_i + b) + \xi_i - 1 \geq 0$$
$$\xi_i \geq 0$$

where $C$ is a penalty parameter on the training error and $\xi_i$ is the non-negative slack variable. Introducing the Karush KuhnTucker (KKT) [98] condition for the optimum constrained function, then $LP$ is transformed to the dual Lagrangian $LD(\alpha)$.

$$LD(\alpha) \quad \underset{\alpha}{\text{Max}} \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j < x_i . x_j >$$

subject to

$$\textstyle\sum_{i=1}^{m} \alpha_i y_i = 0$$
$$0 \leq \alpha_i \leq C \quad (i = 1, 2, \ldots, m)\ .$$

To find the optimal hyperplane, a dual Lagrangian $LD(\alpha)$ must be maximized with respect to non-negative $\alpha_i$. The solution $\alpha_i$ for the dual optimization problem determines the parameters $w^*$ and $b^*$ of the optimal hyperplane. Thus, the optimal hyperplane decision function $f(x) = sign(w^{*T} x + b^*)$ can be written as $f(x) = sign\left(\sum_{i=1}^{m} y_i \alpha_i^* < x_i, x > + b^*\right)$. The nonlinear SVM maps the training samples from the input space into a higher-dimensional feature space via a mapping function $\phi$. In the dual Lagrange, the inner products are replaced by the kernel function $< \phi(x_i), \phi(x_j) >= K(x_i, x_j)$, and the nonlinear SVM dual Lagrangian $LD(\alpha)$ is similar with that in the linear generalized case i.e.

$$\underset{\alpha}{\text{Max}} \qquad \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$
$$0 \le \alpha_i \le C \quad (i = 1, 2, \ldots, m)$$

Followed by the steps described in the linear generalized case, we obtain decision function of the following form

$$\begin{aligned} f(x) &= sgn\left(\sum_{i=1}^{m} y_i \alpha_i^* < \phi(x), \phi(x_i) > +b^*\right) \\ &= sgn\left(\sum_{i=1}^{m} y_i \alpha_i^* K(x, x_i) + b^*\right). \end{aligned} \tag{13.8}$$

Proper kernel parameters setting can improve the SVM classification accuracy. With the RBF kernel, there are two parameters $C$ and $\gamma$ to be determined in the SVM model. The grid search approach [63] is an alternative to finding the best $C$ and $\gamma$ when using the radial basis function (RBF) kernel function. In addition to the proper parameters setting, feature subset selection can improve the SVM classification accuracy. To pursue even small improvement in credit scoring accuracy, many methods have been investigated in the last decade. Artificial Neural Networks (ANNs), genetic algorithms (GA) are the most commonly soft computing method used in credit scoring modelling [134, 135].

## 13.12 Drawbacks

Although credit scoring has significant benefits, its limitations should also be noted. One major problem that can arises when constructing a credit scoring model is that the model may be built using a biased sample of customers who have been granted credit [58]. This may occur because applicants (i.e. potential customers) who are rejected are not included in the data for constructing the model. Hence, the sample is biased (i.e. different from the general population) as good customers are too heavily represented. The credit scoring model built using this sample may not perform well on the entire population since the data used to build the model is different from the data that the model will be applied to. The second problem that can arise when constructing credit scoring models is the change of patterns over time. The key assumption for any predictive modeling is that the past can predict the future [10]. In credit scoring, this means that the characteristics of past applicants who are subsequently classified as good or bad creditors can be used to predict the credit status of new applicants. Sometimes, the tendency for the distribution of the characteristics to change over time is so fast that it requires constant refreshing of the credit scoring model to stay relevant. One of the consequences of credit scoring is the possibility that end-users become so reliant on the technology that they reduce the need for prudent judgment and the need to exercise their knowledge on special cases. In other instances, end-users may unintentionally apply more resources than necessary to work the entire portfolio. This could run into the risk of a self-fulfilling prophecy [84]. In the U.S., a new industry has emerged that is dedicated to help borrowers improve their credit scores by rearranging finances [136], rather than obeying the simple rule: pay your bills on time and keep your debt low. Such score-polishing actions could potentially distort the patterns of credit default. Despite the limitations highlighted above, there is no doubt that credit scoring continues to be a major tool in predicting credit risk in customer lending. It is envisaged that organizations using credit scoring appropriately will gain important strategic advantage and competitive edge over their rivals.



## 13.13 Summary and Additional Notes

- This chapter seeks to give an overview of the techniques that are used and being developed to forecast the financial risk involved in lending to customers. We have sought to give a fairly comprehensive biography of the literature of the credit scoring models. Credit scoring are some of the most important forecasting techniques used in the retail and customer finance areas. As a pure forecasting tool as opposed to a decision making one, credit scoring has mainly been used as a way of forecasting future bad debt in order to set aside appropriate provisioning.
- There are number of different ways of estimating the probability that a company will default during a particular period of time in the future. One involves bond prices, another involves historical data and a third involves equity prices. The default probabilities backed out from bond prices are risk-neutral probabilities. The probabilities backed out from historical data are real-world probabilities. Real world probabilities should be used for scenario analysis and the calculation of credit VaR. Risk-neutral probabilities should be used for valuat-

ing credit sensitive instrument. In general risk-neutral default probabilities are significantly higher than real-world probabilities.

- A Credit VaR measure can be defined as the credit loss that with a certain probability will not be exceeded during a certain time period. It can be defined to take into account only losses arising from defaults. Alternatively, it can be defined so that it reflects the impact of both defaults and credit rating changes.

## 13.14 Exercises

**Exercise 13.1** *Let* $A = \{(2,0),(0,0)\}$ *and* $B = \{(4,4)\}$ *be two sets in* $\mathbf{R}^2$*. It is claimed that the line* $y = 3$ *is the hard margin classifier. Verify this claim analytically.*

**Exercise 13.2** *Let* $x = (0,2,-3)^T$ *and* $(-2,1,4)^T$*. Determine* $\|(x)_+\|_1$*,* $\|(y)_+\|_1$ *and* $\|(x+y)_+\|_1$ *and* $\|(x)_+ + (y)_+\|_1$*. Show that the polynomial kernel of degree 2 is a Mercer kernel.*

**Exercise 13.3** *Let* $A = \{(0,0),(2,0),(0,2)\}$ *and* $B = \{(0,-1),(1,0),(0,1)\}$ *be two datasets with class labels +1 and -1, respectively.*

1. *Use a polynomial kernel of degree 2 to find a soft margin classifier.*
2. *Write the optimization problem (both the primal and dual forms) if a Gaussian kernel with* $\sigma = 1$ *is employed for finding the soft margin classifier.*

**Exercise 13.4** *Perform the simulation of Merton's Model for Credit-VaR for a Bond portfolio when the correlation 0.4.*

# 14
# Monte Carlo Simulation

## 14.1 Introduction

Monte Carlo methods (or Monte Carlo simulations) are a class of computational algorithms that rely on repeated random sampling to compute their results. The use of Monte Carlo simulations in quantitative finance or specifically in financial engineering for valuation of options was probably first suggested by Phelim Boyle in 1977 [19]. Ever since then the Monte Carlo simulation has become an essential tool in pricing the derivative securities and also in risk management. These applications have, in turn, stimulated research into new Monte Carlo methods and renewed interest in some older techniques. This chapter develops the Monte Carlo methods and study their applications in finance. It also uses simulation, which is a fictitious representation of reality, as a vehicle for presenting models and ideas from financial engineering. The subject of Monte Carlo methods can be viewed as a branch of "experimental mathematics" in which one uses random numbers to conduct experiments. Typically experiments/simulations are carried on systems using anywhere from hundreds to billions of random numbers.

A Monte Carlo method is a numerical method based on random sampling that can be used to solve a mathematical or statistical problem. It provides approximate solutions to a variety of mathematical problems by performing statistical sampling experiments. These methods derive their collective name from the fact that Monte Carlo, the capital of Monaco, has many casinos and casino roulette wheels.

One of the most important uses of Monte Carlo methods is in evaluating difficult multi-dimensional integrals for which we have very few methods for computation. Note that these methods only provide an approximation of the actual value. The attempt to minimize this error is the reason for many different Monte Carlo methods. For example, a model of a random process that produces (or mimic) traffic movements on a particular highway is a simulation. Now suppose, we are interested to know the average number of cars passing through that highway or the probability that the waiting time is more than 5 minutes, then in these cases we need the Monte Carlo simulation.

A Monte Carlo simulation uses repeated sampling to determine the properties of some statistical phenomenon. The Monte Carlo simulation technique has formally existed since the early 1940s, where it had its applications in research into nuclear fusion. One important feature of Monte Carlo simulation is that it is possible to estimate the order of magnitude of estimation error, in terms of statistical confidence intervals. Thus, Monte Carlo simulation includes the distribution of the random variables, analysis of the output and efficiency of the simulation.

In this chapter, we focus on generating random numbers and variables, and Monte Carlo techniques, and show their applications in the area of finance. We shall be referring mainly to Glasserman [49] and Jackel [69].

## 14.2 Generating Random Numbers

The essential feature common to all Monte Carlo computations is that at some point we have to substitute for a random variable a corresponding set of actual values, having the statistical properties of the random variable. The values that we substitute are called random numbers on the grounds that they could well have been produced by a suitable random process.



The efficient ways to obtain random numbers using arithmetic operation is suggested by John von Neumann in 1940's. There are two types of random numbers namely *pseudo-random numbers* and *quasi-random numbers*. The random number sequences generated in a deterministic way is called pseudo-random numbers. Any standard random number generator produces a series of uniformly distributed numbers on the interval $[0,1]$. In software MATLAB, the $rand(n,m)$ function generates uniformly distributed pseudo-random numbers in an $(n\times m)$ matrix. A good generator will have a very large repeat cycle nearly $10^9$. Note that, the sequence of numbers generated by the algorithm is not random but deterministic in the sense that the numbers generated by the algorithm reproduce independence and contain no any discernible information on the next value. This property is often referred to as lack of predictability. In other words, the number generated are spread out evenly across the interval $[0,1]$.

We input a number, generally called a *random number seed*, to start with the generation of random numbers sequence. Then the set of mathematical operations are performed on the random number seed. The resultant generated random numbers are called pseudo-random numbers. These random numbers are tested with rigorous statistical tests to ensure that the numbers are random.

The most common method used for pseudo-random number generation is a recursive technique called the *linear congruence generator* or *Lehmner generator*. It is defined by the recursive formula

$$x_{n+1} = (Ax_n + C) \quad (mod \quad M),$$

where the integers $A, C$ and $M$ are parameters that can be adjusted for convenience and to ensure the desired nature of the sequence of pseudo-random numbers, and *mod* stands for modulus operation. This generator is initiated with a seed $x_0$. In general, the random integers so generated should be mapped into the interval $[0, 1]$. For a linear congruence generator taking possible values $x \in \{0, 1, \ldots, M-1\}$, it is suggested to use $u = x/M$ or $(x+0.5)/M$ to get values which are approximately uniformly distributed in the unit interval $[0, 1]$. Note that, there are many tests that can be applied to determine whether the hypothesis of independent uniform variables is credible.

As an illustration, let $x_0 = 4$, $A = 81$, $C = 35$ and $M = 256$ (8 bits). This generates the sequence

$$\begin{aligned} x_1 &= (81 \times 4 + 35) \ (mod\ 256) = 359 \ mod\ 256 = 103 \\ x_2 &= (81 \times 103 + 35) \ (mod\ 256) = 8378 \ mod\ 256 = 186 \\ x_3 &= (81 \times 186 + 35) \ (mod\ 256) = 15101 \ mod\ 256 = 253 \ . \end{aligned}$$

Therefore, the corresponding $u_i's$ are $u_1 = 103/256 = 0.4023$, $u_2 = 0.726$, $u_3 = 0.98828$.

The above repetition is inevitable for a linear congruential generator. There are at most $M$ possible numbers after reduction mod $M$ and once we arrive back at the seed the sequence is destined to repeat itself. In the example above, the sequence cycles after 256 numbers. The length of one cycle, before the sequence begins to repeat itself again, is called the *period of the generator*.

*Quasi-random numbers* are numbers selected from a quasi-random sequence. Quasi-random sequence is a sequence of $n$-tuples that fills $n$-space more uniformly than uncorrelated random points, sometimes also called a low-discrepancy sequence. Quasi-random numbers are useful in computational problems such as quasi-Monte Carlo integration. Although the ordinary uniform random numbers and quasi-random sequences both produce uniformly distributed sequences, there is a major difference between the two. A uniform random generator on $[0, 1)$ produces outputs so that each trial has the same probability of generating a point on equal subintervals, for example $[0, 1/2)$ and $[1/2, 1)$. Therefore, it is possible for $n$ trials to coincidentally all lie in the first half of the interval, while the $(n+1)^{th}$ point still falls within the other half with probability 1/2. This is not the case with the quasi-random sequences, in which the outputs are constrained by a low-discrepancy requirement that has a net effect of points being generated in a highly correlated manner (i.e. the next point "knows" where the previous points are).

**Illustration 14.2.1: Value of pi using Monte Carlo simulation**

We consider a unit circle within a square with sides equal to 2 (see Fig. 14.1).

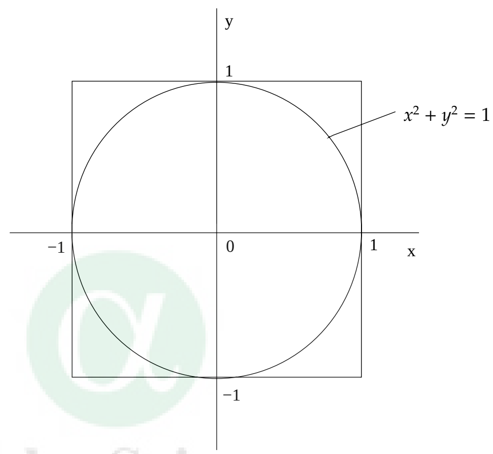





**Fig. 14.1.** Value of pi by Monte Carlo simulation

Now if we pick a random point $(x, y)$ where $x$ and $y$ are between $-1$ and $1$, the probability that this random point lies inside the unit circle is given by the proportion between the area of the unit circle and the square. Thus

$$P(x^2 + y^2 < 1) = \frac{\text{Area of the unit circle}}{\text{Area of the square}} = \frac{\pi}{4} .$$

Therefore if we pick a random point $N$ times in the square and observe that $M$ times the point lies inside the unit circle, the probability that a random point lies inside the unit circle is given by

$$\tilde{P}(x^2 + y^2 < 1) = \frac{M}{N} ,$$

where $\tilde{P}$ indicates that it is a discrete distribution because $M$ and $N$ are integers. But if $N$ becomes very large, then as a consequence of the central limit theorem, the above two probabilities will become equal. This gives

$$\pi = \frac{4M}{N} \; .$$

Therefore to use Monte Carlo simulation, we generate a large number of random $(x, y)$ positions in the square, i.e. $x$ and $y$ are between $-1$ and 1. We next determine which of these positions are inside the circle. Each time it is inside the circle, we add 'one' to the counter. Hence by the methodology of Monte Carlo simulation we have

$$\pi \approx 4 \, \frac{\text{Number of points inside circle}}{\text{Total number of point generated in a square}} \; .$$

## 14.3 Generating Random Variables

In last section we have learned to generate pseudo-random numbers from the uniform distribution $U[0, 1]$. In this section we use this knowledge to learn how to generate a general distributed random numbers. One of the simplest methods of generating random samples from a distribution with cumulative distribution function (CDF) $F(x) = P(X \le x)$ is based on the inverse of the CDF. The CDF is an increasing function, however it is not necessarily continuous. Thus we define the generalized inverse $F^{-}(u) := inf\{x : F(x) \ge u\}, 0 \le u \le 1$. The Fig. 14.2 illustrates its definition. If $F$ is continuous, then $F^{-}(u) = F^{-1}(u)$.

**Theorem 14.3.1** ***(Inversion Method)*** *Let* $U \sim \mathcal{U}[0, 1]$ *and* $F$ *be its CDF. Then* $F^{-}(U)$ *has the CDF* $F$*.*

**Proof.** It is easy to see (e.g. see, Fig. 14.2) that $F^{-}(U) \le x$ is equivalent to $U \le F(x)$. Thus for $U \sim \mathcal{U}[0, 1]$

$$P(F^{-}(U) \le x) = P(U \le F(x)) = F(x) \; ,$$

which shows that $F$ is the CDF of $X$ is $F^{-}(U)$.

□

Inversion method can be summarized as follows

generate $U \sim \mathcal{U}[0, 1]$
Return $X \leftarrow F^{-}(U)$.

**Illustration 14.3.1: Generation of Exponential Distribution**

The exponential distribution with rate $\lambda > 0$ has the CDF

$$F_X(x) = 1 - \exp(-\lambda x) \;\; \text{for} \;\; x \ge 0 \; .$$

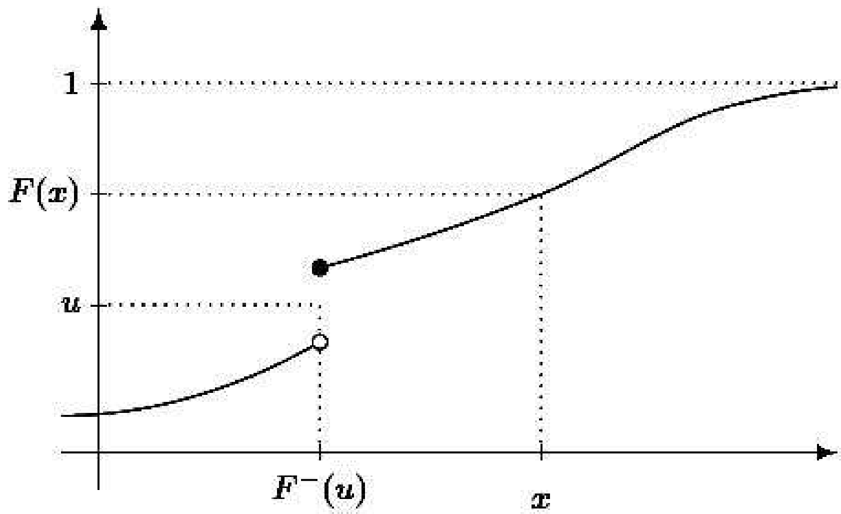



**Fig. 14.2.** Illustration of the definition of the generalized inverse $F^{-}$ of a CDF $F$

Thus

$$F_X^{-}(u) = F_\lambda^{-1}(u) = -\frac{1}{\lambda} ln(1-u) \ .$$

Therefore to generate the given exponential distribution, we first generate $U \sim \mathcal{U}[0,1]$ and then use the inversion method. This is because if $U \sim \mathcal{U}[0,1]$ then $-\frac{1}{\lambda} ln(1-U)$ is exponential distribution with rate $\lambda > 0$. As $U$ and $1-U$ have the same distribution, we can use $-\frac{1}{\lambda} ln(U)$ as well. The details are summarized as follows.

$$\text{generate } U \sim \mathcal{U}[0,1]$$

$$X \leftarrow -\frac{ln(U)}{\lambda}$$

$$(\text{or Return } X \leftarrow -\frac{ln(1-U)}{\lambda}).$$

**Illustration 14.3.2: Generation of a Discrete Distribution**

Let $X$ be a discrete random variable with possible values $a_i$ $(i = 1, 2, \ldots, n)$. We assume that $a_1 < a_2 < \ldots < a_n$. The cumulative distribution function of $X$ is given by

$$q_0 = 0 \ ,$$
$$q_i = \sum_{j=1}^{i} P(X = a_j) = F_X(a_i) \quad (i = 1, 2, \ldots, n) \ .$$

Therefore to generate the given discrete distribution we employ the following steps.
**Step 1.** Generate $U \sim \mathcal{U}[0,1]$.
**Step2.** Find $k \in \{1, 2, \ldots, n\}$ such that $q_{k-1} < U \leq q_k$.
**Step 3.** Set $X = a_k$.

The inversion method is a very efficient tool for generating random numbers. However very few distributions possess CDF whose (generalized) inverse can be evaluated efficiently. For example, CDF of a Gaussian distribution is not even available in the closed form. Note however that the generalized inverse of the CDF is just one possible transformation and that there might be other transformations that yield the desired distribution. An example of such a method is the *Box-Muller method* for generating Gaussian random variables.

**Illustration 14.3.3: Box-Muller Method for Generating Normal Distribution**

This algorithm generates a sample form the standard bivariate normal distribution, each component of which is thus a univariate standard normal variate. This algorithm is based on the below given lemma.

**Lemma 14.3.1** *Let* $X_1, X_2 \sim \mathcal{N}(0,1)$ *and independent, i.e.* $Z = (X_1, X_2) \sim \mathcal{N}(0, \mathbf{I}_2)$, $\mathbf{I}_2$ *being a* $(2 \times 2)$ *identity matrix. Then*

*(i)* $R = X_1^2 + X_2^2$ *is exponential distributed with mean 2, i.e.* $P(R \leq x) = 1 - e^{-x/2}$.
*(ii) Given* $R$, *the point* $(X_1, X_2)$ *is uniformally distributed on the circle of radius* $\sqrt{R}$ *centered at the origin.*
*(iii)* $R$ *and* $\theta = tan^{-1}\left(\frac{X_2}{X_1}\right)$ *are independent random variables.*

The above lemma suggests that to generate $(X_1, X_2)$, we may first generate $R$ and then choose a point uniformally from the circle of radius $\sqrt{R}$. In view of Illustration 14.3.1, we may generate $R$ by setting $R = -2ln(U_1)$ with $U_1 \sim \mathcal{U}[0,1]$.

Next we generate a random point on the circle. For this we generate a random angle $\theta$ between 0 and $2\pi$. This is achieved by setting $\theta = 2\pi U_2$, $U_2 \sim \mathcal{U}[0,1]$. Therefore the corresponding point on the circle has the coordinates $(\sqrt{R}\cos\theta, \sqrt{R}\sin\theta)$. The details of the algorithm are as follows.

generate $U_1, U_2$ independent from $\mathcal{U}[0,1]$
$R \leftarrow -2ln(U_1)$

$$\theta \leftarrow 2\pi U_2$$
$$X_1 \leftarrow \sqrt{R}\cos(\theta), X_2 \leftarrow \sqrt{R}\sin(\theta)$$
$$\text{return } X_1, X_2 \; .$$

The idea of transformation methods, like the inversion method, is to generate random samples from a distribution other than the target distribution and to transform them such that they come from the desired target distribution. In many situations we cannot find such a transformation in closed form. In these cases we have to find other ways of correcting for the fact that we sample from the wrong distribution. The next two algorithms present two such ideas namely the *rejection sampling* and *importance sampling* [102].

**1. Acceptance- Rejection Sampling**

In the sequel we would describe acceptance-rejection Monte Carlo method. The acceptance-rejection method, introduced by Von Neumann, is among the most widely applicable mechanisms for generating random samples. This method generates samples from a target distribution by first generating candidates from a more convenient distribution and then rejecting a random subset of generated candidates. The rejection mechanism is designed so that the accepted samples are indeed distributed according to the target distribution.

The basic idea of rejection sampling is to sample from an instrumental distribution and reject samples that are "unlikely" under the target distribution.

Assume that we want to have a sample from a target distribution whose density $f$ is known to us. The simple idea underlying rejection sampling (and other Monte Carlo algorithms) is the rather trivial identity

$$f(x) = \int_0^{f(x)} 1 du = \int_0^1 1_{0 \le u \le f(x)} du \; .$$

Thus $f(x)$ can be interpreted as the marginal density of a uniform distribution on the area under the density $f(x)$, i.e. $\{(x, u) : 0 \le u \le f(x)\}$. The Fig. 14.3 illustrates this idea. It suggests that we can generate a sample from f by sampling from the area under the curve.

**Illustration 14.3.4: Sampling from a Beta distribution**

The Beta$(a, b)$ distribution $(a, b \ge 0)$ has the density

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1,$$

where $\Gamma(a) = \int_0^{+\infty} t^{a-1}e^{-t}dt$ is the Gamma function. For $a, b > 1$ the Beta$(a, b)$ density is unimodal with mode $\dfrac{(a-1)}{(a+b-2)}$. The Fig. 14.3 shows the density of

Beta$(3, 5)$ distribution. It attains its maximum of $1680/729 \approx 2.305$ at $x = 1/3$. Sampling from the area under the curve (dark gray) corresponds to sampling from the Beta $(3, 5)$ distribution. We use a uniform distribution of the light gray rectangle as proposal distribution. Empty circles denote rejected values, filled circles denote accepted values.

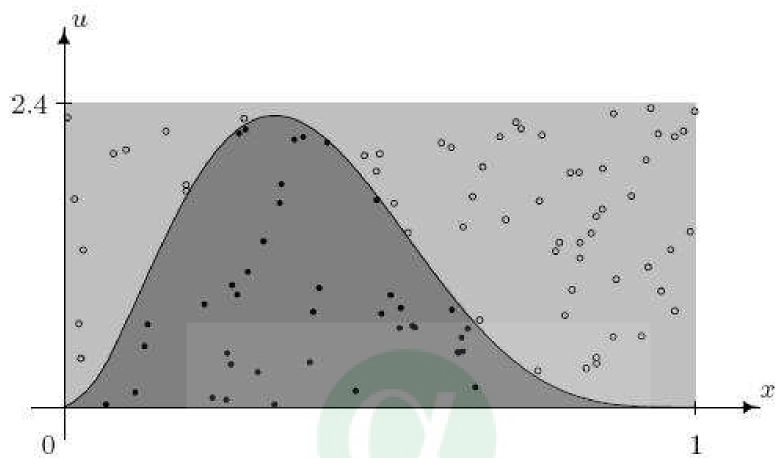



**Fig. 14.3.** Sampling from a Beta(3,5) distribution



Using the above identity we can draw sample from Beta(3, 5) by drawing sample from a uniform distribution on the area under the density $\{(x, u) : 0 \leq u \leq f(x)\}$ (the area shaded in dark gray in Fig. 14.3). We will sample from the light gray rectangle and only keep the samples that fall in the area under the curve.

Mathematically speaking, we sample independently $X \sim \mathcal{U}[0, 1]$ and $U \sim \mathcal{U}[0, 2.4]$. We keep the pair $(X, U)$ if $U < f(X)$, otherwise we reject it. The conditional probability that a pair $(X, U)$ is kept if $X = x$ is $P(U < f(X)/X = x) = P(U < f(x)) = f(x)/2.4$. As $X$ and $U$ were drawn independently we can rewrite our algorithm as

*Draw X from* $\mathcal{U}[0, 1]$ *and accept X with probability* $f(X)/2.4$, *otherwise reject X.*

The method proposed in the above example is based on bounding the density of the Beta distribution by a box. Whilst this is a powerful idea, it cannot be directly applied to other distributions, as the density might be unbounded or have infinite support. However we might be able to bound the density of $f(x)$ by $Mg(x)$, where $g(x)$ is a density that we can easily sample from.

**Algorithm** (Rejection Sampling) Given two densities $f, g$ with $f(x) < Mg(x)$ for all $x$, we can generate a sample from $f$ by employing following steps.
**Step 1.** Draw X from density $g$.
**Step 2.** Accept X as a sample from $f$ with probability $\dfrac{f(X)}{Mg(X)}$, otherwise go back to Step 1. This step is equivalent to the following step.
Generate $U \sim \mathcal{U}[0,1]$. If $U \leq f(x)/Mg(x)$, accept X, otherwise go to Step 1.

The above algorithm can be justified mathematically as detailed below. We have

$$P(X \in \mathcal{X} \text{ and is accepted}) = \int_{\mathcal{X}} g(x)\, \frac{f(x)}{Mg(x)}\, dx = \frac{\int_{\mathcal{X}} f(x)\, dx}{M} ,$$

and thus

$$P(X \text{ is accepted}) = P(X \in \mathcal{X} \text{ and is accepted}) = \frac{1}{M} .$$

$\mathcal{X}$ is the domain of random variable X. This yields

$$\begin{aligned} P(x \in \mathcal{X}/X \text{ is accepted}) &= \frac{P(X \in \mathcal{X} \text{ and is accepted})}{P(X \text{ is accepted})} \\ &= \frac{\int_{\mathcal{X}} f(x)\, dx/M}{1/M} \\ &= \int_{\mathcal{X}} f(x)\, dx . \end{aligned}$$



Thus the density of the values accepted by the algorithm is $f(\cdot)$.

**Remark 14.3.1** *If we know $f$ only up to a multiplicative constant, i.e. if we only know $\pi(x)$, where $f(x) = C\pi(x)$, we can carry out rejection sampling using*

$$\frac{\pi(X)}{Mg(X)}$$

*as probability of accepting X, provided $\pi(x) < Mg(x)$ for all $x$.*

**Illustration 14.3.5: Rejection sampling from $\mathcal{N}(0,1)$ distribution using a Cauchy proposal**

Assume we want to sample from the $\mathcal{N}(0,1)$ distribution with density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right), \quad -\infty < x < \infty,$$

using the Cauchy distribution with density

$$g(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

as instrumental distribution. The smallest $M$ we can choose such that $f(x) \leq Mg(x)$ is $M = \sqrt{2\pi}\,\exp\left(\frac{-1}{2}\right)$.

The Fig. 14.4 illustrates the results. As before, filled circles correspond to accepted values whereas open circles correspond to rejected values. Sampling from the area under the density $f(x)$ (dark gray) corresponds to sampling from the $\mathcal{N}(0,1)$ density. The proposal $g(x)$ is a Cauchy$(0,1)$.

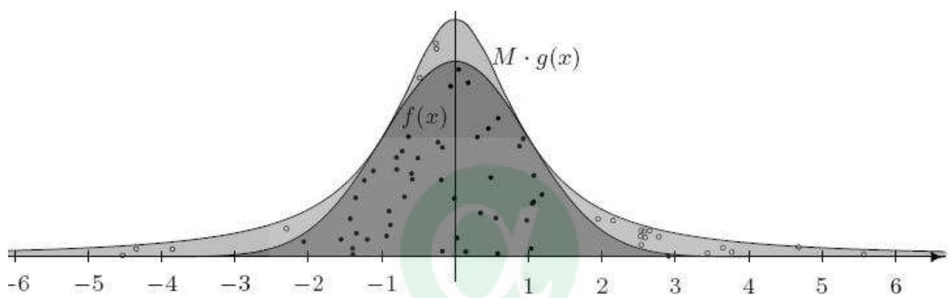



**Fig. 14.4.** Rejection sampling from $\mathcal{N}(0,1)$ distribution using a Cauchy proposal



Note that it is impossible to do rejection sampling from a Cauchy distribution using a $N(0,1)$ distribution as instrumental distribution. This is because there is no $M \in \mathbf{R}$ such that

$$\frac{1}{\pi(1+x^2)} < M\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-x^2}{2}\right),$$

as the Cauchy distribution has heavier tails than the Gaussian distribution.

## 2. Importance Sampling

In rejection sampling we have compensated for the fact that we sampled from the instrumental distribution $g(x)$ instead of $f(x)$ by rejecting some of the values proposed by $g(x)$. Importance sampling is based on the idea of using weights to correct for the fact that we sample from the instrumental distribution $g(x)$ instead of the target distribution $f(x)$.

Importance sampling is based on the identity

$$P(X \in \mathcal{X}) = \int_{\mathcal{X}} f(x)\, dx = \int_{\mathcal{X}} g(x)\, \frac{f(x)}{g(x)}\, dx = \int_{\mathcal{X}} g(x)\, w(x)\, dx\ ,$$

for all $g(\cdot)$, such that $g(x) > 0$ for (almost) all $x$ with $f(x) > 0$.

Suppose we are interested in estimating

$$\alpha = E[h(X)] = \int_{\mathcal{X}} h(x)\, f(x)\, dx = \int_{\mathcal{X}} h(x)\, \frac{f(x)}{g(x)}\, g(x)\, dx = \int_{\mathcal{X}} h(x)\, w(x)\, g(x)\, dx\ .$$

We can have the following algorithm based on the principle of importance sampling.

**Algorithm** (Importance Sampling)
Choose $g$ such that $supp(fh) \subset supp(g)$, where supp is support set i.e. $supp(g) = \{x : g(x) > 0\}$. Follow the below given steps.

**Step 1.** For $i = 1, \ldots, n$: generate $X_i$ from density $g$.

**Step 2.** Set $w(X_i) = \dfrac{f(X_i)}{g(X_i)}$.

**Step 3.** Return either $\hat{\mu} = \dfrac{\sum_{i=1}^{n} w(X_i)\, h(X_i)}{\sum_{i=1}^{n} w(X_i)}$ or $\tilde{\mu} = \dfrac{\sum_{i=1}^{n} w(X_i)\, h(X_i)}{n}$ .

In simulation, certain values of the input random variables have more impact on the parameter being estimated than others. If these influential values are generated by sampling more frequently, then the variance of estimator can be reduced. This is possible by choosing another distribution to the random variables. This method of importance sampling is called a *variance reduction technique*.

To look this differently, we observe that the error associated with a Monte Carlo estimate is proportional to $\sigma$ and inversely proportional to the square root of the number of trials. Hence, there are only two ways of reducing the error in a Monte Carlo estimate. These are either to increase the number of trials or to reduce the variance. We introduce importance sampling techniques that reduce $\sigma$ and there by improve the efficiency of each trial. This is done by changing the probability measure which is a standard tool in financial mathematics.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X$ be a random variable with pdf $f(x)$ under $P$. Suppose we want to estimate $\theta = E[h(x)]$ under $P$. Let $g(x)$ be another pdf of the random variable $X$ under $Q$. Then

$$\theta = E_P[h(x)] = \int h(x)\, f(x)\, dx = \int h(x)\, g(x)\, dQ(x) = E_Q[h(x)]\ ,$$

where $dQ(x) = \frac{f(x)}{g(x)}\, dx$. Note that, $E_P[\theta] = E_Q[\theta]$ where as $Var_Q[\theta] < Var_P[\theta]$. Hence, we should generate the samples of the random variable $X$ using pdf $g(x)$ instead of $f(x)$.

A common pitfall of importance sampling is that the tails of the distributions matter. While $g(x)$ might be roughly the same shape as $f(x)$, serious difficulties arise if $g(x)$ gets small much faster than $f(x)$ out in the tails. In such a case, though it is improbable (by definition) that we will realize a value $x_i$ from the far tails of $g(x)$, if we do so then the Monte Carlo estimator will take a jolt $\dfrac{f(x_i)}{g(x_i)}$ for such an improbable $x_i$ which may be orders of magnitude larger than the typical values $\dfrac{f(x)}{g(x)}$ that we see.

**Illustration 14.3.6: Geometric Distribution with Parameter** $p$.

This is a discrete distribution which describes the number of independent trials necessary to achieve a single success with the probability of a success on each trial is $p$. We know that the probability mass function is

$$p(i) = p(1-p)^i, \quad (i = 1, 2, \ldots)\ ,$$

and the cumulative distribution function is

$$F(x) = P(X \leq x) = 1 - (1-p)^{[x]}, \quad x \geq 0\ ,$$

where $[x]$ denotes the integer part of $x$. We wish to output an integer value of $x$ which satisfies the inequalities

$$F(x-1) < U \leq F(x)\ .$$

Solving these inequalities for integer $x$,we obtain

$$1 - (1-p)^{x-1} < U \leq 1 - (1-p)^x$$

$$(1-p)^{x-1} > 1 - U \geq (1-p)^x$$

$$(x-1)\ ln(1-p) > ln(1-U) \geq x\ ln(1-p)$$

$$(x-1) < \frac{ln(1-U)}{ln(1-p)} \leq x\ .$$

We should therefore choose the smallest integer for $X$ which is greater than or equal to $\dfrac{ln(1-U)}{ln(1-p)}$ or equivalently, $X = 1 + \left[\dfrac{ln(1-U)}{ln(1-p)}\right]$ or $1 + \left[\dfrac{-E}{ln(1-p)}\right]$ where we write $-ln(1-U) = E$, an exponential distributed random variable with parameter 1. In MATLAB, the geometric random number generators is called *geornd*.

## 14.4 Monte Carlo Techniques

There are many considerations when using Monte Carlo techniques to perform approximations. One of the chief concern is to get as accurate an approximation as possible. Thus, for each method it is advisable to discuss their associated error statistics. In this section we shall discuss Monte Carlo methods, namely, *crude Monte Carlo method* and *stratified sampling method* [86, 110] in the context of evaluation of integrals.

Consider the evaluation of following integral

$$I = \int_a^b f(x)\, dx \ .$$

We generate $N$ random sample values, $x_i$, $(i = 1, 2, \ldots, N)$ in $[a, b]$. Next find $f(x_i), (\ i = 1, 2, \ldots, N)$ and obtain

$$I_N = \frac{b-a}{N} \sum_{i=1}^{N} f(x_i) \ .$$

It is well known that $I_N$ is unbiased, i.e. its expectation is $I$ and, because of law of large numbers, $I_N \to I$ with probability 1 as $n \to \infty$. The approximation error is of order $\bigcirc(n^{-1/2})$. This follows from central limit theorem, or can be viewed through the mean square error (MSE) $\sqrt{E(I_N - I)^2} = \sqrt{VarI_N} = \frac{\sigma}{\sqrt{n}}$.

We next describe *stratified sampling method*. Consider again the following integral

$$I = \int_a^b f(x)\, dx \ .$$

The basic principle of stratified sampling method is to divide the interval $[a, b]$ into $n$ subintervals and then to perform a crude Monte Carlo method on each subinterval. Thus we write

$$\int_a^b f(x)\, dx = \int_a^{a_1} f(x)\, dx + \int_{a_1}^{a_2} f(x)\, dx + \ldots + \int_{a_{n-1}}^{b} f(x)\, dx \ ,$$

and then apply crude Monte Carlo method in each of the integrals of the right hand side. The reason we might use this method is that instead of finding variance in one go, we can find the variance by adding up the variances of each subinterval. This method is generally helpful when function is step like or has periods of flat. Thus advantage of stratified sampling method is that we get to split the curve into parts that could have certain advantageous properties when evaluating them on their own.

## 14.5 Applications in Finance

In this section, we present the applications of Monte Carlo simulation in option pricing and risk management. Monte Carlo methods are ideal for pricing options where the payoff is path dependent or options where the payoff is dependent on a basket of underlying assets (rather than just a single asset).

The first application to European call option pricing was by Phelim Boyle in 1977 [19]. As with other option pricing techniques Monte Carlo methods are used to price options using what is essentially a three step process. The three steps are as follows

**Step 1.** Calculate potential future prices of the underlying asset(s).

**Step 2.** Calculate the payoff of the option for each of the potential underlying price paths.

**Step 3.** Discount the payoffs back to today and average them to determine the expected price.

The first step in using Monte Carlo methods is to generate (a large number of) potential future asset prices. This is done by selecting an appropriate (stochastic) model for the time evolution of the underlying asset(s) and then simulating the model through time.

For example, the standard model for evolution of equity prices is given by the Wiener process which is described as

$$S(\triangle t) = S(0) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\triangle t + (\sigma \sqrt{(\triangle t)})\epsilon\right).$$

Here $S(0)$ is the stock price today, $S(\triangle t)$ is the stock price at a (small) time into the future, $\triangle t$ is a small increment of time, $\mu$ is the expected return, $\sigma$ is the expected volatility, $\epsilon$ is (random) number sampled from the standard normal distribution.

Typically many thousands, if not tens of thousands, of simulated paths must be generated to enable an accurate option price to be calculated. The more paths that are generated the longer the simulation will take to be performed and hence the longer the time taken to price the option.

Consider Monte Carlo simulation of European call option on a stock whose current value is $S(0)$. Let $S(t)$ be the stock price at time $t$. The option expires at time $T$ and the strike price is $K$. We assume constant interest rate $r$ and the stock price follows a lognormal distribution with volatility parameter $\sigma$.

We know that, the price of the option is its discounted expected value. The return on this simulation is the discounted difference between the terminal value of the stock and the strike price. Then, we repeat this many a times, averaging the discounted returns to estimate the present value of the option.

We present the steps involved as follows.

**Step 1.** Generate several random price paths.
**Step 2.** Calculate the associated exercise value for each path.
**Step 3.** Find the average payoff.
**Step 4.** Find discounted average payoff.

Now we present the Monte Carlo algorithm for this option pricing. By using random sampling of normal distribution and then using the Black-Scholes model, the Monte Carlo algorithm can be described as follows.

For i = 1:N

$$E(i) = normalrnd(0,1);$$
$$S(i) = S(0)\ \exp((r - 0.5\,\sigma^2)T + \sigma\sqrt{T}E(i));$$
$$V(i) = \exp\left(-rT\,\mathrm{Max}\,(S(i) - K, 0)\right);$$

end

$$\hat{A}_N = \frac{1}{N}\sum_{i=1}^{N} V(i).$$



Note that estimator $\hat{A}_N$ will be unbiased and consistent estimator since $E(\hat{A}_N) = A$ and $\hat{A}_N \to A$ with probability 1 as $n \to \infty$.

In the above mechanism, the payoff is determined by the terminal stock price $S(T)$ and does not otherwise depend on the evolution of $S(t)$ between times 0 and $T$. Each simulated path of the underlying asset consists of two points $S(0)$ and $S(T)$.

In order to obtain a more accurate approximation to sampling we divide the time interval $[0,T]$ into smaller subintervals, we simulate the payoff of a derivative security which may depend explicitly on the values of underlying asset at multiple time points. We present the algorithm as follows. For $0 = t_0 < t_1 < \ldots, t_n = T$,

$$S(i+1) = S(i)\ \exp((r - 0.5\,\sigma^2)(t_{i+1} - t_i) + \sigma\ \sqrt{t_{i+1} - t_i}E(i))\ .$$

We next illustrate the Monte Carlo method for interest rate model.

**Illustration 14.5.1: Interest Rate Model** Consider

$$dr_t = \alpha(b - r_t)dt + \sigma\sqrt{r_t}dW_t,\quad t \in [0,T]\ , \tag{14.1}$$

where $\alpha$ and $b$ are positive constants. If $r_0 > 0$, then $r_t$ will never be negative. If $2\alpha b \geq \sigma^2$, then $r_t$ remains strictly positive for all $t$, at most surely.

A simple Euler discretization of (14.1) suggests simulating $r_t$ at times $0 = t_0 < t_1 < \ldots < t_n = T$ by setting

$$r_{t_{i+1}} = r_{t_i} + \alpha(b - r_{t_i})(t_{i+1} - t_i) + \sigma \sqrt{r_{t_i}^+}\left(\sqrt{t_{i+1} - t_i}\right) Z_{i+1}$$

with $Z_i$, $(i = 1, 2, \ldots, n)$ i.i.d random variables each having $\mathcal{N}(0, 1)$. Here $r_{t_i}^+ = \text{Max}(r_{t_i}, 0)$.

Using probability theory, observe that for $0 < u < t$, given $r_u$, $r_t$ is distributed as

$$r_t = c\left(Z + \sqrt{\lambda}\right)^2 + \chi^2_{d-1}$$

where $Z \sim \mathcal{N}(0, 1)$, $\chi^2$ is chi-square distribution with $d-1$ degrees of freedom and

$$c = \frac{\sigma^2\left(1 - e^{-\alpha(t-u)}\right)}{4\alpha}; \quad d = \frac{4\ b\ \alpha}{\sigma^2}; \ \lambda = \frac{4\ r_u\ \alpha\ e^{-\alpha(t-u)}}{\sigma^2\left(1 - e^{-\alpha(t-u)}\right)} = \frac{r_u e^{-\alpha(t-u)}}{c}\ .$$

We now present the details of algorithm.



**Case 1:** $d > 1$.

For $i = 1, 2, \ldots, n-1$

**Step 1.** $c_i = \dfrac{\sigma^2\left(1 - e^{-\alpha(t_{i+1}-t_i)}\right)}{4\alpha}, \qquad \lambda_i = \dfrac{4\ r_{t_i}\ \alpha e^{-\alpha(t_{i+1}-t_i)}}{c}$

**Step 2.** Generate $Z_i \sim \mathcal{N}(0, 1)$

**Step 3.** Generate $X_i \sim \chi^2_{d-1}$

**Step 4.** $r_{t_{i+1}} \leftarrow c\left(Z_i + \sqrt{\lambda_i}\right)^2 + X_i$

end

**Case 2:** $d \leq 1$.

For $i = 1, 2, \ldots, n-1$

**Step 1.** $c_i = \dfrac{\sigma^2\left(1 - e^{-\alpha(t_{i+1}-t_i)}\right)}{4\alpha}, \qquad \lambda_i = \dfrac{4\ r_{t_i}\ \alpha e^{-\alpha(t_{i+1}-t_i)}}{c}$

**Step 2.** Generate $N \sim$ Poisson $(\lambda/2)$

**Step 3.** Generate $X_i \sim \chi^2_{d+2N}$

**Step 4.** $r_{t_{i+1}} \leftarrow cX_i$

end

**Illustration 14.5.2: Portfolio Optimization**

Let us consider a portfolio consisting of two stocks $A$ and $B$. Let $S_A(0)$ and $S_B(0)$ respectively denote the process of these stocks at time $t = 0$. The prices of these stocks at time $t$ are denoted by $S_A(t)$ and $S_B(t)$. At time $t = 0$, the portfolio value $V(0)$ is given by

$$V(0) = n_A S_A(0) + n_B S_B(0)\ ,$$

where the portfolio consists of $n_A$ units of stock $A$ and $n_B$ units of stock $B$. Also let us suppose $S_A(t)$ and $S_B(t)$ follow the geometric Brownian motion with drift $\mu_A$ and volatility $\sigma_A$ and drift $\mu_B$ and volatility $\sigma_B$ respectively. Further, let $S_A(t)$ and $S_B(t)$ be independent. We are interested in finding the estimate of probability that the value of portfolio drops by more than 15% at maturity date $T$ using Monte Carlo simulation.

The desired probability is $\theta = P\left(\frac{V(T)}{V(0)} \leq 0.85\right)$. The steps for estimating $\theta$ are as follows.

For $i = 1$ to $N$; generate $X_i = (S_A^{(i)}(T), S_B^{(i)}(T))$.
Compute

$$I(X_i) = \begin{cases} 1, & \text{if } \dfrac{n_A S_A^{(i)}(T) + n_B S_B^{(i)}(T)}{n_A S_A(0) + n_B S_B(0)} \leq 0.85 \\ 0, & \text{if otherwise} \end{cases}$$

end.
Set $\hat{\theta}_N = \frac{1}{N}\left(I(X_1) + I(X_2) + \ldots + I(X_N)\right)$.
The estimator $\hat{\theta}_N$ will be unbiased and consistent estimator since $E(\hat{\theta}_N) = \theta$ and $\hat{\theta}_N \to \theta$ with probability 1 as $n \to \infty$.



## 14.6 Summary and Additional Notes

- In this chapter we discussed linear congruential generators to generate pseudo random numbers. However there are few other generators available in the literature. One of these is the mixed congruential generator which is a generalization of the linear congruential generators.
- Most common random numbers that have been talked about in the financial simulations are random numbers generated from uniform distribution and normal distribution. However other random numbers that have been generated from non-uniform continuous distributions using inverse transform are exponential random numbers, Cauchy random numbers, Pareto random numbers, extreme values and logistic random numbers. Some of them are very useful for financial applications that involve fat tail distributions.

- For most of the distributions MATLAB have readily available functions to generate random numbers with those distributions. For e.g. such as normal: normrnd($\mu, \sigma, 1, n$), Students t: trnd($\nu, 1, n$), exponential: exprnd($\lambda, 1, n$), uniform: unifrnd($a, b, 1, n$) or rand($1, n$) if $a = 0, b = 1$, Weibull: weibrnd($a, b, 1, n$), gamma: gamrnd($a, b, 1, n$), Cauchy: $a+b*trnd(1, 1, n)$, Poisson: poissrnd($\lambda, 1, n$), binomial: binornd($m, p, 1, n$),.
- There are variance reduction techniques available in the literature to improve the speed and efficiency of a simulation. Some of these are antithetic sampling, control variates, variance reduction by conditioning and stratified sampling. We may refer to Glasserman et al [50] in this regard.
- Quasi-Monte Carlo methods are purely deterministic, numerical analytic methods in the sense that they do not even attempt to emulate the behavior of independent uniform random variables, but rather cover the space in d-dimensions with fewer gaps than independent random variables would normally admit. Quasi Monte Carlo frequently generates estimates superior to the Monte Carlo methods in many problems of low or intermediate effective dimension. If the dimension is large, but a small number of variables determine most of the variability in the simulation, then we might expect Quasi Monte-Carlo methods to continue to perform well. The common method to generate Quasi Monte Carlo numbers are the Halton sequence [56], the Sobol sequence [123].
- In this chapter we have attempted to cover some of the most basic aspects of Monte Carlo methods in quantitative finance. For further details the readers may refer to Glasserman [49] and Jackel [69].



## 14.7 Exercises

**Exercise 14.1** *Consider the linear congruential generator* $x_{n+1} = Ax_n + C \pmod{M}$ *with* $M = 64$. *Generate the pseudo-random numbers with* $A = 4, C = 1$ *and* $x_0 = 2$. *Determine the period of the generator starting with seed* $x_0 = 2$ *and with seed* $x_0 = 3$.

**Exercise 14.2** *Consider the uniformly distributed random variable X on the interval* $[0, 1]$. *Find a function of X which is uniformly distributed on the interval* $[0, 2]$. *Generate the sequence of random numbers of uniformly distributed on the interval* $[0, 2]$.

**Exercise 14.3** *Using inverse transform method, develop an algorithm for simulation of Poisson distribution with parameter* $\lambda$.

**Exercise 14.4** *Evaluate* $\int_0^1 e^{-x^4} dx$, *by the crude Monte Carlo method. Also, evaluate it by acceptance-rejection method.*

**Exercise 14.5** *Consider a multivariate normal random vector* $X = (X_1, X_2, \ldots, X_n)$ *having mean vector* $(\mu_1, \mu_2, \ldots, \mu_n)$ *and covariance matrix* $\sum$ *of order* $n \times n$ *matrix. We wish to generate this* $n$*-dimensional random vector. The procedure involves a decomposition of* $\sum$ *into factors such that* $A'A = \sum$. *Suppose* $Z = (Z_1, Z_2, \ldots, Z_n)$ *is a vector of independent standard normal random variable, the required* $n$*-dimensional random vector is given by* $X = (\mu_1, \mu_2, \ldots, \mu_n) + ZA$ .

**Exercise 14.6** *Consider a Vasicek interest rate model*

$$dr_t = \alpha(b - r_t)dt + \sigma dW_t, \ t \in [0, T],$$

*where* $\alpha, b$ *and* $\sigma$ *are positive constants. Prove the algorithm for the exact simulation at time* $0 = t_0 < t_1 < \ldots < t_n = T$ *is given by* $r_{t_n} = e^{-\alpha(t_{i+1}-t_i)r_{t_i}} + b\left(1 - e^{-\alpha(t_{i+1}-t_i)}\right) + \sigma\sqrt{\frac{1}{2\alpha}\left(1 - e^{-2\alpha(t_{i+1}-t_i)}\right)}\, Z_{i+1}$ *where* $Z_i$ *is the standard normal sample.*
*(Hint: refer to Chapter 11.)*

# 15
# MATLAB Codes for Selected Problems in Finance

## 15.1 Introduction

MATLAB has a very useful toolbox, namely Financial Toolbox, which contains many standard programs of finance. The aim of presenting the MATLAB codes in this book is to encourage readers to attempt writing codes for similar problems in financial mathematics, taking these codes as building blocks.

% at the beginning of any statement in MATLAB code stands for comment lines, which are added in the codes to facilitate understanding of programs.



## 15.2 Binomial Lattice Model for European and American Call Option/Put Option

```
% Binomial Lattice Model for European and American call
% option/put option  with generic input parameter values.
% Symbols have their usual meanings.

% Starting the code.

% Generic input parameter values.

  S0 = input('S0 = ');
  u  = input('u = ');
  d  = input('d = ');
  T  = input('T = ');
  k  = input('k = ');  % Strike price
  r  = input('r = ');
```

```
  R = 1 + r;

% For instance, we can input
% S0 = 100; u = 1.3; d = 0.9; T = 3; k = 110; R = 1.1.

%risk neutral probability measure is p

p = (R-d)/(u-d);

% Finding of binomial lattice for stock market movements.

BinomialLattice = zeros(T,T);
r = 0;            % power of d in the binomial lattice
for i =T:-1:1     % row generation
    r = 1;
    % i+j = T+1 => j= T+1-i   % constructing a lower triangle
    for j = T+1-i:T           % column generation
        BinomialLattice(i,j) = S0*u^(T-i)*d^(r-1);
        r = r+1;
    end
end disp('Binomial lattice for stock movement ');
disp(BinomialLattice);

% Asking for American or European call/put options.

fprintf('\n 1 : for American options
        \n 2 : for European options \n ');
    choice_A_E = input('Your choices ');
    switch choice_A_E
    case 1
    fprintf('\n 1 : for call option \n 2 : for put option ');
    choice_p_c = input('your choice ');
    c0 = zeros(T,T);
    c0(:,T) = max(BinomialLattice(:,T) - k, 0);
    if choice_p_c == 1
    for j = T-1:-1:1        % j = column traversal
    for i = T:-1:2      % i = row traversal
    if i+j >=T+1    % construct a lower triangle
```

```
c0(i,j) = max(BinomialLattice(i,j) - k,
(1/R)*(p*c0(i-1,j+1)+(1-p)*c0(i,j+1)));
end
end
end
disp('American call option binomial lattice mm');
disp(c0);
else
p0 = zeros(T,T);
p0(:,T) = max(k - BinomialLattice(:,T), 0);
for j = T-1:-1:1        % j = column traversal
for i = T:-1:2      % i = row traversal
if i+j>=T+1
p0(i,j) = max(k - BinomialLattice(i,j),
(1/R)*(p*p0(i,j+1)+(1-p)*p0(i-1,j+1)));
end
end
end
disp('American put option binomial lattice ');
disp(p0);
end
break;
case 2
fprintf('\n 1 : for call option \n 2 : for put option ');
choice_p_c = input('your choice ');
c0 = zeros(T,T);
c0(:,T) = max(BinomialLattice(:,T) - k, 0);
if choice_p_c == 1
for j = T-1:-1:1        % j = column traversal
for i = T:-1:2      % i = row traversal
if i+j >=T+1    % construct lower triangle
c0(i,j) =(1/R)*(p*c0(i-1,j+1)+(1-p)*c0(i,j+1));
end
end
end
disp('European call option binomial lattice ');
disp(c0);
else
```

```
    p0 = zeros(T,T);
    p0(:,T) = max(k-BinomialLattice(:,T),0);
    for j = T-1:-1:1        % j = column traversal
    for i = T:-1:2      % i = row traversal
    if i+j>=T+1
    p0(i,j) =(1/R)*(p*p0(i,j+1)+(1-p)*p0(i-1,j+1));
    end
    end
    end
    disp('European put option binomial lattice ');
    disp(p0);
    end
    break;
    otherwise
    disp('Unknown Choice.');
end
```

## 15.3 CRR Model for European and American Call Option/Put Option



```
% CRR model for European and American call option/put option
% with generic input parameter values.
% Symbols have their usual meanings.

% Starting the code.

% Generic input parameter values.

  S0 = input('S0 = ');
  T  = input('T = ');   % Expiration time
  n  = input('n = ');   % no of periods
  k  = input('k = ');   % Strike price
  r  = input('r = ');
  sigma = input('Sigma = ')     % volatility of the stock price.

% For example we can take
%S0 = 100; T = 4; n = 2; k = 110; r = 0.05; sigma = 0.24.
```

```
deltat = T/n;
R = exp(r*deltat);
u = exp(sigma*sqrt(deltat));
d = 1/u;
p = (R-d)/(u-d);        % risk neutral probability measure.

% Finding of binomial lattice for stock market movements.

BinomialLattice = zeros(T,T);
r=0;            %  power of d in the binomial lattice
for i = T:-1:1  % row generation
    r = 1;
    % i+j = T+1 =>j= T+1-i    % create lower triangle display
    for j = T+1-i:T           % column generation
        BinomialLattice(i,j) = S0*u^(T-i)*d^(r-1);
        r = r+1;
    end
end disp('Binomial lattice for stock movement ');
disp(BinomialLattice);


% Asking choice for American or European call/put option.


fprintf('\n 1 : for American options
    \n 2 : for Europen options \n ');
    choice_A_E = input('Your choices '); switch choice_A_E

    case 1
        fprintf('\n 1 : for call option \n 2 : for put option ');
        choice_p_c = input('your choice ');
        c0 = zeros(T,T);
        c0(:,T) = max(BinomialLattice(:,T) - k,0);
        if choice_p_c == 1
        for j = T-1:-1:1       %j = column traversal
        for i = T:-1:2      %i = row traversal
        if i+j >=T+1    % create lower triangle
```

```
        c0(i,j) = max(BinomialLattice(i,j) - k,
        (1/R)*((1-p)*c0(i-1,j+1)+(p)*c0(i,j+1)));
        end
        end
        end
        disp('American call option binomial lattice mm');
        disp(c0);
        else
        p0 = zeros(T,T);
        p0(:,T) = max(k-BinomialLattice(:,T),0);
        for j = T-1:-1:1        %j = column traversal
        for i = T:-1:2      %i = row traversal
        if i+j>=T+1
        p0(i,j) = max(k - BinomialLattice(i,j),
        (1/R)*(p*p0(i,j+1)+(1-p)*p0(i-1,j+1)));
        end
        end
        end
        disp('American put option binomial lattice ');
        disp(p0);
        end
        break;
    case 2
        fprintf('\n 1 : for call option \n 2 : for put option ');
        choice_p_c = input('your choice ');
        c0 = zeros(T,T);
        c0(:,T) = max(BinomialLattice(:,T)-k,0);
        if choice_p_c == 1
        for j = T-1:-1:1        % j = column traversal
        for i = T:-1:2      % i = row traversal
        if i+j >=T+1    % To make lower triangle
        c0(i,j) =(1/R)*((1-p)*c0(i-1,j+1) + (p)*c0(i,j+1));
        end
        end
        end
        disp('European call option binomial lattice ');
        disp(c0);
        else
```

```
        p0 = zeros(T,T);
        p0(:,T) = max(k - BinomialLattice(:,T),0);
        for j = T-1:-1:1        %j = column traversal
        for i = T:-1:2      %i = row traversal
        if i+j>=T+1
        p0(i,j) = (1/R)*((p)* p0(i,j+1) + (1-p)*p0(i-1,j+1));
        end
        end
        end
        disp('European put option binomial lattice ');
        disp(p0);
        end
        break;
        otherwise
        disp('Unknown Choice.');
end
```

## 15.4 Black Scholes Formula for Dividend and Non Dividend Paying Stock

```
% Black Scholes formula for dividend and non dividend paying stock.
% Symbols have their usual meanings.

% Starting the code.

 s0 =input('Enter the stock price at t = 0 : S0 =  ');
 k = input('Enter the strike price for underlying asset ');
 r = input('Enter the risk free rate of interest ');
 T =  input('Enter the expiry time ');
 sigma=input('Enter the volatility for underlying asset ');

% For example, s0 = 100; k = 125; r = 0.05;
% T = 3/12; sigma = 0.24.

CH = 1;

while CH == 1
```

```
fprintf('1 : for Dividend \n 2 : non Dividend \n ');
choice = input(' choice ');
switch choice       % for Dividend and non dividend
    case 1
        disp('Data for dividend paying stocks ');
        dtime = input('Enter the time in array  ');
        damount = input('Enter the dividend amount in array ');
        rd = -r*dtime;
        erd = exp(rd);
        ds = damount.*erd;
        sa = s0 - sum(sum(ds));
        lsa = (log(sa/k) + (r + (sigma * sigma)/2)*T);
        d1 = lsa/(sigma * sqrt(T));
        d2 = d1 - sigma * sqrt(T);

        fprintf('1 : call option \n2 : put option \n ');
        ch = input(' choice for put or call ');
        if ch == 1          % for call option
        c0 = sa*normcdf(d1) - k * exp(-r*T) * normcdf(d2);
        fprintf('Value of call option = %4.4f\n',c0);
        end
        if ch == 2          % for put option
        p0 = k * exp(-r*T) * normcdf(-d2)- sa * normcdf(-d1);
        fprintf('Value of put option = %4.4f\n',p0);
        end
        if ch ~= 1 && ch ~= 2       % ~ means not equal to
            disp('Invalid choice');
        end
    case 2          % for non dividend
        lso = (log(s0/k) + (r + (sigma. * sigma)/2)*T);
        d1 = lso/(sigma * sqrt(T));
        d2 = d1 - sigma * sqrt(T);
        fprintf('1 : call option \n 2 : put option \n ');
        ch = input(' choice for put or call ');
        if ch == 1 % call option
        c0 = s0 * normcdf(d1) - k * exp(-r*T) * normcdf(d2);
        fprintf('\n Value of call option = %f\n',c0);
        end
```

```
        if ch == 2 % put option
        c0 = s0 * normcdf(d1)- k * exp(-r*T) * normcdf(d2);
        p0 = c0 - s0 + k * exp(-r*T);
        fprintf('\n Value of put option  = %4.4f\n',p0);
        end
        if ch ~= 1 && ch ~= 2
            disp('Invalid choice');
        end
    end
    CH = input('1-continuing option pricing else any other key: ');
end
```

## 15.5 Minimum Variance Portfolio

```
% ri is the column vector of return over the given periods.
% It can be generated either by historical data from the market
% or by simulation.
% Here it is done by simulation i.e.,
% ri is randomly generated from standard normal distribution.

% R = [r1 r2 ... rn]   returns matrix.

% We take n = 10, the number of assets.
% We take 30 periods of return.

R = randn(30,10);   %  returns of assets over period of time.
C = cov(R);         %  Covariance matrix of returns.
e = ones(10,1);

pC = pinv(C);       % Finding inverse of the covariance matrix.
w = pC*e/(e'*pC*e); % Weights of assets in portfolio.

disp('Minimum variance portfolio is ');
disp(w);
```

## 15.6 Markowitz Efficient Frontier

```
%  ri is the column vector of return over the given periods,
%  which can be generated either by historical data from the
%  market or by simulation.
%  Here it is taken by simulation randomly generated
%  from standard normal distribution.
%  R = [r1 r2 ... rn]   returns matrix;
%  Here we have taken n = 10, number of assets,
%  and there are 30 periods of return.

% Start the code.

% ri = input('Enter the return vector of the asset i ');

R = randn(30,10);   % Returns of assets over period of time;
C = cov(R);         % Covariance matrix of the returns.
e = ones(10,1);

pC = pinv(C);       % Finding inverse of the covariance matrix.

wmin = pC*e/(e'*pC*e);

disp('Minimum variance portfolio is ');
disp(wmin);

disp('Expected return of each assets over the period ');
meanR = mean(R);
disp(meanR);

mu1 = meanR * wmin;
disp('Expected return at minimum variance portfolio ');
disp(mu1);
fprintf('Enter the expected return  greater than %f',mu1);
mu2 = input(' mu2 = ');

% We calculate the portfolio for the given mu2.
```

```
A = [mu2  meanR * pC * e ; 1  e' * pC * e];
B = [meanR * pC * meanR' mu2; e' * pC * meanR'  1];

D = [meanR * pC * meanR' meanR * pC * e ; e'*pC * meanR' e' * pC * e];

w2 = (det(A)*(pC * meanR') + det(B) * (pC * e))/(det(D));

disp('Another portfolio ');
disp(w2);

% Using the two fund theorem, we find the efficient frontier.

port = zeros(10,1);
effr = zeros(10,1);
effsigma = zeros(10,1);
for i=1:11
    lambda =(i-1)/10 ;
    port = (1-lambda)* wmin + (lambda*w2);
    effr(i,1)= meanR * port;
    effsigma(i,1) = port'* C * port;
end plot(effsigma, effr);
xlabel('Sigma');
ylabel('Return');
```



## 15.7 Marowitz Efficient Frontier: User Specified Data

```
% The program calls in-built MATLAB function 'quadprog'
% for solving a Markowitz quadratic program model.

% Symbols have their usual meanings.
% Start the code.

% Set-up
if ~exist('quadprog')
   msgbox('The optimization toolbox is required to run
            this demo.', 'Product dependency')
```

```
    return
end


n = input('Enter the number of assets = ');
str = ['Enter the expected returns of 'num2str(n)'
        Companies in 1 x 'num2str(n)' matrix form = '];
returns = input(str);

str1 = ['Enter the standard deviation in 1 x 'num2str(n)'
        matrix form = '];
STDS = input(str1);

str2 = ['Enter the correlation matrix in 'num2str(n)' x 'num2str(n)'
        matrix form = '];
correlations = input(str2);


% Convert to variance-covariance matrix.
% Convert the standard deviation and correlation.

covariances = corr2cov(STDS, correlations);

% Compute and plot efficient frontier for 20 portfolios.
% The MATLAB financial toolbox in built function ‘portopt’
% output the mean-variance efficient frontier with user-specified
% covariance and returns.


portopt(returns, covariances, 20)

% Randomize assets weights.
% Randomly generate the asset weights for 1000 portfolios,
% a uniformly distributed 1000 x n matrix of
% pseudo-random numbers in (0,1).

rand('state', 0)
weights = rand(1000, n);
```

```
% Normalized the weights so that their sum is 1.

total = sum(weights, 2);       % sum of square of weights.
total = total(:, ones(n,1));
weights = weights. / total;

% Compute expected returns and risks for each portfolio,
% using a MATLAB financial toolbox in-built
% function 'portstats'

[portRisk,portReturn]=portstats(returns,covariances,weights);

% Now plot the efficient frontier.

hold on
plot(portRisk, portReturn, '.r')
title('Mean-Variance efficient frontier for random portfolios')
hold off
```

## 15.8 CAPM with Generic Input Parameters



```
% First we generate returns of 20 companies,
% for last one year (12 months) using simulation
% Also generate return for market portfolio for last one year.
% But one can easily generalize the program.

% Start the code.

% MR = market return,
% it can taken from historical data of any market.
% Data collections/genration


MR = randn(52,1);  % Random weekly returns for market;
R  = randn(52,20); % weekly returns of assets a1...a20 for one year;
MMR = mean(MR);
```

```
% C(i) = Covariance(MR,R(i)) of the i-th asset.

C = zeros(20,1);

for i = 1:20;
    MERR = mean(MR. * R(:,i));
    C(i,1) = MERR - MMR * mean(R(:,i));
end

VMR = var(MR);

% Calculating the beta coefficient.

% beta(i) = beta coefficient of i-th asset.
% beta(i) = C(i)/var(MR);

beta = zeros(20,1);
for i = 1:20
    beta(i) = C(i,1)/VMR;
end

% Calculating the return  of each asset form CAPM
% r_i = risk free return + beta_i *(E(M)- risk free return);
% rf = input('Enter the risk free interest rate per week ');

rf = 0.005;         % Assumed per week risk-free-return.
r = zeros(size(beta));

for i = 1:length(beta)
    r(i) = rf + beta(i)*(MMR-rf);
end

fprintf('\tAseets \t  beta \t\t asset_return_pm \n');

for i =1 :20
    fprintf('\t  a%d \t %f \t\t  %f\n',i, beta(i),r(i));
```

```
end
```

## 15.9 Geometric Brownian Motion using Simple Random Walk

```
% The Geometric_brownian(N,r,sigma,T)
% simulates a Geometric Brownian motion on [0,T].

% Recall: If X is a Brownian motion then exp(X) is a
%  geometric Brownian motion.

function [Y] = geometric_brownian(N,r,sigma,T)

% N = input('number of step =');
N = 1000;
T = input('Enter length of interval =');
r = input('Enter interest rate = ');
sigma = input('Enter diffusion coefficient = ');

t = (0:1:N)'/N;   % t is the column vector [0 1/N 2/N ... 1]

W = [0; cumsum(randn(N,1))]/sqrt(N);
% cumsum function is running sum of normal N(0,1/N) variables.

t = t * T;
W = W * sqrt(T);
X = (r-(sigma^2)/2)* t + sigma * W;
Y = exp(X);

plot(t,Y);
title(['Sample Path of Geometric Brownian motion with
diffusion coefficient='num2str(sigma),'Interest rate='num2str(r)])
xlabel(['Time'])
```

## 15.10 Binomial Lattice for Bond Price and Interest Rate Modeling

```
% Binomial lattice model for bond price & Interest rate modeling.
% Symbols have their usual meanings.

% Start the code.

% Generic input parameter values.

  r00 = input('r00 = ');
  u   = input('u = ');
  d   = input('d = ');
  T   = input('T = ');
  k   = input('k = ');    % Strike price

% For example, r00 = 0.07; u = 1.3; d = 0.9; T = 5; k = 90;

%  Finding binomial lattice for interest rate modeling

IntLattice = zeros(T,T);

r=0;              % power of d in the Binomial lattice
for i = T:-1:1    % row generation
    r = 1;
    % i+j = T+1 =>j= T+1-i
    % it will make a lower triangle matrix

    for j = T+1-i:T     % column generation
        IntLattice(i,j) = r00*u^(T-i)*d^(r-1);
        r = r+1;
    end
end disp('Short interest rate dynamics')
disp(IntLattice);

% Finding bond price which is the underlying
% asset for call/put options.
```

```
p = zeros(T+1,T+1);
x = 1;       % We want find in percentage later will be converted
p(:,T+1) = x;
for j = T:-1:1          % j = column traversal
    for i = T+1:-1:2    %i = row traversal
    if i+j >=T+2    % To make a lower triangle
    p(i,j) =(1.0/(1 + IntLattice(i-1,j)))*
    (0.5*p(i-1,j+1)+0.5*p(i,j+1));
    end
    end
end

disp('Underlying assets bond price ') disp(p)

% Calculation of Call option where underlying assets is bond price.

 C = zeros(T,T);
 for i = T+1:-1:2
     C(i-1,T) = max(p(i,T)*100-k,0);
 end
 for j = T-1:-1:1       % j = column traversal
 for i = T:-1:2      % i = row traversal
 if i+j >=T+1    % To make a lower triangle
 C(i,j) =(1.0/(1 + IntLattice(i,j)))*
 (0.5*C(i,j+1)+0.5*C(i-1,j+1));
 end
 end
end

disp('Binomial lattice for call price ');
disp(C);

% Calculation of Put option where underlying assets is bound price.

 P = zeros(T,T);
 for i = T+1:-1:2
     P(i-1,T) = max(k - p(i,T)*100, 0);
 end
```

```
for j = T-1:-1:1       % j = column traversal
for i = T:-1:2      % i = row traversal
if i+j >=T+1    % construct a lower triangle
P(i,j) =(1.0/(1 + IntLattice(i,j)))*
(0.5*P(i-1,j+1)+0.5*P(i,j+1));
end
end
end
disp('Binomial lattice for put option ');
disp(P);
```

## 15.11 European/American Call Options via Monte Carlo Method

```
% S(i) Stock price is consider as Geometric Brownian motion.
% Simulate the solution of Black-Scholes model
% stochastic differential equation:
% ds(t) = mu * s(t)dt + sigma * s(t) * dw(t).
%  mu = r = expected return =  drift rate.

% Symbols have their usual meanings.

% Start the Code

% Generic input parameter values.
  S0  = input('S0 = ');
  T   = input('T = ');%Expiration time
  k   = input('k = ');
  r   = input('r = ');
  sigma = input('Sigma = ')     %  volatility of stock price.
  N  =  input('No of simulations');

% For example, we can take, S0 = 100; T = 4; k = 110;
% r = 0.05; sigma = 0.24; N = 100;
E = zeros(N,1);
S = zeros(N,1);
V = zeros(N,1);
```

```
fprintf('\n 1 : for European options \n 2 : for American options \n ');
choice_A_E = input('Your choices '); switch choice_A_E
case 1
fprintf('\n 1 : for call option \n 2 : for put option ');
   choice_p_c = input('your choice ');
     if choice_p_c == 1
     for i=1:N
     E(i) = normrnd(0,1);
     S(i) = S0*exp((r-(1/2)*sigma^2)*T + sigma*sqrt(T)*E(i));
     V(i) = (exp(-r*T))* max(S(i) - k,0);
     end
     disp('call price ');
     c0 =(1/N) * sum(V);
     disp(c0)
    else
    for i=1:N
    E(i)= normrnd(0,1);
    S(i)= S0*exp((r-(1/2)*sigma^2)*T + sigma*sqrt(T)*E(i));
    V(i)=(exp(-r*T))* max(k - S(i),0);
    end
    p0 =(1/N) * sum(V);
    disp('call price ');
    disp(p0)
    end
case 2
fprintf('\n 1 : for call option \n 2 : for put option ');
  choice_p_c = input('your choice ');
   if choice_p_c == 1
   for i = 1:N
   E(i) = rand(0,1);
   S(i) = S0*exp((r-(1/2)* sigma^2)* T + sigma*sqrt(T)*E(i));
   V(i) = (exp(-r*T))* max(S(i)- k,0);
   end
   c0 =(1/N) * sum(V)
   else
   for i=1:N
   E(i) = normrnd(0,1);
   S(i) = S0*exp((r-(1/2)* sigma^2)* T + sigma*sqrt(T)*E(i));
```

```
    V(i) = (exp(-r*T))* max(k - S(i),0);
    end
    p0 =(1/N)*sum(V)
   end
end
```

# References


1. F. Aitsahila, Y. -C. Sheu and P. M. Pardalos, Optimal execution of time-constrained portfolio transaction,*in Computation Methods in Financial Engineering*, Part 1 (J. Erricos, J. Kontoghiorghes, B. Rustem, P. Winker, editors), pp. 95–102, 2008.
2. R. Almgren and N. Chriss, Optimal execution of portfolio transactions, *Journal of Risk*, Vol. 3(2), pp. 5–39, 2000.
3. F. Anderson, H. Mausser, D. Rosen, and S. Uryasev, Credit risk optimization with conditional value at risk criterion. *Mathematical Programming*, Vol. 89, pp. 273–291, 2000.
4. D. Applebaum, *Levy Processes and Stochastic Calculus*, Cambridge University Press, $2^{nd}$ edition, 2009.
5. P. Artzner, F. Delbaen, J. M. Eber and D. Heath, Coherent measures of risk, *Mathematical Finance*, Vol. 9(3), pp. 203–228, 1999.
6. R. B. Avery, R. W. Rostic, P. S. Calem and G. B. Canner, Credit scoring: statistical issues and evidence from credit-bureau files, *Real Estate Economics*, Vol. 28(3), pp. 523–547, 2000.
7. M. Bartholomew-Biggs, *Nonlinear Optimization with Financial Applications*, Kluwer Academic Publishers, 2005.
8. M. Baxter and A. Rennie, *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge University Press, 1996.
9. A. Ben-Tal, T. Margalit and A. N. Nemirovski, Robust modeling of multistage portfolio problems, *in High Performance Optimization* (H. Frenk, K. Roos, T. Terlsky, and S. Zheng editors), Kluwer Academic Publishers, pp. 303–328, 2002. http://www2.isye.gatech.edu/~nemirovs/rob_portf.pdf.
10. M. J. A. Berry and G. S. Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Sons, 2000.
11. D. Bertsimas and A. Lo, Optimal control of execution costs, *Journal of Financial Markets*, Vol. 1, pp. 1–50, 1998.
12. D. Bertsimas, P. Hummel and A. Lo, Optimal control of execution costs for portfolios, *Computing in Science and Engineering*, Vol. 1(6), pp. 40–53, 1999.

13. F. Black, E. Derman and W. Toy, A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts*, Vol. 22, pp.33–39, 1990.
14. F. Black and P. Karasinski, Bond and option pricing when short rates are lognormal, *Financial Analysts*, Vol. 47, pp. 52–59, 1991.
15. F. Black and M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy*, Vol. 81, pp. 637–659, 1973.
16. M. E. Blume and D. B. Kein, The valuation of callable bonds, Wharton Financial Institution Center,
http://finance.wharton.upenn.edu/~rlwctr/papers/8914.PDF
17. E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet and T. Roncalli, Copulas for finance: A reading guide and some applications, 2000.
http://thierry-roncalli.com/download/copula-survey.pdf
18. E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet and T. Roncalli, Copulas: an open field for risk management, 2003.
http://www.thierry-roncalli.com/download/copula-rm.pdf
19. P. Boyle, Options: a Monte Carlo approach, *Journal of Financial Economics*, Vol. 4, pp. 323–338, 1977.
20. P. P. Boyle and T. Vorst, Option replication in discrete time with transaction costs, *Journal of Finance*, Vol. 47, pp. 271-293, 1992.
21. D. Brigo and F. Mercurio, *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*, Springer Finance, $2^{nd}$ edition, 2006.
22. N. Bucay and D. Rosen, Credit risk of an international bond portfolio: a case study, *ALGO Research Quarterly*, Vol. 2(1), pp. 9–29, 1999.
23. X. Cai, K. -L. Teo, X. Yang and X. Y. Zhou, Portfolio optimization under a minimax rule, *Management Science*, Vol. 46, pp. 957–972, 2000.
24. A. J. G. Cairns, *Interest Rate Models*, Princeton University Press, 2004.
25. M. Capinski and T. Zastawniak, *Mathematics for Finance: An Introduction to Financial Engineering*, $2^{nd}$ edition, Springer 2010.
26. M. M. Chawla, On the bionomial tree method for the valuation of options, *International Journal of Applied Mathematics*, Vol. 20(2), pp. 235–251, 2007.
27. M. M. Chawla, On mean variance portfolio optimization, *International Journal of Applied Mathematics*, Vol. 21(3), pp. 473–494, 2008.
28. Y.-W. Chen and C.-J. Lin, Combining SVMs with various feature selection strategies, 2005. Available from http://www.csie.ntu.edu.tw/ cjlin/ papers/features.pdf.
29. U. Cherubini, E. Luciano and W. Vecchiato, *Copula Methods in Finance*, John Wiley & Sons, 2004.

http://www.elan.com.mx/biblioteca/FINANCE%20MATERIAL%20CM/Wiley%20Copula%20Methods%20in%20Finance.pdf.

30. V. K. Chopra and W. T. Ziembia, *The effect of errors in means, variances, and co-variances on optimal portfolio choices, Worldwide Asset and Liability Modeling* (W. T. Ziembia and J. M. Mulvey editors), Cambridge University Press, 1998.
31. N. A. Chriss, *Black-Scholes and Beyond Option Pricing Models*, McGraw Hill, 1997.
32. G. Cornuejols and R. Tütüncü, *Optimization Methods in Finance*, Cambridge University Press, 2007.
33. J. C. Cox, S. A. Ross and M. Rubinstein, Option pricing: a simplified approach, *Journal of Financial Economics*, Vol. 7(3), pp. 229-263, 1979.
34. J. C. Cox, J. E. Ingersoll and S. A. Ross, A theory of the term structure of interest rates, *Econometrica*, Vol. 53, pp. 385–407, 1985.
35. R. Deaves and M. Parlar, A generalized bootstrap method to determine the yield curve, Macmaster University, 1999. http://www.stats.uwo.ca/events/Colloquium/W09/pdf/Rodrigo.pdf
36. X. -T. Deng, Z. -F. Li and S. -Y. Wang, A minimax portfolio selection strategy with equilibrium, *European Journal of Operational Research*, Vol. 166, pp. 278–292, 2005.
37. D. Dentcheva and A. Ruszczynski, Portfolio optimization with stochastic dominance constraints, *Journal of Banking and Finance*, Vol. 30, pp. 438–451, 2006.
38. E. Derman, I. Kani, and N. A. Chriss, Goldman Sachs–quantitative strategies research notes, implied trinomial trees of the volatility smile, 1996. http:// www.ederman.com/new/docs/gs-implied-trinomial-trees.pdf.
39. D. Duffie and J. Pan. An overview of Value-at-Risk, *Journal of Derivatives*, Vol. 4, pp. 7–49, 1997.
40. P. Eizek and K. Komorad, Implied trinomial trees, SFB 649 Discussion Paper 2005-007, http://edoc.hu-berlin.de/series/sfb-649-papers/2005-7/PDF/7.pdf.
41. E. J. Elton, M. J. Gruber, S. J. Brown and W. Goetzmann, *Modern Portfolio Theory Investment Analysis*, $8^{th}$ edition, John Wiley & Sons, 2009.
42. P. Embrechts, A.J. McNeil and D. Straumann, Correlation and dependency in risk management: properties and pitfalls, *Risk Management: Value at Risk and Beyond*( M. Dempster and H.K. Moffatt editors), Cambridge University Press, 2000.
43. T. W. Epps, *Pricing Derivative Securities*, World Scientific Publishing, 2000.

44. C.I. Fábián, G. Mitra and D. Roman, Processing second order stochastic dominance models using cutting plane representations, *Mathathematical Programming* (Ser A), Vol. 130(1), pp. 33–57, 2011.
45. A. Fensterstock, Credit scoring and the next step, *Business Credit*, Vol. 107(3), pp. 46–49, 2005.
46. D. Filipović, Interest Rate Models, http://www.mathematik.uni-muenchen.de/~filipo/ZINSMODELLE/zinsmodelle1.pdf
47. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons, 1970. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0708563.
48. M. Garman and S. Kohlhagen, Foreign currency option values, *Journal of International Money and Finance*, Vol. 2, pp. 231–237, 1983.
49. P. Glasserman, *Monte Carlo methods in Financial Engineering*, Springer, 2003.
50. P. Glasserman, P. Heidelberger and P. Shahabuddin, Variance reduction techniques for estimating Value-at-Risk, Management Science, Vol. 46(10), pp. 1349–1364, 2000.
51. C. Goffman, *Introduction to Real Analysis*, A Harper International Edition, 1966.
52. A. J. Goldman and A. W. Tucker, *Linear Equalities and Related Systems*, Priceton University Press, pp. 53–97, 1956.
53. M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering*, Birkhäuser, 2002.
54. L. Güntay, N.R. Prabhala and H. Unal, Callable bonds and hedging, Wharton Financial Institution Center, http://fic.wharton.upenn.edu/fic/papers/02/0213.pdf
55. N. Gupta and A. Chaudhary, Optimal trading strategies, M.Tech. Dissertation, Department of Mathematics, IIT Delhi, 2009.
56. J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numerical Mathematics*, Vol. 2, 84–90, 1960.
57. H. Han and X. Wu, A fast numerical method for the Black-Scholes equation of American options *SIAM Journal on Numerical Analysis*, Vol. 41, pp. 2081–2095, 2003.
58. D. J. Hand, Modelling consumer credit risk, *IMA Journal of Management Mathematics*, Vol. 12(1), pp. 139–155, 2001.
59. L. Haris, *Trading and Exchanges Market Microstructure for Practitioners*, Oxford University Press, 2003.

60. M. Haugh, Continuous-time short rate models, Term structure models: IEOR E4710, 2010. http://www.columbia.edu/~mh2078/cts_shortrate_models.pdf
61. D. Heath, R. Jarrow and A. Morton, Bond pricing and the term structure of interest rates: a new method for contingent claim valuation, *Econometrica*, Vol. 60, pp. 77–105, 1992.
62. S. Heston, A closed form solution for options with stochastic volatality and applications to bond and currency options, *Review of Financial Studies*, Vol. 6, pp. 327–343, 1993.
63. C. W. Hsu, C. C. Chang and C. J. Lin, A practical guide to support vector classification, 2003.
http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.
64. G. Huberman and W. Stanzl, Optimal liquidity trading, *Review of Finance*, Vol. 9, pp. 165-200, 2005.
65. J. C. Hull, *Options, Futures and Other Derivatives*, $8^{th}$ edition, Prentice Hall, 2012.
66. J. Hull and A. White, Pricing interest rate derivative securities, *Review of Financial Studies*, Vol. 3, pp. 573–592, 1990.
67. P. J. Hunt and J. E. Kennedy, *Financial Derivatives in Theory and Practice*, John Wiley & Sons, 2004.
68. R. Jarrow and A. Rudd, *Option Pricing*, Irwin Professional Publishers, 1983.
69. P. Jăckel, *Monte Carlo Methods in Finance*, John Wiley & Sons, 2002.
70. N. Ju and Zhong, An approximate formula for pricing American options.
http://www.rhsmith.umd.edu/faculty/nju/JZ99JD.pdf
71. T. Joro and P. Na, Portfolio perference evaluation in mean-variance-skewness framework, *European Journal of Operation Research*, Vol. 175, pp. 516–542, 2006.
72. O. Josef, Models of interest rate evolution-Vasicek and CIR models, *Journal of Applied Mathematics*, Vol. 2(3), pp. 117–125, 2009.
73. L. Kaelbling, M. Littman and A. Moore, Reinforcement learning: a survey, *Journal of Artificial Intelligence Research*, Vol. 4, pp. 237–285, 1996.
74. P. Kall, *Stochastic Linear Programming*, Springer, 1976.
75. I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, $2^{nd}$ edition, Springer, 1991.
76. R. Khemchandani, N. Gupta, A. Chaudhary and S. Chandra, Optimal execution with weighted impact functions: a quadratic programming approach, *Optimization Letters*, 2012. DOI 10.1007/s11590-012-0441-4
77. R. Kissell and M. Glantz, Optimal trading strategies: quantitative approaches for managing market impact and trading risk, AMACON, 2003.

78. F. C. Klebaner, *Introduction to Stochastic Calculus with Applications*, $3^{rd}$ edition, Imperial College Press, 2011.
79. H. Konno and K. Suzuki, A mean-variance-skewness optimization model, *Journal of Operation Research Society of Japan*, Vol. 38, pp. 135–187, 1995.
80. H. Konno and H. Yamazaki, Mean- absolute deviation portfolio optimization model and its application, *Management Science*, Vol. 37, pp. 519–531, 1991.
81. H. Leland, Option pricing and replication with transactions costs, *Journal of Finance*, Vol. 40, pp. 1283–1301, 1985.
82. D. Leisen and M. Reimer, Binomial models for option valuation - examining and improving convergence, Discussion Paper no. B 309, 1995. http://www.smartquant.com/references/OptionPricing/option31.pdf
83. H. Levy, Stochastic Dominance and expected utility: survey and analysis, *Management Science*, Vol. 38, pp. 555–593, 1992.
84. P. Lucas, Score updates, *Collections and Credit Risk*, Vol. 7(10), pp. 22–25, 2002.
85. D. G. Luenberger, *Investment Science*, Oxford University Press, 1998.
86. J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2001.
87. R. L. McDonald, *Derivatives Markets*, $2^{nd}$ edition, Addison Wesley Pearson, 2006.
88. D. Martin, Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance*, Vol. 1, pp. 249-276, 1977.
89. R. Mansini, W. Ogryczak and S. Uryasev, Conditional value at risk and related linear programming model for portfolio optimization, *Annals Operations Research*, Vol. 152, pp. 227–256, 2007.
90. H. M. Markowitz, Portfolio selection, *Journal of Finance*, Vol. 7, pp. 77–91, 1952.
91. H. Mausser and D. Rosen, Beyond VaR: from measuring risk to managing risk, *ALGO Research Quarterly*, Vol. 1(2), pp. 5–20, 1998.
92. A. Mele, Lecture Notes in Financial Economics, 2011, http://www.antoniomele.org/files/fin_eco.pdf.
93. R. Merton, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics*, Vol. 3, pp. 125–144, 1976.
94. R. C. Merton, Theory of rational option pricing, *Bell Journal of Economics and Management Science*, Vol. 4, pp. 141–183, 1973.
95. L. J. Mester, What's the point of credit scoring?, 1997. http://www.phil.frb.org/research-and-data/publications/business-review/1997/september-october/brso97lm.pdf.

96. C. Michel, D. Galai and R. Mark, A comparative analysis of current credit risk models, *Journal of Banking and Finance*, Vol. 24, pp. 59–117, 2000.
97. T. Mikosch, *Elementary Stochastic Calculus (with Finance in View)*, World Scientific Publication, 1998.
98. O. L. Mangasarian. *Nonlinear Programming*, SIAM, Philadelphia, 1994.
99. J. P. Morgan, *CreditMetrics*, Technical Document, 1997. http://www.ma.hw.ac.uk/~mcneil/F79CR/CMTD1.pdf
100. J. Morrison, Introduction to survival analysis in business, *The Journal of Business Forecasting Methods and Systems.* Vol. 23(1), pp. 18–22, 2004.
101. Y. Nevmyvaka, Y. Feng and M. Kearns, *Reinforcement learning for optimized trade execution*, Proceedings of International Conference on Machine Learning, 2006.
102. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, 1998.
103. W. Ogryczak and A. Ruszczynski, From stochastic dominance to mean-risk models: semideviations as risk measures, *European Journal of Operation Research*, Vol. 116, pp. 33–50, 1999.
104. W. Ogryczak and A. Ruszczynski, On consistency of stochastic dominance and mean-semideviations models, *Mathematical Programming*, Vol. 89, pp. 217–232, 2001.
105. J. Ohlson, Spring Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, Vol. 18, pp. 109-131, 1980.
106. S. Park, Solving the mystery of credit scoring models, *Business Credit*, Vol. 106(3), pp. 43–47, 2004.
107. A. F. Perold, The implementation shortfall: paper versus reality, *Journal of Portfolio Management*, Vol. 14, pp. 4–9, 1988.
108. S. Perrakisa and J. Lefoll, Option pricing and replication with transaction costs and dividends, http://www.smartquant.com/references/OptionPricing/option2.pdf
109. M. Pritsker, Evaluating value at risk methodologies, *Journal of Financial Services Research*, Vol. 12(2/3), pp. 201–242, 1997.
110. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, $2^{nd}$ edition, Springer, 2004.
111. R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *The Journal of Risk*, Vol. 2, pp. 21–41, 2000.
112. R. T. Rockafellar, Coherent approaches to risk in optimization under uncertainty: tutorials in operational research, *INFORMS*,

pp. 38–61, 2007. http://www.math.washington.edu/~rtr/papers/rtr-RiskTutorial_INFORMS2007.pdf
113. S. Roman, *Introduction to the Mathematics of Finance: From Risk Management to Options Pricing*, Springer, 2004.
114. S. M. Ross, *An Introduction to Mathematical Finance: Options and Other Topics*, Cambridge University Press, 1999.
115. M. Rubinstein, Implied binomial tree, *Journal of Finance*, Vol. 49(3), pp. 771–818, 1994.
http://www.haas.berkeley.edu/groups/finance/WP/rpf232.pdf
116. A. Ruszczy'nski and R.J. Vanderbei, Frontiers of stoschastically nondominated portfolios, *Econometrika*, Vol. 71, pp. 1287–1297, 2003.
117. P. A. Samuelson, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review*, Vol. 6, pp. 41 -50, 1965.
118. A. L. Sandler S. E. McGinn and J. L. F. Barloon, Lending scrutiny of credit score-based underwriting systems, ABA Bank Compliance (March/April) Vol. 38, 2000.
119. W. F. Sharpe, Capital asset prices - a theory of market equilibrium under conditions of risk. *Journal of Finance* Vol. 13(3), pp. 425-442, 1964.
120. W. F. Sharpe, G. J. Alexander and J. V. Bailey, *Investments*, Prentice Hall International, 1998.
121. S. E. Shreve, *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer, 2004.
122. S. E. Shreve, *Stochastic Calculus for Finance II: Continuous Time Models*, Springer, 2004.
123. I. M. Sobol, On the distribution of points in a cube and the approx imate evaluation of integrals, *USSR Computational Mathematics and Mathematical Physics*, Vol. 7, pp. 86–112, 1967.
124. M. S. Sodhi, LP modelling for asset liability management: a survey of choices and simplifications, *Operations Research*, Vol. 53, pp. 181–196, 2005.
125. H. M. Soner, S. E. Shreve and J. Cvitanic, There is no nontrivial portfolio for option pricing with transaction costs, *The Annals of Applied Probability*, Vol. 5, pp. 327-355, 1995.
126. J. B. Sousa, M. L. Esquivelb and R. M. Gaspar, Machine learning Vasicek model calibration with Gaussian processes, http://www.dm.fct.unl.pt/sites/www.dm.fct.unl.pt/files/preprints/2010/4_10.pdf
127. P. Sӧderlind, Lecture Notes in Finance 2 (MiQE/F, MSc course at UNISG),

2011. http://home.datacomm.ch/paulsoderlind/Courses/OldCourses/Fin2MiQEFAll.pdf

128. S. Sriboonchitta, W. K. Wong, S. Dhompongsa and H. T. Nguyen, *Stochastic Dominance and Applications to Finance, Risk and Economics*, CRC Press, 2009.
129. M. C. Steinbach, Markowitz revisted: mean variance models in financial portfolio analysis *SIAM Review*, Vol. 43(1), pp. 31–85, 2001.
130. R. E. Steuer, Y. Qi and M. Hirschberger, Portfolio optimization: new capabilities and future methods, *Zeitschrift fr Betriebswirtschaft*, Vol. 76, pp. 199–219, 2006.
131. W. Sun, A. C. Fan, L. -W. Chen, T. Schouwenaars and M. A. Albota, Optimal Rebalancing Strategy Using Dynamic Programming for Institutional Portfolios, Available at SSRN 2004.
132. R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, 1998.
133. California Debt and Investment Advisory Commission, Understanding interest rate swap math pricing, 2007. http://www.treasurer.ca.gov/cdiac/publications/math.pdf
134. L. C. Thomas, A survey of credit and behavioral scoring forecasting financial risk of lending to consumers, *International Journal of Forecasting*, Vol. 16, pp. 163–167, 2000.
135. L. C Thomas, D. Edelman and J. N. Crook, *Credit Scoring and its Application*, SIAM Monograph on Mathematical Modelling and its Computation, 2002.
136. Y. Tian, A modified lattice approach to option pricing, *Journal of Futures Markets*, Vol. 13(5), pp. 563–577, 1993.
137. H. Timmons, The cracks in credit scoring: as loan defaults rise, worries grow about how creditors pick borrowers, Business Week (November) 25, 2002.
138. R. S. Tsay, *Analysis of Financial Time Series*, John Wiley & Sons, 2005.
139. R. H. T*ü*t*ü*nc*ü* and M. Koenig, Robust asset allocation, *Annals of Operation Research*, Vol. 132, pp. 157–87, 2004.
140. S. Uryasev and R.T. Rockafellar, Optimization of conditional Value-at-Risk, Research Report 99-4, ISE Dept., University of Florida, 1999.
141. V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
142. J. R. Varma, Bond valuation and the pricing of interest rate options in India, *The ICFAI Journal of Applied Finance*, Vol. 2(2), pp. 161–176, 1996.
143. O. A. Vasicek, An equilibrium characterization of the term structure, *Journal of Financial Economics*, Vol. 15, pp. 177–188, 1977.

144. P. -C. G. Vassiliou, *Applied Stochastic Finance*, Vol. 2, John Wiley & Sons, 2012.
145. F. de Weert, *Exotic Options Trading*, John Wiley & Sons, 2008.
146. G. West, Interest Rate Derivatives: Lecture Notes, 2010. http://www.finmod.co.za/ird.pdf.
147. P. Wilmott, *Introduces Quantitative Finance*, John Wiley & Sons, 2007.
148. L. Y. Yu, X. D. Ji and S. Y. Wang, Stochastic programming models in financial optimization: a survey, *Advance Modeling and Optimization*, Vol. 5, pp. 1-26, 2003.
149. R. Zagst, *Interest-Rate Management*, Springer, 2002.
150. X. Y. Zhou and D. Li, Continuous time mean- variance selection: a stochastic LQ framework, *Applied Mathematics and Optimization*, Vol. 42, pp. 19–33, 2000.

# Index