# Introduction to Mathematical Finance

*Discrete Time Models*

Stanley R. Pliska

First published 1997
Reprinted 1997, 1998, 1999, 2000 (twice), 2001

Blackwell Publishers Inc
350 Main Street, Malden, Massachusetts 02148, USA

Blackwell Publishers Ltd
108 Cowley Road, Oxford OX4 1JF, UK



*Library of Congress Cataloging in Publication Data* has been applied for

*British Library Cataloguing in Publication Data*
A CIP catalogue record for this book is available from the British Library

ISBN 1–55786–945–6

*Commissioning Editor:* Rolf Janke
*Desk Editor:* Linda Auld
*Production Manager/Controller:* Lisa Parker
*Text Designer* Lisa Parker

Typeset in Times on 10/11.5 pt.
by Pure Tech India Ltd., Pondicherry, India
Printed and bound in Great Britain
by MPG Books Ltd, Bodmin, Cornwall

This book is printed on acid-free paper.

# Contents

# Preface

## Aims and Audience

This book, which has grown out of research conducted by the author with J. Michael Harrison in 1981, is designed to serve as a textbook for advanced undergraduate and beginning graduate students who seek a rigorous yet accessible introduction to the modern financial theory of security markets. This is a subject that is taught in both business schools and mathematical science departments, and it is also a subject that is widely and extensively utilized in the financial industry. The derivatives industry has roughly $20 trillion in notional principal outstanding as this book goes to press, and the portfolio management industry is probably even bigger. Mathematics play crucial roles in both these areas. Consequently, financial practitioners (especially 'rocket scientists,' quants, financial engineers, etc.) may find this book useful for their theoretical background.

The full theory of security markets requires knowledge of continuous time stochastic process models, measure theory, mathematical economics, and similar prerequisites which are generally not learned before the advanced graduate level. Hence a proper study of the complete theory of security markets requires several years of graduate study (or equivalent, sink or swim, experience). However, by restricting attention to discrete time models of security prices it is possible to acquire an introduction without making a big investment in the advanced mathematics. In fact, while living in a discrete time world it is possible to learn virtually all of the important financial concepts. The purpose of this book is to provide such an introductory study.

There is still a lot of mathematics in this book. The reader should be comfortable with calculus, linear algebra, and probability theory that is based on calculus (but not necessarily measure theory). Random variables and expected values will be playing important roles. The book will develop important notions concerning discrete time stochastic processes; prior knowledge here will be useful but is not required. Presumably the reader will be interested in finance and thus will come with some rudimentary knowledge of stocks, bonds, options, and financial decision making. The last topic involves utility theory, of course; hopefully the reader will be

familiar with this and related topics of introductory microeconomic theory. Some exposure to linear programming would be advantageous, but those lacking this knowledge can make do with the appendix and independent study.

The aim of this book is to provide a rigorous treatment of the financial theory while maintaining a casual style. There is an emphasis on computational examples, and exercises are provided to check understanding and provide supplemental information. Readers seeking institutional knowledge about securities, derivatives, and portfolio management should look elsewhere, but those seeking a careful introduction to financial engineering will find that this is a useful and comprehensive introduction to the subject.

## Brief Summary of This Book

This book consists of seven chapters, each divided into a number of sections. Important equations, fundamental statements, examples, and exercises are labeled with numbers by chapter. For example, equation 2.1 is the first equation in chapter 2.

This summary will point out which subjects are most important, and why (usually because I think something is of fundamental importance rather than a narrow result of limited or temporary consequence). It will also indicate topics that are new, at least in their treatment. Arguably, there are no new results in this book, but, like Monday morning quarterbacking, we can look backwards and see better ways to say and do things. Hopefully, the book will successfully do this, thereby conveying a clear understanding of some fundamental ideas about security markets.

The first two chapters are devoted to single period models. Most of the important concepts in this book are introduced here, making sections 1.1.–1.5 and 2.1–2.3 especially important. Section 2.4 is a modern treatment of the important mean-variance portfolio analysis. Sections 1.6 and 2.5–2.7 are extensions and ventures into significant topics that are a bit out of this book's mainstream.

The rest of this book is devoted to multiperiod models. This builds on the single period results, emphasizing what is new and different. The redundant material is kept concise in order to spare the patience of the stronger reader (but such readers will still find it worthwhile to refer to the first two chapters). Chapter 3 describes the basic elements of securities market models and introduces important notions such as dividend processes and the binomial model. Chapter 4 is devoted to derivatives, including forwards and futures; all the sections here are of fundamental interest. Chapter 5 attends to optimal consumption and investment problems. Sections 5.2 and 5.4 are the most important ones here (of course I might be biased, for the ideas originated from my research in 1982 and 1986), because they deal with the risk neutral computational approach. Sections 5.5–5.8 are extensions and special cases.



Interest rate derivatives have become extremely important in recent years. Chapter 6 is devoted to this subject, covering examples of key derivatives such as caps and swaptions and explaining how discrete time interest rate models are used for derivative valuation.

Chapter 7 provides a brief look at models with infinite sample spaces. This seemingly innocuous extension leads to significant mathematical complications and technicalities, and so this chapter will be most appealing to readers whose interests lean in the direction of abstract mathematics.

## Suggested Readings

The aim here is to provide some suggestions for further study, not to give an account of which researchers are responsible for specific results. Most of the references that will be mentioned are books, and some of these have very comprehensive bibliographies of old research. This discussion is for the reader who wishes to learn more mathematical finance, not history.

I will begin with the prerequisites, starting with basic probability theory. Feller (1968, 1971) is a classic still worth reading. I used Olkin, Gleser, and Derman (1980) for teaching probability courses in the 1970s and 80s. More recent texts on basic probability theory include Ross (1997a), Karr (1993), and Pitman (1993). All these texts assume the reader knows some calculus, but measure theory is not needed.

This book uses a lot of linear algebra and matrix theory, another subject where the newer books are no better than a classic, namely, Gantmacher (1959). Nevertheless, here are some newer books: Brown (1991), Roman (1992), Lay and Guardino (1997), and Riess et al. (1997).

Growing out of linear algebra is the subject of linear programming, the problem of maximizing or minimizing a linear objective function subject to some linear constraints. The appendix provides a quick overview of this subject as well as a list of good references. A closely related subject, quadratic programming, involves similar optimization problems, differing only in that the objective function is a quadratic function. Even more general are convex optimization problems, also called nonlinear programming problems, where the objective function is not necessarily quadratic. Such problems arise in finance when a portfolio manager seeks to maximize expected utility. Some good references include Jeter (1986), Hayhurst (1987), Bazaraa (1993), and Rockafellar (1997).

One subject of mathematics that should not be ignored is introductory analysis. This has to do with things like convergence, open and closed sets, functions, and limits. The books by Bartle and Sherbert (1992), Mikusinski and Mikusinski (1993), Berberian (1994), and Browder et al. (1996) are popular texts for this area.

Many of the preceding mathematical topics are covered in the introductions to the mathematics that are useful in economics by Klein (1973), Chiang (1974), and Ostaszewski (1993). These books are highly

recommended, because while giving primary emphasis to the mathematical tools, they also explain how the math is used in economics, thereby providing some economic background for the study of financial markets. In this same category, but focusing more narrowly on the application of optimization theory to economics, is the book by Dixit (1990).

So much for the prerequisites. It is not necessary to be an expert in any of the preceding areas, but you should be familiar with them.

I now turn to three areas that are developed in this book and that the reader may wish to investigate further. For discrete time stochastic processes (random walks, Bernoulli processes, Markov chains, martingales, etc.) there are several introductory books to choose from: Hoel, Port, and Stone (1972), Cinlar (1975), Karlin and Taylor (1975, 1981), Taylor and Karlin (1984), Ross (1995), (1997b), Kijima (1997) and Norris (1998). At a more advanced level (some measure theory may be used) one should be aware of the classics by Doob (1953), Neveu (1975) and Revuz (1984) as well as the more recent books by Durrett (1991) and Williams (1991).

Another mathematical topic that is developed in this book is dynamic programming. This has to do with the optimal control of a stochastic process. In the common situation where the process is Markovian, this topic is called Markov decision theory. Here one can do no better than look at the work by Bertsekas (1976), Denardo (1982), Whittle (1982, 1983), and Puterman (1994).

Finally I come to financial economics. Until recently, most of the books on the theory of security markets were written by finance professors and thus tended to emphasize economic theory at the expense of probability modeling. Markowitz (1990) is the definitive reference on single period portfolio management. Ingersoll (1987) and Duffie (1992) provide good, broad treatments of both discrete and continuous time models. Other books containing some general treatments of discrete time models, but presented in an older fashion, are Jarrow (1988), Huang and Litzenberger (1988), and Eatwell, Milgate, and Newman (1989). Meanwhile, three excellent books with a narrower focus, namely, on discrete and continuous time models of derivatives, are by Cox and Rubenstein (1985), Hull (1993), and Jarrow and Turnbull (1996). Also worth looking at is the book by Wilmott, Dewynne, and Howison (1993), which studies option pricing from the partial differential equation perspective, and the one by Dixit and Pindyck (1994), which studies capital investment decisions by firms and thus covers some of the same ground as one would when investing in securities.

In the last few years a variety of finance books have been written by mathematicians. These tend to emphasize probabilistic rather than economic arguments. Luenberger (1998) and Panjer et al. (1998) take very broad, introductory perspectives. Baxter and Rennie (1996), Lamberton and Lapeyre (1996), Elliott and Kopp (1998), and Mikosch (1998) provide good introductions to the continuous time theory after first developing some discrete time theory. All of these might be sensible for a first year graduate course.



At a more advanced, research level, one can find comprehensive treatments by Duffie (1988), Dothan (1990), Merton (1990), Musiela and Rutkowski (1997), Bingham and Kiesel (1998), and Bjork (1998). Korn (1997) and Karatzas and Shreve (1998) provide advanced studies of optimal consumption/investment problems. Rebonato (1998) focuses on models of interest rate derivatives.

## Final Remarks

The first two printings of this book had a number of typographical errors. These are listed on my web page: *www.uic.edu/~srpliska*. If you discover any errors in this printing, please bring them to my attention by contacting me at: *srpliska@uic.edu*.

A solutions manual for all the exercises in this book has been prepared for instructors who adopt this book for classroom use. Instructors should contact me about this, and I will mail a copy free of charge.

Stanley R. Pliska

# Acknowledgments

This book grew out of lecture notes, first organized in a careful fashion for a 1991 PhD class in Japan, while I was a Yamaichi Visiting Professor of Finance at Tsukuba University. I am indebted to Masaaki Kijima for making this experience possible. The work continued in 1992 while I was a Distinguished Visiting Fellow at the London School of Economics (thank you Michael J. P. Selby) and a Visiting Scholar at the University of Warwick (thank you Stewart Hodges).

A preliminary version of the book was tried out in a 1994 PhD class at the University of Illinois at Chicago. The reaction of the students was very useful, especially the feedback from Bill Francis and Rashida Dahodwala.

In January, 1995, I took a close-to-final version to the Program on Financial Mathematics at the Isaac Newton Institute for Mathematical Sciences, University of Cambridge (I am indebted to Chris Rogers for making this rewarding experience possible). There, parts of the book were used for a course, and copies of the whole book were made available to visiting researchers. The feedback I received and the hospitality of the Institute while I was there for six months as a Prudential Distinguished Visiting Fellow were important factors in the final stages of book preparation. In particular, careful comments by Abel Cadenillas and Peter Lakner, two visiting researchers who took copies for 1995–96 courses at their respective universities, were especially helpful. Also, Ruediger Kiesel provided some useful feedback.

I am very grateful to Andriy L. Turinskiy, a graduate mathematics student at my university, for helping me prepare the solutions manual. I also owe thanks to various people, especially Tomasz Bielecki, John Fuqua, and Edward Kao, for pointing out some typographical errors in the first two printings of this book.

# 1 Single Period Securities Markets



## 1.1 Model Specifications

Single period models are obviously unrealistic representations of complex, time-varying, random phenomena such as stock and bond prices. But they have the virtues of being mathematically simple as well as being able to illustrate many of the important economic principles associated with even the most complex, continuous time models. Hence single period models are worth studying for introductory purposes.

The following elements of the basic, single period model are specified as data:

- Initial date $t = 0$ and terminal date $t = 1$, with trading and consumption possible at these two dates.
- A finite sample space $\Omega$ with $K < \infty$ elements:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$$

  Here each $\omega \in \Omega$ should be thought of as a possible state of the world, the value of which is unknown at time $t = 0$ but which becomes apparent to the investors at time $t = 1$.

- A probability measure $P$ on $\Omega$, with $P(\omega) > 0$ for all $\omega \in \Omega$.
- A *bank account* process $B = \{B_t : t = 0, 1\}$, where $B_0 = 1$ and $B_1$ is a random variable.[1] The bank account process will be distinguished from the other securities because its time $t = 1$ price $B_1(\omega)$ will be assumed to be strictly positive for all $\omega \in \Omega$. Usually, in fact, $B_1 \geqslant 1$, in which case $B_1$ should be thought of as the time $t = 1$ value of the bank account when \$1 is deposited at time $t = 0$ and $r \equiv B_1 - 1 \geqslant 0$ should be thought of as the *interest rate*. For many applications the quantities $r$ and $B_1$ are taken to be deterministic scalars. If necessary for a particular application,

however, $B_1$ can be a positive random variable with $r$ violating the constraint $r \geqslant 0$.

- A *price process* $S = \{S_t : t = 0, 1\}$, where $S_t = (S_1(t), S_2(t) \ldots, S_N(t)), N < \infty$, and $S_n(t)$ is the time $t$ price of security $n$. For many applications these $N$ risky securities are stocks. The time $t = 0$ prices are positive scalars that are known to the investors, whereas the time $t = 1$ prices are non-negative random variables whose values become known to the investors only at time $t = 1$. When $N = 1$, it is convenient to simply write $S_t$ for the time $t$ price.

Having specified all the data describing the model, the next step is to define several quantities of interest. A *trading strategy* $H = (H_0, H_1, \ldots, H_N)$ describes an investor's portfolio as carried forward from time $t = 0$ to time $t = 1$. In particular, the scalar $H_0$ is the number of dollars invested in the savings account, and for $n \geqslant 1$ the scalar $H_n$ is the number of units of security $n$ (for example, shares of stock) held between times 0 and 1. In general, $H_n$ can be positive or negative (negative means borrowing or selling short), but sometimes there are constraints specified for the trading strategies to be admissible (for example, $H_n \geqslant 0$ for $n \geqslant 1$; that is, no short selling of the risky securities).

The *value process* $V = \{V_t : t = 0, 1\}$ describes the total value of the portfolio at each point in time. By simple bookkeeping this is

$$V_t \equiv H_o B_t + \sum_{n=1}^{N} H_n S_n(t), \qquad t = 0, 1$$

Note that the value process depends on the choice of the trading strategy $H$ and that $V_1$ is a random variable.

The *gains process* $G$ is a random variable that describes the total profit or loss generated by the portfolio between times 0 and 1. Since $H_n(S_n(1) - S_n(0))$ is the net profit due to investment in the $n$th security (similarly for the bank account), the gains process is

$$G \equiv H_0 r + \sum_{n=1}^{N} H_n \Delta S_n$$

where, by standard notation, $\Delta S_n \equiv S_n(1) - S_n(0)$.

A simple calculation verifies that

$$V_1 = V_0 + G \tag{1.1}$$

Hence equation (1.1) says that any change in the value of the portfolio must be due to a profit or loss in the investment and not, for example, due to the addition of funds from an outside source.

The movement of the security prices relative to each other will be important to study, so it is convenient to normalize the prices in such a way that the bank account becomes constant. In other words, we are going to make the bank account the *numeraire*. We do this by defining the *discounted price process* $S^* = \{S_t^* : t = 0, 1\}$ by setting $S_t^* \equiv (S_1^*(t), \ldots, S_N^*(t))$ and



$$S_n^*(t) \equiv S_n(t)/B_t, \qquad n = 1, \ldots, N; \quad t = 0, 1$$

the *discounted value process* $V^* = \{V_t^* : t = 0, 1\}$ by

$$V_t^* \equiv H_0 + \sum_{n=1}^{N} H_n S_n^*(t), \qquad t = 0, 1$$

and the *discounted gains process* $G^*$ by the random variable

$$G^* \equiv \sum_{n=1}^{N} H_n \Delta S_n^*$$

where, as one should guess, $\Delta S_n^* \equiv S_n^*(1) - S_n^*(0)$. With some more elementary bookkeeping, one eventually obtains

$$V_t^* = V_t / B_t, \qquad t = 0, 1 \tag{1.2}$$

as well as the discounted counterpart of equation (1.1), namely,

$$V_1^* = V_0^* + G^* \tag{1.3}$$

**Example 1.1** Suppose $K = 2$, $N = 1$, $r = 1/9$, $S_0 = 5$, $S_1(\omega_1) = 20/3$ and $S_1(\omega_2) = 40/9$. Then $B_1 = 1 + r = 10/9$, $S_1^*(\omega_1) = 6$, and $S_1^*(\omega_2) = 4$. For an arbitrary trading strategy $H$ we have $V_0 = V_0^* = H_0 + 5H_1$ as well as

$$V_1 = (10/9)H_0 + H_1 S_1 \qquad V_1^* = H_0 + H_1 S_1^*$$
$$G = (1/9)H_0 + H_1(S_1 - 5) \qquad G^* = H_1(S_1^* - 5)$$

Hence in state $\omega_1$

$$V_1 = (10/9)H_0 + (20/3)H_1 \qquad V_1^* = H_0 + 6H_1$$
$$G = (1/9)H_0 + (5/3)H_1 \qquad G^* = H_1$$

whereas in state $\omega_2$

$$V_1 = (10/9)H_0 + (40/9)H_1 \qquad V_1^* = H_0 + 4H_1$$
$$G = (1/9)H_0 - (5/9)H_1 \qquad G^* = -H_1$$

It is easy to verify that equations (1.1) to (1.3) hold for both $\omega \in \Omega$.

**Example 1.2** With everything else the same as in example 1.1, take $K = 3$ and set $S_1(\omega_3) = 30/9$, so that $S_1^*(\omega_3) = 3$. The other quantities of interest are left to the reader. Although this was a simple modification, it will be shown later that we have substantially changed the character of this model.

**Example 1.3** For a simple model featuring two risky securities, suppose $K = 3$, $r = 1/9$ and the price process is as follows:

| $n$ | $S_n(0)$ | $S_n(1)$ | | |
|---|---|---|---|---|
| | | $\omega_1$ | $\omega_2$ | $\omega_3$ |
| 1 | 5 | 60/9 | 60/9 | 40/9 |
| 2 | 10 | 40/3 | 80/9 | 80/9 |

It follows that the discounted price process is given by

| $n$ | $S_n^*(0)$ | $S_n^*(1)$ | | |
|---|---|---|---|---|
| | | $\omega_1$ | $\omega_2$ | $\omega_3$ |
| 1 | 5 | 6 | 6 | 4 |
| 2 | 10 | 12 | 8 | 8 |

The other quantities of interest are left to the reader.

**Example 1.4** Again, and as will be shown later, a small modification will create a model having substantially different character. With everything else the same as in example 1.3, we take $K = 4$ and set the prices in state $\omega_4$ to be $S_1(1) = 20/9$ and $S_2(1) = 120/9$. Now the discounted price process is:

| $n$ | $S_n^*(0)$ | $S_n^*(1)$ | | | |
|---|---|---|---|---|---|
| | | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ |
| 1 | 5 | 6 | 6 | 4 | 2 |
| 2 | 10 | 12 | 8 | 8 | 12 |

**Exercise 1.1** Verify (1.2).

**Exercise 1.2** Verify (1.3).

**Exercise 1.3** Specify $V$, $V^*$, $G$ and $G^*$ for

(a) Example 1.2
(b) Example 1.3
(c) Example 1.4

## 1.2 Arbitrage and other Economic Considerations

In order for the single period model to be reasonable from the economic standpoint, it must satisfy various criteria. For example, the model would be unreasonable if the investors were certain to be able to make a profit on a transaction, without any risk of losing money or even of failing to make a gain. Such would be the case if there existed a dominant trading strategy.

A trading strategy $\hat{H}$ is said to be *dominant* if there exists another trading strategy, say $\tilde{H}$, such that $\hat{V}_0 = \tilde{V}_0$ and $\hat{V}_1(\omega) > \tilde{V}_1(\omega)$ for all $\omega \in \Omega$. In



other words, both trading strategies start with the same amount of money, but the dominant one is certain to end up with more.

If $H$ is a trading strategy satisfying $V_0 = 0$ and $V_1(\omega) > 0$ for all $\omega \in \Omega$, then $H$ is dominant because it dominates the strategy which starts with zero money and does no investment at all. Conversely, if the trading strategy $\hat{H}$ dominates the trading strategy $\tilde{H}$, then by defining a new trading strategy $H = \hat{H} - \tilde{H}$ it follows by the linearity in the definition of $V$ that $V_0 = \hat{V}_0 - \tilde{V}_0 = 0$ and $V_1(\omega) = \hat{V}_1(\omega) - \tilde{V}_1(\omega) > 0$ for all $\omega \in \Omega$. In other words, the following is true:

**(1.4)** There exists a dominant trading strategy if and only if there exists a trading strategy satisfying $V_0 = 0$ and $V_1(\omega) > 0$ for all $\omega \in \Omega$.

Note that the condition in (1.4) is unreasonable from the economic standpoint; an investor starting with zero money should not have a guaranteed way of ending up with a positive amount of money. Hence a securities market model having a dominant trading strategy cannot be a realistic one.

Not surprisingly, if there exists a dominant trading strategy, then there exists a trading strategy which can transform a strictly negative initial wealth into a non-negative wealth. To see this, suppose $H$ satisfies the condition in (1.4). Then by (1.2) and the fact that $B_t > 0$, one has $V_0^* = 0$ and $V_1^*(\omega) > 0$ for all $\omega \in \Omega$. So by (1.3), $(H_1, \ldots, H_N)$ must be such that $G^*(\omega) > 0$ for all $\omega \in \Omega$. Now define a new strategy $\tilde{H}$ by setting $\tilde{H}_n = H_n$ for $n = 1, \ldots, N$ and

$$\tilde{H}_0 = -\sum_{n=1}^{N} H_n S_n^*(0) - \delta$$

where

$$\delta \equiv \min_{\omega} G^*(\omega) > 0$$

It follows from the definition of $\tilde{V}_t^*$ that $\tilde{V}_0^* = -\delta < 0$ and $\tilde{V}_1^*(\omega) = \tilde{V}_0^* + \tilde{G}^*(\omega) = -\delta + \tilde{G}^*(\omega) \geqslant 0$ for all $\omega \in \Omega$. Hence by (1.2), again, $\tilde{H}$ is as desired.

Conversely, suppose there is a trading strategy such as $\tilde{H}$. Then by reversing the preceding argument one sees that $(\tilde{H}_1, \ldots, \tilde{H}_N)$ is such that $\tilde{G}^*(\omega) > 0$ for all $\omega \in \Omega$. Hence upon setting $H_n = \tilde{H}_n$ for $n = 1, \ldots, N$ and

$$H_0 = -\sum_{n=1}^{N} \tilde{H}_n S_n^*(0)$$

it follows that the new trading strategy $H$ satisfies $V_0 = 0$ and $V_1(\omega) > 0$ for all $\omega \in \Omega$. In view of (1.4), this means there is another equivalent condition:

**(1.5)** There exists a dominant trading strategy if and only if there exists a trading strategy satisfying $V_0 < 0$ and $V_1(\omega) \geqslant 0$ for all $\omega \in \Omega$.

The existence of a dominant trading strategy is unsatisfactory from another standpoint: it leads to illogical pricing. For reasons which will soon

become clear, it is often useful to interpret $V_1(\omega)$ as the time $t = 1$ payoff of a contract or claim when state $\omega$ pertains, in which case $V_0$ can be interpreted as the time $t = 0$ price of this claim. But if the trading strategy $\hat{H}$ dominates $\tilde{H}$, then the contingent claims $\hat{V}$ and $\tilde{V}$ have the same prices even though the former claim has a strictly greater payoff in every state $\omega$. This is not consistent with reality.

The pricing of claims will be logically consistent if there is a *linear pricing measure*, that is, a non-negative vector $\pi = (\pi(\omega_1), \ldots, \pi(\omega_K))$ such that for every trading strategy $H$ you have

$$V_0^* = \sum_\omega \pi(\omega) V_1^*(\omega) = \sum_\omega \pi(\omega) V_1(\omega)/B_1(\omega)$$

Now the illogical pricing associated with dominant trading strategies no longer exists; each claim has a unique price, and a claim that pays more than another in every state will have a higher time $t = 0$ price.

If there is a linear pricing measure $\pi$, then by its definition and that of $V_t^*$ one has

**(1.6)** $$H_0 + \sum_{n=1}^{N} H_n S_n^*(0) = \sum_\omega \pi(\omega) \left[ H_0 + \sum_{n=1}^{N} H_n S_n^*(1)(\omega) \right]$$

Taking $H_1 = \ldots = H_N = 0$, it can be seen that the linear pricing measure must satisfy $\pi(\omega_1) + \ldots + \pi(\omega_K) = 1$; thus one can interpret $\pi$ as a probability measure on the sample space $\Omega$. Taking for arbitrary $i \in \{1, \ldots, N\}$ a trading strategy with $H_n = 0$ for all $n \neq i$, one sees that this equation implies

**(1.7)** $$S_n^*(0) = \sum_\omega \pi(\omega) S_n^*(1)(\omega), \quad n = 1, \ldots, N$$

Conversely, suppose $\pi$ is a probability measure on $\Omega$ satisfying (1.7). Then (1.6) is satisfied, and it follows that:

**(1.8)** The vector $\pi$ is a linear pricing measure if and only if it is a probability measure on $\Omega$ satisfying (1.7).

Since a linear pricing measure $\pi$ can be taken to be a probability measure, (1.7) says that the initial price of each security is equal to the *expectation*[2] under $\pi$ of the final discounted price. Similarly, by the original definition of $\pi$, the initial value $V_0$ of any portfolio is equal to the expectation under $\pi$ of the final discounted value of the portfolio.

It turns out there exists a close relationship between the concepts of dominant trading strategies and linear pricing measures:

**(1.9)** There exists a linear pricing measure if and only if there are no dominant trading strategies.

This important principle can be verified with linear programming duality theory.[3] In particular, let $\pi \in \mathbb{R}^K$ be a column vector, let $Z \in \mathbb{R}^{N+1}$ denote the column vector



$$Z = \begin{pmatrix} S_1^*(0) \\ \vdots \\ S_N^*(0) \\ 1 \end{pmatrix}$$

and let $\mathbb{Z}$ denote the $(N+1) \times K$ matrix

$$\mathbb{Z} \equiv \begin{pmatrix} S_1^*(1, \omega_1) \ldots S_1^*(1, \omega_K) \\ \vdots \qquad\qquad \vdots \\ S_N^*(1, \omega_1) \ldots S_N^*(1, \omega_K) \\ 1 \qquad \ldots \qquad 1 \end{pmatrix}$$

Then by (1.8) the existence of a linear pricing measure implies the existence of a solution to the linear program

**(1.10)** $$\begin{aligned} &\text{maximize} && (0, \ldots, 0)\pi \\ &\text{subject to} && \mathbb{Z}\pi = Z \\ & && \pi \geqslant 0. \end{aligned}$$

By duality theory there must exist a solution $h = (h_1, \ldots, h_{N+1})$ to the *dual linear* program

**(1.11)** $$\begin{aligned} &\text{minimize} && hZ \\ &\text{subject to} && h\mathbb{Z} \geqslant 0 \end{aligned}$$

and the two optimal objective values must coincide (in which case they obviously equal zero). Now interpret the solution $h$ of (1.11) as a trading strategy, with the last component of $h$ corresponding to $H_0$. The objective function in (1.11) says that $V_0^* = 0$, whereas the constraint says that $V_1^*(\omega) \geqslant 0$ for all $\omega \in \Omega$. Since the minimizing strategy $h$ has an objective value equal to zero, there cannot be any trading strategies with $V_0 < 0$ and $V_1(\omega) \geqslant 0$ for all $\omega \in \Omega$. Hence by (1.5) the existence of a linear pricing measure implies there cannot be any dominant trading strategies.

Conversely, if there are no dominant trading strategies, then (1.11) has a solution, namely, $h = 0$. It follows by duality theory that (1.10) has a solution $\pi$ which, as explained above, can be taken as the linear pricing measure.

To summarize matters up to this point, securities market models that permit dominant trading strategies are unreasonable from the economic point of view. Moreover, models without dominant strategies are reasonable, it would seem, because they are accompanied by linear pricing measures. Hence it makes sense to concentrate attention on the latter kind of model. But before agreeing to drop from consideration all models with dominant trading strategies, it is worth mentioning that one can have an even less reasonable securities market model.

It is said that the *law of one price* holds for a securities market model if there do not exist two trading strategies, say $\hat{H}$ and $\tilde{H}$, such that

$\hat{V}_1(\omega) = \tilde{V}_1(\omega)$ for all $\omega \in \Omega$ but $\hat{V}_0 > \tilde{V}_0$. In other words, if the law of one price holds, then there is no ambiguity about the time $t = 0$ price of any claim. On the other hand, the law of one price does not hold if there are two different trading strategies that yield the same time $t = 1$ payoff but the initial values of the two corresponding portfolios are different. This notion was mentioned above, just following principle (1.5).

Notice that if there do not exist two distinct trading strategies yielding the same payoff at time 1, then automatically the law of one price holds. On the other hand, if $\hat{H}$ and $\tilde{H}$ are as in the preceding paragraph, then $\hat{V}_1^* = \tilde{V}_1^*$ and $\hat{V}_0^* > \tilde{V}_0^*$ which, in turn, imply $\hat{G}^*(\omega) < \tilde{G}^*(\omega)$ for all $\omega \in \Omega$. Defining a new trading strategy $H$ by taking $H_n = \tilde{H}_n - \hat{H}_n$ for $n = 1, \ldots, N$ yields $G^*(\omega) > 0$ for all $\omega \in \Omega$. Finally, taking $H_0 = -\sum H_n S_n^*(0)$ leads to $V_0 = 0$ and $V_1(\omega) > 0$ for all $\omega \in \Omega$. Hence by (1.4) the following is true:

(1.12) If there are no dominant trading strategies, then the law of one price holds. The converse, however, is not necessarily true.

In other words, if the law of one price fails to hold, then there will exist a dominant trading strategy. The converse is not necessarily true, because, as will be illustrated in example 1.5 that follows, you can have a dominant trading strategy for a model that satisfies the law of one price. Thus failure of the law of one price is, in a sense, worse than having dominant trading strategies.

**Example 1.5** For a trivial example where the law of one price fails to hold, suppose $K = 2$, $N = 1$, $r = 1$, $S_0 = 10$, and $S_1(\omega_1) = S_1(\omega_2) = 12$. Hence $V_1$ is constant on $\Omega$, and for any scalar $\lambda$ there is an infinite number of trading strategies with $V_1 = \lambda$, each of which has a different value of $V_0$.

Now suppose $S_1(\omega_2)$ is changed to the value 8. For any $X \in \mathbb{R}^2$ there is a unique $H$ (and thus a unique time $t = 0$ price) such that $V_1 = X$, so the law of one price must hold. However, the trading strategy $H = (10, -1)$ satisfies $V_0 = 0$ and $V_1 = (8, 12)$, so it must be a dominant trading strategy.

Returning to the category of models that are without dominant trading strategies, it is clear that such models cannot have trading strategies that start with zero wealth and are certain to have a strictly positive amount of wealth at time $t = 1$. But what about trading strategies that start with zero wealth, cannot lose any money, and end up with a strictly positive amount of wealth at time $t = 1$ in at least one of the states $\omega$, but not all? In other words, investors would have the possibility of being able to make a profit on a transaction without being exposed to the risk of incurring a loss. Such an investment opportunity is called an arbitrage opportunity, and it is unreasonable from the economic standpoint.

Formally, an *arbitrage opportunity* is some trading strategy $H$ such that



(a) $V_0 = 0$.
(b) $V_1 \geqslant 0$, and
(c) $EV_1 > 0$.

Note that an arbitrage opportunity is a riskless way of making money: you start with nothing and, without any chance of going into debt, there is a chance of ending up with a positive amount of money. If such a situation were to exist, then everybody would 'jump in' with this trading strategy, affecting the prices of the securities. This economic model would not be in equilibrium. Hence for our single period model to be sensible from the economic standpoint, there cannot exist any arbitrage opportunities.

The following principle is true by (1.4) and example 1.6, which follows.

(1.13) If there exists a dominant trading strategy, then there exists an arbitrage opportunity, but the converse is not necessarily true.

**Example 1.6** Suppose $K = 2$, $N = 1$, $r = 0$, $S_0 = 10$, $S_1(\omega_1) = 12$, and $S_1(\omega_2) = 10$ (with one stock, the subscript denotes time). The trading strategy $H = (-10, 1)$ is an arbitrage opportunity, because $V_0 = 0$ and $V_1 = (2, 0)$. However, there are no dominant trading strategies, because $\pi = (0, 1)$ is a linear pricing measure.

From (1.2) and the fact that $B_t > 0$ for all $t$ and $\omega$, it follows easily that $H$ is an arbitrage opportunity if and only if

(a) $V_0^* = 0$,
(b) $V_1^* \geqslant 0$, and
(c) $EV_1^* > 0$.

In fact, there is still another equivalent condition:

(1.14) $H$ is an arbitrage opportunity if and only if
(a) $G^* \geqslant 0$, and
(b) $EG^* > 0$.

To see this, suppose $H$ is an arbitrage opportunity. By (1.3), $G^* = V_1^* - V_0^*$, so by the preceding remark $G^* \geqslant 0$ and $EG^* = EV_1^* - EV_0^* = EV_1^* > 0$. Conversely, suppose (a) and (b) in (1.14) are satisfied by some trading strategy $\hat{H}$. Then consider the strategy $H = (H_0, \hat{H}_1, \ldots \hat{H}_N)$, where

$$H_0 = -\sum_{n=1}^{N} \hat{H}_n S_n^*(0)$$

Under $H$ one has $V_0^* = 0$. Moreover, by (1.3) one has $V_1^* = V_0^* + G^* = G^*$. Hence (a) and (b) in (1.14) imply $V_1^* \geqslant 0$ and $EV_1^* > 0$, in which case $H$ is an arbitrage opportunity by the preceding remark.

In summary, and as illustrated in figure 1.1, all single period securities market models can be classified into four categories: (1) there are no arbitrage opportunities, (2) there are arbitrage opportunities but no dominant

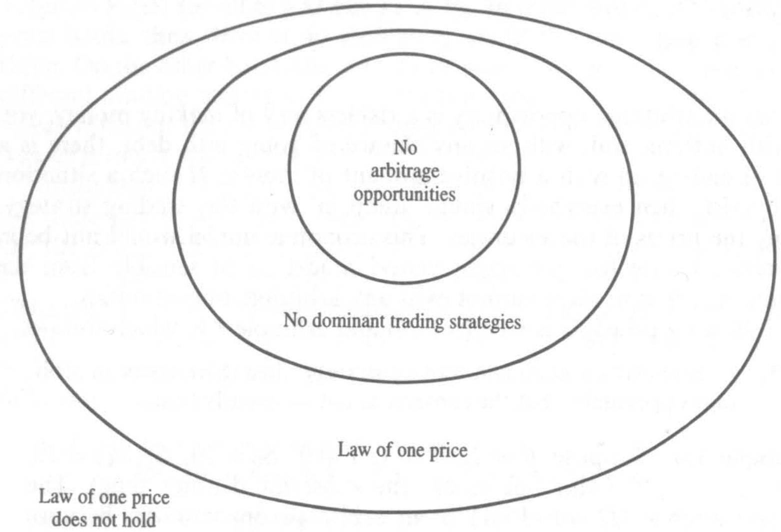


**Figure 1.1** Classification of securities market models

trading strategies, (3) there are dominant trading strategies but the law of one price holds, and (4) the law of one price does not hold. And only the first category is reasonable from the economic point of view.

Unfortunately, it is not so easy to check directly whether a model has any arbitrage opportunities, at least when there are two or more risky securities. But there is an important necessary and sufficient condition for the model to be free of arbitrage opportunities. This condition involves the discounted price process and something called a risk neutral probability measure, which is a special kind of linear pricing measure. It will be the subject of the next section.

---

**Exercise 1.4** Consider the model with $K = 3$, $N = 2$, $r = 0$, and the following security prices:

| $n$ | $S_n(0)$ | $S_n(1)(\omega_1)$ | $S_n(1)(\omega_2)$ | $S_n(1)(\omega_3)$ |
|---|---|---|---|---|
| 1 | 4 | 8 | 6 | 3 |
| 2 | 7 | 10 | 8 | 4 |

Show that there exist dominant trading strategies and that the law of one price holds.

**Exercise 1.5** Show for example 1.3 that there are no dominant trading strategies but there exists an arbitrage opportunity.

---



## 1.3 Risk Neutral Probability Measures

In the preceding section it was explained that if there exists a linear pricing measure, then there cannot be any dominant trading strategies, although there can still be arbitrage opportunities. In order to rule out arbitrage opportunities, we need a little bit more: there must exist a linear pricing measure which gives strictly positive mass to every state $\omega \in \Omega$.

A probability measure $Q$ on $\Omega$ is said to be a *risk neutral* probability measure if

(a) $Q(\omega) > 0$, all $\omega \in \Omega$, and
(b) $E_Q[\Delta S_n^*] = 0$, $n = 1, 2, \ldots, N$.

Here the notation $E_Q[X]$ means the expected value of the random variable $X$ under the probability measure $Q$. Note that

$$E_Q[\Delta S_n^*] = E_Q[S_n^*(1) - S_n^*(0)] = E_Q[S_n^*(1)] - S_n^*(0),$$

so $E_Q[\Delta S_n^*] = 0$ is equivalent to

$$E_Q[S_n^*(1)] = S_n^*(0), \qquad n = 1, 2, \ldots, N \tag{1.15}$$

This is essentially the same as (1.7) and says that under the indicated probability measure the expected time $t = 1$ discounted price of each risky security is equal to its initial price. Hence a risk neutral probability measure is just a linear pricing measure giving strictly positive mass to every $\omega \in \Omega$.

We now come to a very important result.

**(1.16)** There are no arbitrage opportunities if and only if there exists a risk neutral probability measure $Q$.

Before proving this result, it is worthwhile to look at some examples and provide some intuition.

**Example 1.1 (continued)** We want to find strictly positive numbers $Q(\omega_1)$ and $Q(\omega_2)$ so that (1.15) is satisfied, that is,

$$5 = 6Q(\omega_1) + 4Q(\omega_2)$$

Also $Q$ must be a probability measure, so it must satisfy

$$1 = Q(\omega_1) + Q(\omega_2)$$

It is easy to see that $Q(\omega_1) = Q(\omega_2) = 1/2$ satisfies both equations, so this is a risk neutral probability measure, and by (1.16) there cannot be any arbitrage opportunities.

Of course, with this simple example it is easy to see from the discounted price process that there cannot be any arbitrage opportunities. Indeed, principle (1.16) is easy to understand in the case where there is a single risky security (i.e., $N = 1$). From the definition, there is an arbitrage opportunity if and only if one can take a position $H_1$ in the discounted price process $S^*$ that will possibly gain but cannot lose. This means that either

$\Delta S^* \geqslant 0$ with $\Delta S^*(\omega) > 0$ for at least one $\omega \in \Omega$ or $\Delta S^* \leqslant 0$ with $\Delta S^*(\omega) < 0$ for at least one $\omega \in \Omega$. Clearly, in both cases it is impossible to find a strictly positive probability measure satisfying (1.15). On the other hand, if neither of these two cases applies, then one can find a risk neutral probability measure and there are no arbitrage opportunities.

**Example 1.2 (continued)** The system of equations to be solved, namely,

$$5 = 6Q(\omega_1) + 4Q(\omega_2) + 3Q(\omega_3)$$
$$1 = Q(\omega_1) + Q(\omega_2) + Q(\omega_3)$$

involves three unknowns but only two equations, so we will solve for two of the unknowns in terms of the third, say $Q(\omega_1)$. Thus this system will be satisfied for an arbitrary real number $Q(\omega_1)$ if

$$Q(\omega_2) = 2 - 3Q(\omega_1) \quad \text{and} \quad Q(\omega_3) = -1 + 2Q(\omega_1)$$

Now for $Q$ to be a strictly positive probability measure we must have $Q(\omega_i) > 0$ for all $i$. Using the preceding two equations, this leads to three inequalities for $Q(\omega_1)$, including $Q(\omega_1) > 0$. In view of its equation, $Q(\omega_2) > 0$ if and only if $Q(\omega_1) < 2/3$. Similarly, $Q(\omega_3) > 0$ if and only if $Q(\omega_1) > 1/2$. Hence our solution will be a strictly positive probability measure if and only if $1/2 < Q(\omega_1) < 2/3$. In other words, $Q = (\lambda, 2 - 3\lambda, -1 + 2\lambda)$ is a risk neutral probability measure for each value of the scalar $\lambda$ satisfying $1/2 < \lambda < 2/3$, and there are no arbitrage opportunities.

**Example 1.3 (continued)** We seek a solution of

$$5 = 6Q(\omega_1) + 6Q(\omega_2) + 4Q(\omega_3)$$
$$10 = 12Q(\omega_1) + 8Q(\omega_2) + 8Q(\omega_3)$$
$$1 = Q(\omega_1) + Q(\omega_2) + Q(\omega_3)$$

There exists a unique solution to these equations, namely, $Q(\omega_1) = Q(\omega_3) = 1/2$, $Q(\omega_2) = 0$. This is a linear pricing measure, but this solution is not strictly positive, so there does not exist a risk neutral probability measure. By (1.16), therefore, there must exist an arbitrage opportunity. It takes a bit of work to find one; we will come back to this example later.

Example 1.3 illustrates why the intuition which worked for the case of a single risky security does not work when there are two or more risky securities. Looking at the discounted price process for the first security, it is clear that we can find a strictly positive probability measure $Q$ satisfying $E_Q[S_1^*(1)] = 5$. Similarly for the second risky security. The problem, however, is that we cannot find a single strictly positive probability measure that will simultaneously work for both securities. The interactions between these two securities permit arbitrage opportunities even though, taken individually, the securities seem acceptable. And it is these kinds of interactions which make the intuitive understanding of principle (1.16) much more difficult when there are two or more risky securities.



These three examples illustrate the three kinds of situations that can arise: either (1) there is a unique risk neutral probability measure, (2) there are infinitely many risk neutral probability measures, or (3) there are no risk neutral probability measures.

We now return to the explanation of (1.16) for the case where $N \geqslant 2$. For a general, single period model, consider the set

$$\mathbb{W} = \{X \in \mathbb{R}^K : X = G^* \text{for some trading strategy H}\}$$

One should think of $\mathbb{W}$ as a set of random variables, and because of (1.3) one should think of each $X \in \mathbb{W}$ as a possible time $t = 1$ discounted wealth when the initial value of the investment is zero. Note that $\mathbb{W}$ is actually a linear subspace of $\mathbb{R}^K$, that is, for any $X, \hat{X} \in \mathbb{W}$ and any scalars $a$ and $b$ one also has $aX + b\hat{X} \in \mathbb{W}$.

Next, consider the set

$$\mathbb{A} \equiv \{X \in \mathbb{R}^K : X \geqslant 0, X \neq 0\}$$

This is just the non-negative orthant of $\mathbb{R}^K$. In view of (1.14) it is apparent that there exists an arbitrage opportunity if and only if $\mathbb{W} \cap \mathbb{A} \neq \emptyset$, that is, if and only if the subspace $\mathbb{W}$ intersects with the non-negative orthant of $\mathbb{R}^K$. Hence to find an arbitrage opportunity in a model for which there is no risk neutral probability measure, one can use linear algebra to characterize $\mathbb{W}$ quantitatively and then compute a vector in its intersection with $\mathbb{A}$.

Now corresponding to the subspace $\mathbb{W}$ is the *orthogonal subspace*

$$\mathbb{W}^\perp \equiv \{Y \in \mathbb{R}^K : X \cdot Y = 0 \quad \text{for all} \quad X \in \mathbb{W}\}$$

where $X \cdot Y = X(\omega_1)Y(\omega_1) + \ldots + X(\omega_K)Y(\omega_K)$ denotes the *inner product* of $X$ and $Y$. If you consider the geometric picture for the case $K = 2$ (see figure 1.2) or even the case $K = 3$, it should be easy to believe that $\mathbb{W} \cap \mathbb{A} = \emptyset$ implies the existence of a ray in $\mathbb{W}^\perp$ along which every component of every point not at the origin is strictly positive.[4] In particular, along this ray there will exist one point whose components sum to one, in which case this point can be interpreted as a probability measure. In other words, denoting

$$\mathbb{P}^+ \equiv \{X \in \mathbb{R}^K : X_1 + \ldots + X_K = 1, X_1 > 0, \ldots, X_K > 0\}$$

the geometry suggests that $\mathbb{W} \cap \mathbb{A} = \emptyset$ if and only if $\mathbb{W}^\perp \cap \mathbb{P}^+ \neq \emptyset$.

Since $\Delta S_n^* \in \mathbb{W}$ for all $n$, it follows that any element of the set $\mathbb{W}^\perp \cap \mathbb{P}^+$ is actually a risk neutral probability measure. Conversely, if $Q$ is any risk neutral probability measure, then for any $G^* \in \mathbb{W}$ (with corresponding trading strategy $H$) we have

$$\textbf{(1.17)} \qquad E_Q G^* = E_Q\left[\sum_{n=1}^{N} H_n \Delta S_n^*\right] = \sum_{n=1}^{N} H_n E_Q[\Delta S_n^*] = 0$$

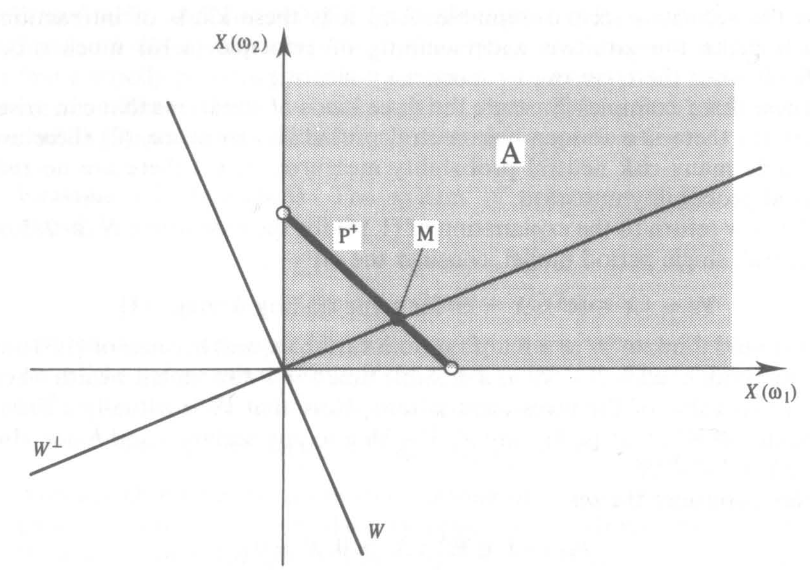


**Figure 1.2** Geometric interpretation of the risk neutral probability measures

so $Q \in \mathbb{W}^{\perp} \cap \mathbb{P}^{+}$. Thus if we let $\mathbb{M}$ denote the set of all risk neutral probability measures, we have that

$$\mathbb{M} = \mathbb{W}^{\perp} \cap \mathbb{P}^{+}$$

Moreover, by the geometric intuition used above we conjecture that $\mathbb{W} \cap \mathbb{A} = \emptyset$ if and only if $\mathbb{M} \neq \emptyset$. This conjecture, of course, is the same as principle (1.16).

In order to make this argument more rigorous and apply it to the case of general $K$, it is convenient to use a version of the Hahn–Banach theorem called the *separating hyperplane theorem*. Consider the set

$$\mathbb{A}^{+} = \{X \in \mathbb{A} : EX \geqslant 1\}$$

This is a closed and convex[5] subset of $\mathbb{R}^K$, and the absence of arbitrage opportunities implies $\mathbb{W}$ and $\mathbb{A}^{+}$ are disjoint. Hence by the separating hyperplane theorem there exists some $Y \in \mathbb{W}^{\perp}$ such that $X \cdot Y > 0$ for all $X \in \mathbb{A}^{+}$. For each $k = 1, \ldots, K$ we can find a vector $X$ in $\mathbb{A}^{+}$ whose $k$th component is positive and other components are zeros, so every component of $Y$ must be strictly positive. By setting $Q(\omega_k) = Y(\omega_k)/[Y(\omega_1) + \ldots + Y(\omega_K)]$, it is clear that $Q$ is a probability measure with $Q \in \mathbb{W}^{\perp}$. Since $\Delta S_n^* \in \mathbb{W}$ for all $n$, we conclude that $Q$ is a risk neutral probability measure.

What about the converse of (1.16)? This is easy. If $Q$ is a risk neutral probability measure, then, as explained above, for an arbitrary trading strategy $H$ we have equation (1.17), which shows that $G^*$ cannot satisfy



both $G^* \geqslant 0$ and $EG^* > 0$. Hence by (1.14) there cannot be any arbitrage opportunities, and so our conjecture and principle (1.16) are verified.

**Example 1.3 (continued)** We want to compute $\mathbb{W} \cap \mathbb{A}$, which we know is non-empty. Knowing $S_n^*$, one computes $\Delta S_n^*$ to be as follows:

| $n$ | $\Delta S_n^*(\omega_1)$ | $\Delta S_n^*(\omega_2)$ | $\Delta S_n^*(\omega_3)$ |
|---|---|---|---|
| 1 | 1 | 1 | −1 |
| 2 | 2 | −2 | −2 |

It follows that

$$\mathbb{W} = \{X \in \mathbb{R}^3 : X = (H_1 + 2H_2, H_1 - 2H_2, -H_1 - 2H_2) \quad \text{for some} \quad H_1, H_2 \in \mathbb{R}\}$$

Notice that $X_1 + X_3 = 0$ for all $X \in \mathbb{W}$. Conversely, given any vector $X$ with $X_1 + X_3 = 0$, one can readily find a unique trading strategy $H$ with $G^* = X$. Hence

$$\mathbb{W} = \{X \in \mathbb{R}^3 : X_1 + X_3 = 0\}$$

that is,

$$\mathbb{W}^{\perp} = \{Y \in \mathbb{R}^3 : Y = (\lambda, 0, \lambda) \text{for some } \lambda \in \mathbb{R}\}$$

Now comparing $\mathbb{W}$ and $\mathbb{A}$ we see that

$$\mathbb{W} \cap \mathbb{A} = \{X \in \mathbb{R}^3 : X_1 = X_3 = 0, X_2 > 0\}$$

So starting with any positive number $X_2$, we compute the trading strategy $H$ which gives rise to the time $t = 1$ portfolio value $(0, X_2, 0)$. This will be the solution of

$$H_1 + 2H_2 = 0$$
$$H_1 - 2H_2 = X_2$$

namely, $H_1 = X_2/2$ and $H_2 = -X_2/4$. Finally, upon setting

$$H_0 = -H_1 S_1^*(0) - H_2 S_2^*(0) = -(X_2/2)(5) - (-X_2/4)(10),$$

one obtains $H_0 = 0$. It is apparent that $H = (0, X_2/2, -X_2/4)$ is an arbitrage opportunity for every $X_2 > 0$.

**Exercise 1.6** Show that $\mathbb{W}$ and $\mathbb{W}^{\perp}$ are linear subspaces.

**Exercise 1.7** Specify $\mathbb{W}$ and $\mathbb{W}^{\perp}$ in the case of

(a) Example 1.1.
(b) Example 1.2.
(c) Example 1.4.

**Exercise 1.8** Determine either all the risk neutral probability measures or all the arbitrage opportunities in the case of example 1.4.

**Exercise 1.9** Suppose $K = 2$, $N = 1$, and the interest rate is a scalar parameter $r \geqslant 0$. Also, suppose $S_0 = 1$, $S_1(\omega_1) = u$ ('up'), and $S_1(\omega_2) = d$ ('down'), where the parameters $u$ and $d$ satisfy $u > d > 0$. For what values of $r$, $u$, and $d$ does there exist a risk neutral probability measure? Say what this measure is. For the complementary values of these parameters, say what all the arbitrage opportunities are.

**Exercise 1.10** Let $A$ denote the $(K+1) \times (K+2N)$ matrix

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \Delta S_1^*(\omega_1) & -\Delta S_1^*(\omega_1) & \Delta S_2^*(\omega_1) & \cdots & -\Delta S_N^*(\omega_1) & -1 & 0 & \cdots & 0 \\ \Delta S_1^*(\omega_2) & -\Delta S_1^*(\omega_2) & \Delta S_2^*(\omega_2) & \cdots & -\Delta S_N^*(\omega_2) & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \Delta S_1^*(\omega_K) & -\Delta S_1^*(\omega_K) & \Delta S_2^*(\omega_K) & \cdots & -\Delta S_N^*(\omega_K) & 0 & 0 & \cdots & -1 \end{bmatrix}$$

and let $b$ denote the $(K+1)$-component column vector $(1, 0, \dots, 0)'$. Show that

$$Ax = b, \quad x \geqslant 0, \quad x \in \mathbb{R}^{K+2N}$$

has a solution if and only if there exists an arbitrage opportunity.

**Exercise 1.11** Farkas's Lemma, a variation of the separating hyperplane theorem, says that given an $m \times n$ matrix $A$ and an $m$-dimensional column vector $b$, either

$$Ax = b, \quad x \geqslant 0, \quad x \in \mathbb{R}^n$$

has a solution or

$$yA \leqslant 0, \quad yb > 0, \quad y \in \mathbb{R}^m$$

has a solution, but not both. Use this and the results of exercise 1.10 to show that if there are no arbitrage opportunities, then there exists a risk neutral probability measure.

## 1.4 Valuation of Contingent Claims

A *contingent claim* is a random variable $X$ representing a payoff at time $t = 1$. You can think of a contingent claim as part of a contract that a buyer and a seller make at time $t = 0$. The seller promises to pay the buyer the amount $X(\omega)$ at time $t = 1$ if $\omega \in \Omega$ turns out to be the true state of the world. Hence, when viewed at time $t = 0$, the payoff $X$ is a random variable, and so the problem of interest is to determine the time $t = 0$ value of this payoff. In other words, what is the fair price that the buyer should pay the seller at time $t = 0$ in order for the two parties to be happy with their contract?



Now one might suppose that the value of a contingent claim would depend on the risk preferences and utility functions of the buyer and seller, but in a great many cases this is not so. It turns out that by the arguments of *arbitrage pricing theory* there is often a unique, correct, time $t = 0$ value for the contingent claim, a value that does not depend on the risk preferences of the parties who buy and sell this claim.

Here is the argument. A contingent claim $X$ is said to be *attainable* or *marketable* if there exists some trading strategy $H$, called the *replicating portfolio*, such that $V_1 = X$. In this case one says that $H$ *generates* $X$. Now suppose the time $t = 0$ price $p$ of $X$ is such that $p > V_0$. Then an astute individual would sell the contingent claim for $p$ at time $t = 0$, follow the trading strategy $H$ at a time $t = 0$ cost of $V_0$, and pocket the difference $p - V_0$. This individual has made a riskless profit, because at time $t = 1$ the value $V_1$ of the portfolio corresponding to $H$ is exactly equal to the obligation $X$ of the contingent claim in every state of the world. In other words, if $p > V_0$, then this astute individual could lock in a profit of $p - V_0$ by investing in a portfolio that provides exactly the right value to settle the obligation on the contingent claim.

Similarly, if $p < V_0$, then an astute individual would follow the trading strategy $-H$, thereby collecting the amount $V_0$ at time $t = 0$, and purchasing the contingent claim for the amount $p$, thereby locking in a risk free profit of $V_0 - p$. At time $t = 1$ the amount collected $X$ is exactly what is needed to settle the obligation $V_1$ associated with the trading strategy $-H$. Again, if $p < V_0$, then this astute individual could lock in a riskless profit of $V_0 - p$.

If $p = V_0$, then apparently we cannot use $H$ to create a riskless profit. So does this mean that $V_0$ is the correct value of $X$? Not necessarily, for suppose there is a second trading strategy, say $\hat{H}$, such that $\hat{V}_1 = X$ but $\hat{V}_0 \neq V_0$. Then even if $p = V_0$, one could use $\hat{H}$ and the argument above to lock in a riskless profit, thereby implying the different price $\hat{V}_0$. The problem here, of course, is that the law of one price does not hold. So for $V_0$ to be the unique, logical, time $t = 0$ price of $X$, it is necessary to assume that the law of one price does indeed hold. In this case we say that $V_0$ is the price of $X$ as implied by arbitrage pricing theory.

As explained in section 1.2, if there are no arbitrage opportunities, then there are no dominant trading strategies, and if there are no dominant trading strategies, then the law of one price holds. Thus by (1.16) the existence of a risk neutral probability measure implies the law of one price. Alternatively, we can see this directly from the following, very important calculation:

**(1.18)** If $Q$ is any risk neutral probability measure, then for every trading strategy $H$ one has

$$V_0 = V_0^* = E_Q V_0^* = E_Q[V_1^* - G^*] = E_Q V_1^* - E_Q\left[\sum_{n=1}^{N} H_n \Delta S_n^*\right]$$

$$= E_Q V_1^* - \sum_{n=1}^{N} H_n E_Q[\Delta S_n^*] = E_Q V_1^* - 0 = E_Q V_1^* = E_Q[V_1/B_1]$$

In other words, under $Q$ the expected, discounted, time $t = 1$ value of any portfolio is equal to its initial value. So if there is a positive probability that the portfolio will go up in value, then there also must be a positive probability of going down in value, and vice versa. Moreover, there is no way you can have two trading strategies $H$ and $\hat{H}$ with both $V_1 = \hat{V}_1$ and $V_0 \neq \hat{V}_0$, so the law of one price must hold.

Notice for future reference that the calculation in (1.18) does not depend on the choice of $Q$, because $V_1^*$ is the time $t = 1$ discounted value of the portfolio under some trading strategy. In other words, for a model where there are two or more risk neutral probability measures, $E_Q V_1^*$ is constant with respect to such $Q$.

Returning to the contingent claim $X$, by the arguments near the beginning of this section we have the following important *valuation concept*:

**(1.19)** If the law of one price holds, then the time $t = 0$ value of an attainable contingent claim $X$ is $V_0 = H_0 B_0 + \sum_{n=1}^{N} H_n S_n(0)$, where $H$ is the trading strategy that generates $X$.

If we have the stronger condition that the model is free of arbitrage opportunities, then we have the following, sensational result:

**(1.20)** *Risk neutral valuation principle*: If the single period model is free of arbitrage opportunities, then the time $t = 0$ value of an attainable contingent claim $X$ is $E_Q[X/B_1]$, where $Q$ is any risk neutral probability measure.

This follows immediately from (1.2), (1.18), (1.19), and the fact that $B_0 = 1$. We now turn to several examples.

**Example 1.1 (continued)** Suppose $r = 1/9$, $X(\omega_1) = 7$, and $X(\omega_2) = 2$. Then the time $t = 0$ value of $X$ is

$$E_Q[X/B_1] = (1/2)(9/10)7 + (1/2)(9/10)2 = 4.05$$

providing $X$ is attainable. How do we check this? One way is to try to compute the trading strategy $H$ that generates $X$. This can be done by solving

$$X/B_1 = V_1^* = V_0^* + G^* = 4.05 + H_1 \Delta S_1^*$$

There is one unknown, $H_1$, and two equations, one for each $\omega$, but both equations give the same solution, namely, $H_1 = 2.25$. To determine $H_0$ one can solve



$$4.05 = V_0 = H_0 + H_1 S_1 = H_0 + (2.25)(5)$$

to obtain $H_0 = -7.2$.

In summary, the contingent claim $X$ is indeed attainable. To generate it you start with 4.05, you borrow 7.2 at the riskless interest rate $r = 1/9$, and you use the sum $4.05 + 7.2 = 11.25$ to purchase $11.25 \div 5 = 2.25$ shares of the risky asset. At time $t = 1$ you must pay $(7.2)(10/9) = 8$ to settle the loan. The amount of money remaining in the portfolio will depend on $\omega$: in state $\omega_1$ this will be $V_1 = (2.25)(20/3) - 8 = 7$, whereas in state $\omega_2$ this will be $V_1 = (2.25)(40/9) - 8 = 2$. If the time $t = 0$ value of this contingent claim were different from 4.05, then you could use this trading strategy in the manner discussed at the beginning of this section to lock in a riskless profit.

**Example 1.7** For a general securities model, taking

$$X(\omega) = \begin{cases} 1, & \omega = \hat{\omega} \\ 0, & \omega \neq \hat{\omega} \end{cases}$$

for some $\hat{\omega} \in \Omega$ leads to the time $t = 0$ price (if $X$ is attainable)

$$E_Q[X/B_1] = \sum_{\omega} Q(\omega)X(\omega)/B_1(\omega) = Q(\hat{\omega})/B_1(\hat{\omega})$$

For this reason $Q(\hat{\omega})/B_1(\hat{\omega})$ is sometimes called the *state price* for state $\hat{\omega} \in \Omega$. Thus the time $t = 0$ price of an attainable contingent claim is simply the weighted sum across the states of the payoffs under $X$, with the weights being the state prices.

**Example 1.8 – Call Options** Suppose $N = 1$ and $X$ has the form

$$X = (S_1 - e)^+ = \max\{0, S_1 - e\}$$

where $e$ is a specified number called the *exercise price* or the *strike price*. Hence $X$ is the contingent claim corresponding to the right to purchase the risky security at time $t = 1$ for the amount $e$. If it turns out that $S_1 \geqslant e$, then at time $t = 1$ this right will be worth the difference $S_1 - e$, and so the option should be exercised. On the other hand, if $S_1 \leqslant e$, then at time $t = 1$ this right will be worth nothing, and so the option should not be exercised. If $X$ is attainable, then its time $t = 0$ price is

$$E_Q[X/B_1] = \sum_{\omega \in \Omega'} Q(\omega)[S_1(\omega) - e]/B_1(\omega)$$

where $\Omega' \equiv \{\omega \in \Omega : S_1(\omega) \geqslant e\}$.

**Example 1.1 (continued)** With $r = 1/9$ and $e = 5$, the time $t = 1$ value of the call option is

$$X(\omega) = \begin{cases} 5/3, & \omega = \omega_1 \\ 0, & \omega = \omega_2 \end{cases}$$

Hence if $X$ is attainable, then its time $t = 0$ value is

$$E_Q[X/B_1] = (1/2)(9/10)(5/3) = 0.75$$

To check whether $X$ is attainable, we shall try to compute a trading strategy that generates $X$. We solve the system of two equations (one for each state)

$$V_1 = H_0 B_1 + H_1 S_1 = X$$

for the two unknowns and obtain $H_1 = 0.75$ and $H_0 = -3$. So, indeed, $V_0 = H_0 + H_1 S_0 = -3 + (0.75)(5) = 0.75$ is the time $t = 0$ price of $X$.

**Example 1.9 – Put options** Suppose $N = 1$ and $X$ has the form

$$X = (e - S_1)^+ = \max\{0, e - S_1\}$$

Then $X$ is the contingent claim that gives the owner the right to sell the risky security at time $t = 1$ for the amount $e$. This option should be exercised if and only if $S_1 < e$.

**Example 1.2 (continued)** Consider an arbitrary contingent claim $X = (X_1, X_2, X_3)$. This claim is marketable if and only if $V_1 = H_0 B_1 + H_1 S_1 = X$ for some pair of numbers $H_0$ and $H_1$, that is, there exists a solution to the system of equations

$$\begin{aligned} \omega_1 &: \quad (10/9)H_0 + (20/3)H_1 = X_1 \\ \omega_2 &: \quad (10/9)H_0 + (40/9)H_1 = X_2 \\ \omega_3 &: \quad (10/9)H_0 + (30/9)H_1 = X_3 \end{aligned}$$

Since there are three equations with only two unknowns, perhaps there is no solution. Let's see. Using the third equation to substitute for $H_0$ in the first two gives

$$H_1 = (3X_1 - 3X_3)/10 \quad \text{and} \quad H_1 = (9X_2 - 9X_3)/10$$

Hence the contingent claim is attainable if and only if these two values of $H_1$ are the same, that is, if and only if

$$X_1 - 3X_2 + 2X_3 = 0 \qquad (1.21)$$

This example illustrates the general principle that not all the contingent claims are attainable whenever the underlying model has multiple risk neutral probability measures, a principle that will be developed in the next section.



**Exercise 1.12** For example 1.1 with $r = 1/9$, what is the price of a put option with exercise price $e = 5$? What trading strategy generates this contingent claim?

**Exercise 1.13 – Put-Call parity** Suppose the interest rate $r$ is a scalar, and let $c$ and $p$ denote the prices of a call and put, respectively, both having the same exercise price $e$. Show that either both are marketable or neither is marketable. Use risk neutral valuation to show that in the former case one has

$$c - p = S_0 - e/(1 + r)$$

## 1.5 Complete and Incomplete Markets

Just because, as will be assumed throughout this section, there exists a risk neutral probability measure, it does not necessarily follow that one can use the risk neutral valuation principle to determine the time $t = 0$ price of a contingent claim. The problem, of course, is that the contingent claim might not be marketable, in which case it is not clear what its time $t = 0$ price should be. In particular, there is no reason to be sure that $E_Q[X/B_1]$ is the correct value. We therefore need a convenient method for checking whether a contingent claim is indeed marketable. One method, as illustrated with example 1.1 in the preceding section, is to try to compute a generating trading strategy by solving a system of linear equations. A solution to such a system will exist if and only if the contingent claim is marketable. But there exist alternative methods.

The model is said to be *complete* if every contingent claim $X$ can be generated by some trading strategy. Otherwise, the model is said to be *incomplete*. It turns out there are simple ways to check whether a model is complete. One way is to understand when the system of linear equations mentioned just above will always have a solution.

**(1.22)** Suppose there are no arbitrage opportunities. Then the model is complete if and only if the number of states in $\Omega$ equals the number of independent vectors in $\{B_1, S_1(1), \ldots, S_N(1)\}$.

To see this, define the $K \times (N + 1)$ matrix $A$ by

$$A = \begin{bmatrix} B_1(\omega_1) & S_1(1)(\omega_1) & \cdots & S_N(1)(\omega_1) \\ B_1(\omega_2) & S_1(1)(\omega_2) & \cdots & S_N(1)(\omega_2) \\ \vdots & \vdots & & \vdots \\ B_1(\omega_K) & S_1(1)(\omega_K) & \cdots & S_N(1)(\omega_K) \end{bmatrix}$$

and consider column vectors $H = (H_0, H_1, \cdots, H_N)'$ and $X = (X_1, \ldots, X_K)$. Then the model is complete if and only if the system $AH = X$ has a solution

$H$ for every $X$. By linear algebra, this last fact will be true if and only if the matrix $A$ has rank $K$, that is, this matrix has $K$ independent columns.

**Example 1.1 (continued)** The matrix

$$A = \begin{bmatrix} 10/9 & 20/3 \\ 10/9 & 40/9 \end{bmatrix}$$

has two independent rows, so this model is complete

**Example 1.10** Suppose we take example 1.1 and add a second risky security with $S_2(0) = 54$, $S_2(1)(\omega_1) = 70$, and $S_2(1)(\omega_2) = 50$. Note that $Q = (1/2, 1/2)$ is still a risk neutral probability measure because $54 = (1/2)(9/10)70 + (1/2)(9/10)50$. Now

$$A = \begin{bmatrix} 10/9 & 20/3 & 70 \\ 10/9 & 40/9 & 50 \end{bmatrix}$$

but this still has rank two. Hence this augmented model is still complete, although the risky securities are redundant.

**Example 1.2 (continued)** The matrix

$$A = \begin{bmatrix} 10/9 & 20/3 \\ 10/9 & 40/9 \\ 10/9 & 10/3 \end{bmatrix}$$

has rank two, whereas $K = 3$, so this model is incomplete. Now we saw earlier that the risk neutral probability measures are of the form $Q = (\lambda, 2 - 3\lambda, -1 + 2\lambda)$, where $\lambda$ is any scalar satisfying $1/2 < \lambda < 2/3$. Suppose we take any such Q and then use the formula from the risk neutral valuation principle (1.20):

$$E_Q[X/B_1] = \lambda(9/10)X_1 + (2 - 3\lambda)(9/10)X_2 + (-1 + 2\lambda)(9/10)X_3$$

If $X$ is marketable, then this value will be the same for all $\lambda$ because it must coincide with $V_0$ under the generating trading strategy. Note that this value is the same if and only if equation (1.21) holds. Moreover, recall from the discussion of (1.21) that a contingent claim is marketable if and only if (1.21) holds. Putting this together, we see that a contingent claim in this model is marketable if and only if $E_Q[X/B_1]$ is the same value under every risk neutral probability measure. It turns out that this necessary and sufficient condition holds in general.

As stated earlier, throughout this section it will be assumed that $\mathbb{M} \neq \emptyset$, where $\mathbb{M}$ is the set of all risk neutral probability measures. Now if the contingent claim $X$ is attainable, then $E_Q[X/B_1]$ is constant with respect to all $Q \in \mathbb{M}$. This is because, as already discussed in connection with (1.18), one has $V_0 = E_Q[X/B_1]$ for all $Q \in \mathbb{M}$, where $V_0$ is the initial value of the replicating portfolio.



To show the converse, it suffices to suppose that the contingent claim $X$ is *not* attainable and then demonstrate that $E_Q[X/B_1]$ does *not* take the same value for all $Q \in \mathbb{M}$. Consider the $K \times (N+1)$ matrix $A$, the $(N+1)$-dimensional column vector $H$, and the $K$-dimensional column vector $X$ as described above in connection with (1.22). If X is not attainable, then there is no solution $H$ to the system $AH = X$. By a slightly modified version of Farkas's Lemma (see exercise 1.11), it follows that there must exist a row vector $\pi = (\pi_1, \ldots, \pi_K)$ satisfying

$$\pi A = 0, \quad \pi X > 0.$$

Let $\hat{Q} \in \mathbb{M}$ be arbitrary, and let the scalar $\lambda > 0$ be small enough so that

$$Q(\omega_k) \equiv \hat{Q}(\omega_k) + \lambda \pi_k B_1(\omega_k) > 0, \quad \text{all} \quad k = 1, \ldots, K.$$

Since $\pi$ times the 'zeroth' column of $A$ is zero, it follows that the quantity $Q$ which was just defined is actually a probability measure giving positive probability to each state $\omega \in \Omega$. Moreover, for any discounted price process $S_n^*$ we have

$$\begin{aligned} E_Q S_n^*(1) &= \sum Q(\omega_k)[S_n(1, \omega_k)]/[B_1(\omega_k)] \\ &= \sum \hat{Q}(\omega_k)[S_n(1, \omega_k)]/[B_1(\omega_k)] + \lambda \sum \pi_k S_n(1, \omega_k) \\ &= \sum \hat{Q}(\omega_k) S_n^*(1, \omega_k) \end{aligned}$$

where we used the fact that $\pi$ times the '$n$th' column of $A$ is zero. But $\hat{Q} \in \mathbb{M}$, so $\sum \hat{Q}(\omega_k) S_n^*(1)(\omega_k) = S_n^*(0)$, in which case we realize that $Q \in \mathbb{M}$.

It remains to show that the expected value of $X/B_1$ under $Q$ is different from the expected value under $\hat{Q}$. Denote $\delta \equiv \pi X$ and note that $\delta > 0$. Then

$$\begin{aligned} E_Q[X/B_1] &= \sum Q(\omega_k) X(\omega_k)/[B_1(\omega_k)] \\ &= \sum \hat{Q}(\omega_k) X(\omega_k)/[B_1(\omega_k)] + \lambda \sum \pi_k X(\omega_k) \\ &= E_{\hat{Q}}[X/B_1] + \lambda\delta \end{aligned}$$

In other words, $E_Q[X/B_1] \neq E_{\hat{Q}}[X/B_1]$ since $X$ is not attainable. In summary, therefore, we have the following important result.

**(1.23)** The contingent claim $X$ is attainable if and only if $E_Q[X/B_1]$ takes the same value for every $Q \in \mathbb{M}$.

Notice that if $\mathbb{M}$ is a singleton and $X$ is an arbitrary contingent claim, then trivially $E_Q[X/B_1]$ takes the same value for all $Q \in \mathbb{M}$, in which case $X$ must be attainable and the model must be complete. On the other hand, suppose every contingent claim $X$ is attainable but $\mathbb{M}$ contains two distinct risk neutral probability measures, say $Q$ and $\hat{Q}$. In this case there must exist some state $\omega_k$ with $Q(\omega_k) \neq \hat{Q}(\omega_k)$, so take the contingent claim $X$ defined by

$$X(\omega) = \begin{cases} B_1(\omega_k), & \omega = \omega_k \\ 0, & \text{otherwise} \end{cases}$$

Then

$$E_Q[X/B_1] = Q(\omega_k) \neq \hat{Q}(\omega_k) = E_{\hat{Q}}[X/B_1]$$

But this contradicts (1.23), which says that if $X$ is attainable, then $E_Q[X/B_1]$ takes the same value for all $Q \in \mathbb{M}$. Hence if the model is complete, then $\mathbb{M}$ cannot have more than one element. We can combine these observations as follows.

**(1.24)** The model is complete if and only if $\mathbb{M}$ consists of exactly one risk neutral probability measure.

To summarize matters, if the model is complete then we know how to price all the contingent claims. Moreover, if the model is not complete then we know how to price some of the contingent claims, namely, all the attainable ones. But what about the claims that are not attainable in an incomplete model? For such a claim we cannot pinpoint its time $t = 0$ price, but it turns out that at least we can identify an interval within which a fair, reasonable value for the time $t = 0$ price must fall.

For the rest of this section we shall be considering an incomplete model and we shall focus on an arbitrary contingent claim $X$ that is not attainable. Consider the quantity

$$V_+(X) \equiv \inf\{E_Q[Y/B_1] : Y \geqslant X, Y \text{ is attainable}\}$$

and refer to figure 1.3 throughout this discussion. The choice of $Q \in \mathbb{M}$ here does not really matter since it is only being used to compute the price of attainable contingent claims. Note that $\lambda B_1$ is an attainable contingent claim for all values of the scalar $\lambda$ and that $\lambda B_1 \geqslant X$ for all large enough values of $\lambda$; hence $V_+(X)$ is well defined and finite. Notice also that $V_+(X)$ is bounded below by $\sup\{E_Q[X/B_1] : Q \in \mathbb{M}\}$.

The quantity $V_+(X)$ is important because it is a good upper bound on the fair price of $X$. This follows from an arbitrage argument that is similar to the one discussed in the preceding section. If $X$ could be sold for a greater amount, say $p > V_+(X)$, then one should make use of the trading strategy that replicates $Y$, which is any attainable contingent claim satisfying $Y \geqslant X$ and $p > E_Q[Y/B_1] \geqslant V_+(X)$. In particular, one should sell $X$ at time $t = 0$, use part of the proceeds to purchase for the amount $E_Q[Y/B_1]$ the portfolio which replicates $Y$, and pocket the difference $p - E_Q[Y/B_1]$ as a riskless profit. At time $t = 1$ the value of the portfolio $Y$ will always be enough to cover the obligation $X$ of the contingent claim. Hence $V_+(X)$ is the price of the cheapest portfolio that can be used to hedge a short position in the contingent claim $X$.

The unattainable contingent claim $X$ cannot trade at a price higher than $V_+(X)$, or else there will exist an arbitrage opportunity. Similarly, this contingent claim cannot trade at a price lower than $V_-(X)$, where



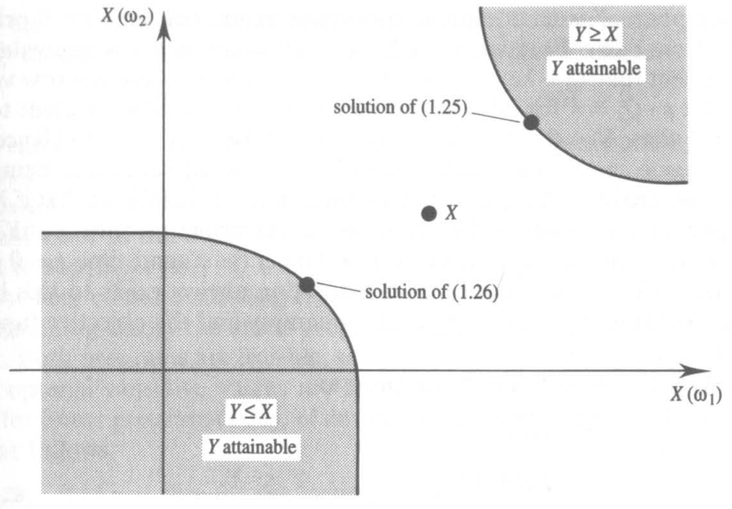


**Figure 1.3** Determining fair prices for unattainable contingent claim X

$$V_-(X) \equiv \sup\{E_Q[Y/B_1] : Y \leqslant X, Y \text{ is attainable}\}$$

As with $V_+(X)$, the quantity $V_-(X)$ is well defined and finite with $V_-(X) \leqslant \inf\{E_Q[X/B_1] : Q \in \mathbb{M}\}$. The fair price (or prices) of $X$ must be in the interval $[V_-(X), V_+(X)]$. We therefore are interested in computing $V_+(X)$ as well as any attainable contingent claim $Y \geqslant X$ satisfying $V_+(X) = E_Q[Y/B_1]$, and similarly for $V_-(X)$.

Consider the linear program

**(1.25)**

$$\begin{aligned} &\text{minimize to} \quad \lambda \\ &\text{subject to} \quad Y \geqslant X \\ &\qquad U - Y/B_1 = 0 \\ &\qquad \lambda - U \cdot Q_1 = 0 \\ &\qquad \vdots \qquad \vdots \\ &\qquad \lambda - U \cdot Q_J = 0 \\ &\lambda \in \mathbb{R}, \quad Y \in \mathbb{R}^K, \quad U \in \mathbb{R}^K \end{aligned}$$

Here $Q_j \in \mathbb{M} = \mathbb{W}^\perp \cap \mathbb{P}^+, j = 1 \ldots, J$, are chosen to be independent vectors, thereby forming a basis of $\mathbb{W}^\perp$, which is assumed to have dimension $J$. This means that the subspace $\mathbb{W}$ of discounted gains has dimension $K - J$ and can be expressed as

$$\mathbb{W} = \{X \in \mathbb{R}^K : X \cdot Q_j = 0 \quad \text{for} \quad j = 1, \cdots, J\}$$

Now suppose $Y$ is an attainable contingent claim with time $t=0$ price $\lambda$, and set $U = Y/B_1$. Because $V_1^* = V_0^* + G^*$, this statement is equivalent to the statement that $U - \lambda e \in \mathbb{W}$ and $U = Y/B_1$, where $e$ here is a row vector of 1's. But $e \cdot Q_j = 1$ for all $j$, so this statement, in turn, is equivalent to the statement that $U = Y/B_1$ and $U \cdot Q_j - \lambda = 0$ for $j = 1, \cdots J$. Hence the feasible region in the linear program (1.25) can be interpreted as being the set of all attainable contingent claims $Y$ with $Y \geqslant X$. It follows that if $\lambda$ and $Y$ are part of an optimal solution of this linear program, then $V_+(X) = \lambda$ and $Y$ is an attainable contingent claim with $Y \geqslant X$ and time $t=0$ price equal to $V_+(X)$. Note that an optimal solution always exists to this linear program because the feasible region is nonempty and the objective function is bounded below.

Similarly, if you solve the linear program

$$
\text{(1.26)} \qquad \begin{aligned}
&\text{maximize} && \lambda \\
&\text{subject to} && Y \leqslant X \\
& && U - Y/B_1 = 0 \\
& && \lambda - U \cdot Q_1 = 0 \\
& && \vdots \qquad \vdots \\
& && \lambda - U \cdot Q_J = 0 \\
& && \lambda \in \mathbb{R}, \quad Y \in \mathbb{R}^K, \quad U \in \mathbb{R}^K
\end{aligned}
$$

and obtain an optimal solution $(\lambda, Y, U)$, then $V_-(X)$ and $Y$ is an attainable contingent claim with $Y \leqslant X$ and time $t=0$ price equal to $V_-(X)$.

It turns out that not only does linear programming enable us to completely solve for the quantities of interest, but it gives us something extra as a bonus. Consider another linear program:

$$
\text{(1.27)} \qquad \begin{aligned}
&\text{maximize} && \sum_{k=1}^{K} [X(\omega_k)/B_1(\omega_k)]\psi_k \\
&\text{subjet to} && \theta_1 + \cdots + \theta_J = 1 \\
& && \psi_1 - Q_1(\omega_1)\theta_1 - \cdots - Q_J(\omega_1)\theta_J = 0 \\
& && \vdots \qquad \vdots \qquad \qquad \vdots \qquad \vdots \\
& && \psi_K - Q_1(\omega_K)\theta_1 - \cdots - Q_J(\omega_K)\theta_J = 0 \\
& && \psi \in \mathbb{R}^K \qquad \theta \in \mathbb{R}^J \qquad \psi \geqslant 0
\end{aligned}
$$

If $(\psi, \theta)$ is an arbitrary feasible solution, then

$$\psi = \theta_1 Q_1 + \theta_2 Q_2 + \cdots + \theta_J Q_J,$$

which is non-negative. Moreover, with $e$ a row vector of 1's we have

$$e \cdot \psi = \theta_1 e \cdot Q_1 + \cdots + \theta_J e \cdot Q_J = \theta_1 + \cdots + \theta_J = 1$$



since each $Q_j$ is a probability measure, so $\psi$ can be interpreted as a probability measure too. For any discounted price process $S_n^*$ we have

$$E_\psi[\Delta S_n^*] = \theta_1 Q_1 \cdot \Delta S_n^* + \cdots + \theta_J Q_J \cdot \Delta S_n^* = 0$$

since each $Q_j \in \mathbb{M}$. Hence $\psi$ can be interpreted as a linear pricing measure (but not, necessarily, as a risk neutral probability measure). In other words the feasible region can be interpreted as the closure of $\mathbb{M}$. It follows that the optimal value of the objective function in linear program (1.27) is precisely equal to $\sup\{E_Q[X/B_1] : Q \in \mathbb{M}\}$.

Now here comes a startling result of fundamental importance. By linear programming theory the linear programs (1.25) and (1.27) are duals of each other. Both programs are feasible, so by linear programming duality theory, their optimal objective values are equal to each other. Analogous results hold for linear program (1.26), of course, so all these results can be summarized as follows:

**(1.28)** If $\mathbb{M} \neq \emptyset$, then for any contingent claim $X$ one has

$$V_+(X) = \sup\{E_Q[X/B_1] : Q \in \mathbb{M}\} \quad \text{and}$$
$$V_-(X) = \inf\{E_Q[X/B_1] : Q \in \mathbb{M}\}.$$

Of course, if $X$ is attainable, then $V_+(X) = V_-(X)$ is its usual time $t=0$ price.

**Example 1.2 (continued)** Consider the contingent claim $X = (30, 20, 10)$. This is not attainable, because it does not satisfy equation (1.21). Recalling that $\mathbb{M}$ consists of all probability measures of the form $Q = (q, 2-3q, -1+2q)$ where $1/2 < q < 2/3$, it is straightforward to compute (making a slight and obvious change of notation for this particular example) $E_q[X/B_1] = 27 - 9q$. Hence

$$V_+(X) = \sup_q E_q[X/B_1] = \sup_q \{27 - 9q\} = 27 - 9(1/2) = 22\ 1/2$$

and

$$V_-(X) = \inf_q E_q[X/B_1] = \inf_q \{27 - 9q\} = 27 - 9(2/3) = 21$$

Upon solving the linear program (1.25) one obtains the attainable contingent claim corresponding to $V_+(X)$; this is $Y = (30, 20, 15)$, as can be verified by checking equation (1.21) and checking that the time $t=0$ price of $Y$ is indeed 22 1/2. Similarly, the attainable contingent claim corresponding to $V_-(X)$ is verified to be $Y = (30, 50/3, 10)$.

**Exercise 1.14** Explain why the model in example 1.4 is not complete. Characterize the set of all the attainable contingent claims. Compute $V_+(X)$ and $V_-(X)$ for $X = (40, 30, 20, 10)$.

**Exercise 1.15** Use (1.23) to verify whether there are any values of the exercise price $e$ such that the call option is attainable for the model in example 1.2. Similarly, specify which put options are attainable. Assume $r = 1/9$.

**Exercise 1.16** Just after linear program (1.25) it was asserted that one can choose $Q_j \in \mathbb{M} = \mathbb{W}^\perp \cap \mathbb{P}^+$, $j = 1, \cdots, J$, to be independent vectors, thereby forming a basis of $\mathbb{W}^\perp$, which is assumed to have dimension $J$. Use linear algebra to carefully verify this assertion. Compute $Q_j$ vectors for example 1.4.

## 1.6 Risk and Return

With $Q$ a risk neutral probability measure and $\omega \in \Omega$, recall that $Q(\omega)/B_1(\omega)$ is sometimes called the state price of $\omega$. For this reason, the random variable

$$L(\omega) \equiv \frac{Q(\omega)}{P(\omega)}$$

is called the *state price vector* or the *state price density*. The main result to be shown in this section is that the risk premium of an arbitrary portfolio is proportional to the covariance[6] between a return corresponding to the state price density and the return for the portfolio, a result that resembles a principal finding of the capital asset pricing model.

Assuming the time $t = 0$ price $S_n(0)$ is strictly positive, the *return* $R_n$ for risky security $n$ is defined to be the random variable

$$R_n \equiv \frac{S_n(1) - S_n(0)}{S_n(0)}, \qquad n = 1, \cdots, N$$

Similarly, the return corresponding to the bank account is defined by

$$R_0 \equiv \frac{B_1 - B_0}{B_0} = r$$

The returns are useful quantities for a variety of purposes, one of which is that if you know the time $t = 0$ prices and the returns, then you can compute time $t = 1$ prices. Since prices are non-negative one has $R_n \geqslant -1$, with equality if and only if $S_n(1) = 0$. It is left as an exercise to verify that the gain for a portfolio can be written as

(1.29) $$G = H_0 B_0 R_0 + \sum_{n=1}^{N} H_n S_n(0) R_n$$

Hence the gain for a portfolio is a weighted combination of the underlying returns, each weight being the amount of money invested at time $t = 0$ in the corresponding security.



The returns can also be used to compute risk neutral probability measures. Since

$$\begin{aligned} S_n^*(1) - S_n^*(0) &= \frac{S_n(1) - B_1 S_n(0)}{B_1} \\ &= \frac{[1 + R_n]S_n(0) - [1 + R_0]S_n(0)}{1 + R_0} \\ &= S_n(0)\left(\frac{R_n - R_0}{1 + R_0}\right) \end{aligned}$$

it follows from (1.15) that

**(1.30)** If $Q$ is a probability measure with $Q(\omega) > 0$ for all $\omega \in \Omega$, then $Q$ is a risk neutral probability measure if and only if

$$E_Q\left(\frac{R_n - R_0}{1 + R_0}\right) = 0, \qquad n = 1, \ldots, N$$

Notice that when the interest rate $R_0 = r$ is deterministic, the equation in (1.30) becomes simply

$$E_Q[R_n] = r, \qquad n = 1, \ldots, N$$

This is one example of many situations where, under the assumption of a deterministic interest rate, one has a nice, and often important, relationship involving returns. Therefore, this assumption will be in force for the balance of this section, as will be the assumption that there exists a risk neutral probability measure $Q$.

The mean return for security $n$, denoted $\bar{R}_n = E[R_n]$, often plays an important role. For example, it is easy to see that $\mathrm{cov}(R_n, L) = E[R_n L] - E[R_n]E[L] = E_Q[R_n] - E[R_n] = r - \bar{R}_n$. In other words,

(1.31) $$\bar{R}_n - r = -\mathrm{cov}(R_n, L), \qquad n = 1, \cdots, N$$

The difference $\bar{R}_n - r$ here is called the *risk premium* for the security; normally this is positive because investors usually insist that the expected returns of risky securities be higher than the riskless return $r$. Thus (1.31) says that the risk premium of a security is related to the correlation[7] between the security's return and the state price density.

Consider the return $R$ of a portfolio corresponding to an arbitrary trading strategy $H = (H_0, H_1, \ldots, H_N)$. Assuming $V_0 > 0$, this is

$$R = \frac{V_1 - V_0}{V_0}$$

Using $S_n(1) = S_n(0)[1 + R_n]$ and the definition of $V_t$ one obtains

(1.32) $$R = \frac{H_0}{V_0} r + \sum_{n=1}^{N} \left[\frac{H_n S_n(0)}{V_0}\right] R_n$$

If you interpret $H_o/V_o$ as the fraction of money invested in the savings account (recall $B_0 = 1$) and $H_nS_n(0)/V_0$ as the fraction of money invested in the $n$th security, then (1.32) says that the return on the portfolio is a convex combination of the returns of the individual securities. Using (1.31), (1.32), and some basic properties of the covariance, it is straightforward to verify that

$$\bar{R} - r = -\mathrm{cov}(R, L) \tag{1.33}$$

where, of course, $\bar{R} = E[R]$.

Now fix two scalars $a$ and $b$ with $b \neq 0$, and assume the contingent claim $a + bL$ is attainable, that is, suppose there exists some trading strategy $H'$ such that $V_1' = a + bL$. Since $V_0'(1 + R') = a + bL$ (here $V'$ and $R'$ denote the value and return processes, respectively, corresponding to $H'$), one can substitute for $L$ and use the properties of the covariance relationship to verify that

$$\mathrm{cov}(R, L) = \frac{V_0'}{b}\,\mathrm{cov}(R, R')$$

($R$ still corresponds to an arbitrary trading strategy). Hence (1.33) can be rewritten as

$$\bar{R} - r = -\frac{V_0'}{b}\,\mathrm{cov}(R, R') \tag{1.34}$$

In particular, in the special case where you choose $H = H'$, (1.34) says that

$$\bar{R}' - r = -\frac{V_0'}{b}\,\mathrm{cov}(R', R') = -\frac{V_0'}{b}\,\mathrm{var}(R')$$

Using this to substitute for $V_0'/b$ in (1.34), where now we are back to an arbitrary trading strategy $H$, we obtain the following:

**(1.35)** Suppose for scalars $a$ and $b$ the contingent claim $a + bL$ is generated by some portfolio having return $R'$ and suppose the interest rate $r$ is deterministic. Let $R$ be the return of an arbitrary portfolio. Then

$$\bar{R} - r = \frac{\mathrm{cov}(R, R')}{\mathrm{var}(R')}(\bar{R}' - r)$$

The ratio $\mathrm{cov}(R, R')/\mathrm{var}(R')$ is called the *beta* of the trading strategy $H$ with respect to the trading strategy $H'$. This result says that the risk premium of $H$ is proportional to the risk premium of $H'$, with the proportionality constant being this beta. Or from a slightly different perspective, (1.35) says that the risk premium is proportional to its beta with respect to a linear transformation of the state price density. This result resembles the traditional capital asset pricing model, only here $H'$ corresponds to a linear transformation of the state price density instead of the market portfolio.

Notice that with a deterministic interest rate $r$ and with arbitrary scalars $a$ and $b$ ($b \neq 0$), the contingent claim $a + bL$ is attainable if and only if the



state price density $L$ is. This is because $H_0(1 + r) + \sum H_nS_n(1) = a + bL$ if and only if

$$\frac{1}{b}\left[H_0 - \frac{a}{1 + r}\right](1 + r) + \sum_{n=1}^{N} \frac{1}{b} H_nS_n(1) = L$$

**Exercise 1.17** Verify equation (1.29), both in general and by applying it to example 1.1.

**Exercise 1.18** Assuming the time $t = 0$ price is strictly positive, the *discounted return* $R_n^*$ is defined by $R_n^* \equiv [S_n^*(1) - S_n^*(0)]/S_n^*(0)$ for $n = 1, \ldots, N$. Show that

(a) $G^* = \sum_{n=1}^{N} H_nS_n^*(0)R_n^*$

(b) $R_n^* = \dfrac{R_n - R_0}{1 + R_0} \quad n = 1, \ldots, N$

(c) The strictly positive probability measure $Q$ is a risk neutral probability measure if and only if $E_Q[R_n^*] = 0$ for $n = 1, \ldots, N$.

**Exercise 1.19** Analyze the risk and return properties of example 1.1 assuming $P(\omega_1) = p$ for a general parameter $0 < p < 1$.

(a) What are $R_1$ and $\bar{R}_1$?
(b) What is $L$?
(c) Verify (1.31) for $n = 0$ and 1.

From now on suppose $H = (H_0, H_1) = (1, 3)$.

(d) What are $R$ and $\bar{R}$?
(e) Verify (1.32).
(f) Verify (1.33).
(g) What are $H'$, $V_0'$, and $R'$?
(h) Verify (1.34).
(i) Verify (1.35).

## NOTES

1 If $X$ is a *random variable*, this means $X$ is a real-valued function on the sample space $\Omega$. In other words, we know the value $X(\omega)$ for each state of the world $\omega \in \Omega$.

2 The *expected value* (also called the *mean* or *average*) of the random variable $X$ is denoted $EX$ or $E[X]$ and defined to be

$$EX \equiv \sum_{k=1}^{K} X(\omega_k)P(\omega_k)$$

More generally, if $f$ is a real-valued function on the real line,

$$Ef(X) \equiv \sum_{k=1}^{K} f\big(X(\omega_k)\big)P(\omega_k)$$

In particular, for scalars $a$ and $b$, $E[aX+b] = aEX + b$.

3 See the appendix on linear programming.

4 In other words, for some $\hat{Y} \in \mathbb{W}^\perp$ whose components are all strictly positive, the ray is of the form $\{Y \in \mathbb{R}^N : Y = \lambda\hat{Y}, \lambda > 0, \lambda \in \mathbb{R}\}$.

5 *Closed* means that if $\{X_l\}$ is a sequence of points in $\mathbb{A}^+$ that converges to some $X \in \mathbb{R}^K$, then $X \in \mathbb{A}^+$; *convex* means that for any $X, \hat{X} \in \mathbb{A}^+$ and any scalar $\lambda$ with $0 < \lambda < 1$, then $\lambda X + (1-\lambda)\hat{X} \in \mathbb{A}^+$.

6 For two random variables $X$ and $Y$, the *covariance* $\text{cov}(X, Y)$ is defined to be $E[XY] - E[X]E[Y]$. Note that $\text{cov}(X - E[X], Y) = \text{cov}(X, Y)$. Moreover, given three random variables $X$, $Y$, and $Z$ and two scalars $a$ and $b$, one has $\text{cov}(aX + bZ, Y) = a\,\text{cov}(X, Y) + b\,\text{cov}(Z, Y)$.

7 The *variance*, denoted $\text{var}(X)$, of a random variable $X$ is defined by $\text{var}(X) \equiv E[X^2] - (E[X])^2 = E[(X - E[X])^2]$. The *standard deviation* of $X$ is $\sigma_X \equiv \sqrt{\text{var}(X)}$. The *correlation* between the random variables $X$ and $Y$ (assuming $\sigma_X > 0$ and $\sigma_Y > 0$) is defined by $\rho(X, Y) \equiv \text{cov}(X, Y)/(\sigma_X\sigma_Y)$. Hence the risk premium for security $n$ equals $-\rho(R_n, L)\sigma_{R_n}\sigma_L$.

# 2 Single Period Consumption and Investment



## 2.1 Optimal Portfolios and Viability

This chapter is concerned with the problem of choosing the best trading strategy for the purpose of transforming wealth invested at time $t = 0$ into time $t = 1$ wealth. With some variations of this problem that will be considered in later sections, a portion of the wealth is consumed at time $t = 0$. The problem is to compute an optimal trading strategy, and for this a measure of performance is needed.

The measure of performance that will be used here is that of expected utility. In particular, suppose $u : \mathbb{R} \times \Omega \to \mathbb{R}$ is a function such that $w \to u(w, \Omega)$ is differentiable, concave, and strictly increasing for each $\omega \in \Omega$. If $w$ is the value of the portfolio at time $t = 1$ and $\omega$ is the state, then $u(w, \omega)$ will represent the *utility* of the amount $w$. Hence our measure of performance will be the expected utility of terminal wealth, that is,

$$Eu(V_1) = \sum_{\omega\in\Omega} P(\omega)u(V_1(\omega), \omega)$$

Note that the utility function $u$ can depend explicitly on both the terminal wealth $w$ and the state $\omega$. However, for many applications it suffices for $u$ to depend only on the wealth, in which case $u$ is simply a concave, strictly increasing function with a single argument.

Let $\mathbb{H}$ denote the set of all trading strategies, that is, $\mathbb{H} = \mathbb{R}^{N+1}$, the linear space of all vectors of the form $(H_0, H_1, \ldots, H_N)$. Let $v \in \mathbb{R}$ be a specified scalar representing the initial, time $t = 0$ wealth. We are interested in the following optimal portfolio problem:

(2.1)
$$\begin{array}{ll} \underset{H \in \mathbb{H}}{\text{maximize}} & Eu(V_1) \\ \text{subject to} & V_0 = v \end{array}$$

Since $V_1 = B_1 V_1^*$ and $V_1^* = V_0^* + G^*$, this is the same as

(2.2)
$$\text{maximize} \quad E[u(B_1\{v + H_1 \Delta S_1^* + \ldots + H_N \Delta S_N^*\})]$$

Notice that if there exists an arbitrage opportunity, then there cannot exist a solution to (2.1). In other words, if $\hat{H}$ is a solution and $H$ is an arbitrage opportunity, then setting $\tilde{H} = \hat{H} + H$ gives

$$v + \sum_{n=1}^{N} \tilde{H}_n \Delta S_n^* = v + \sum_{n=1}^{N} \hat{H}_n \Delta S_n^* + \sum_{n=1}^{N} H_n \Delta S_n^* \geqslant v + \sum_{n=1}^{N} \hat{H}_n \Delta S_n^*$$

where the inequality follows because $H$ is an arbitrage opportunity. In fact, this inequality is actually strict for at least one $\omega \in \Omega$. Since $u$ is strictly increasing in wealth and since $P(\omega) > 0$ for all $\omega \in \Omega$, this means the objective value in (2.2) is strictly greater under $\tilde{H}$ than under $\hat{H}$. This contradicts the assertion that $\hat{H}$ is an optimal solution of (2.2), in which case the following must be true:

**(2.3)** If there exists an optimal solution of the portfolio problem (2.1) or (2.2), then there are no arbitrage opportunities.

In other words, (2.3) says that if there exists an optimal solution to (2.1) or (2.2), then there exists a risk neutral probability measure. By a result that is somewhat surprising, there exists an explicit relationship between any such solution and the risk neutral probability measures. This relationship can be derived from the first order conditions necessary for optimality. To see this, rewrite the objective function in (2.2) as

$$\sum_{\omega \in \Omega} P(\omega) u(B_1(\omega)\{v + H_1 \Delta S_1^*(\omega) + \ldots + H_N \Delta S_N^*(\omega)\}, \omega)$$

so that the first order necessary condition can be expressed as

(2.4)
$$\begin{aligned} 0 &= \frac{\partial E[u(B_1\{v + H_1 \Delta S_1^* + \ldots + H_N \Delta S_N^*\})]}{\partial H_n} \\ &= \sum_{\omega \in \Omega} P(\omega) u'(B_1(\omega)\{v + H_1 \Delta S_1^*(\omega) + \ldots + H_N \Delta S_N^*(\omega)\}, \omega) B_1(\omega) \Delta S_n^*(\omega) \\ &= E[B_1 u'(V_1) \Delta S_n^*], \qquad n = 1, \ldots, N \end{aligned}$$

where $u'$ denotes the partial derivative of $u$ with respect to the first argument. Hence if $(H, V)$ is a solution of (2.2), then it must satisfy this system of $N$ equations. But recall the condition which a risk neutral probability measure must satisfy:

(2.5)
$$0 = E_Q[\Delta S_n^*] = \sum_{\omega \in \Omega} Q(\omega) \Delta S_n^*(\omega), \qquad n = 1, \ldots, N$$



Comparing (2.4) and (2.5) it is apparent that upon setting $Q(\omega) = P(\omega)B_1(\omega)u'(V_1(\omega), \omega)$ one has obtained a measure satisfying (2.5). Note that $Q(\omega) > 0$ for all $\omega$ since $u$ is strictly increasing. However, $Q(\omega_1) + \ldots + Q(\omega_K)$ is not necessarily equal to one, so $Q$ is only a probability measure up to a constant. It is easy to see what this constant should be, and so we have the following:

**(2.6)** If $(H, V)$ is a solution of the optimal portfolio problem (2.1) or (2.2), then

$$Q(\omega) \equiv \frac{P(\omega) B_1(\omega) u'(V_1(\omega), \omega)}{E[B_1 u'(V_1)]}, \qquad \omega \in \Omega$$

defines a risk neutral probability measure.

Rewriting (2.6) slightly in the case where $B_1 = 1 + r$ is constant, we obtain $L(\omega) = Q(\omega)/P(\omega) = u'(V_1(\omega), \omega)/E[u'(V_1)]$. In other words, when the interest rate is deterministic, the *state price density* is proportional to the *marginal utility of terminal wealth.*

What about the converse? If there exists a risk neutral probability measure $Q$, then does the optimal portfolio problem (2.1) have a solution? Not necessarily, for some $u$ and $v$ may be such that no solution $H$ exists. However, one can always find some $u$ and $v$ such that a solution $H$ does exist. Formalizing this idea, we will say that the model is *viable* if there exists a function $u : \mathbb{R} \times \Omega \to \mathbb{R}$ and an initial wealth $v$ such that $w \to u(w, \omega)$ is concave and strictly increasing for each $\omega \in \Omega$ and such that the corresponding optimal portfolio problem (2.1) has an optimal solution $H$.

**(2.7)** The securities market model is viable if and only if there exists a risk neutral probability measure $Q$.

In view of (2.6), to verify this principle it suffices to assume the existence of a risk neutral probability measure, cleverly select $u$ and $v$, and then demonstrate the existence of a solution of (2.2). The choice of $u$ will be

$$u(w, \omega) = w \frac{Q(\omega)}{P(\omega) B_1(\omega)}$$

while $v$ will be arbitrary. Now for an arbitrary $(H_1, \ldots, H_N)$ we have

$$\begin{aligned} E[u(B_1\{v &+ H_1 \Delta S_1^* + \ldots + H_N \Delta S_N^*\}, \omega)] \\ &= \sum P(\omega) B_1(\omega)\{v + H_1 \Delta S_1^* + \ldots + H_N \Delta S_N^*\} Q(\omega)/[P(\omega) B_1(\omega)] \\ &= \sum Q(\omega)\{v + H_1 \Delta S_1^* + \ldots + H_N \Delta S_N^*\} \\ &= v + H_1 E_Q[\Delta S_1^*] + \ldots + H_N E_Q[\Delta S_N^*] = v \end{aligned}$$

so every vector $(H_1, \ldots, H_N)$ gives rise to the same objective value in (2.2). Equivalently, every trading strategy with initial wealth $v$ gives rise to the same objective value in (2.1), which means that all such trading strategies are optimal. Hence (2.7) is true by this clever choice of utility function.

The optimal portfolio problem (2.1) or (2.2) is a standard convex optimization problem, so one can use standard techniques to compute a solution. One such approach is to work with the necessary equations (2.4), a system of $N$ equations and $N$ unknowns. Unfortunately, as seen in the following example, these equations can be nonlinear in $H$ and thus difficult to solve.

**Example 2.1** Suppose $N = 2$, $K = 3$, $r = 1/9$, and the discounted price process is as follows:

| $n$ | $S_n^*(0)$ | $S_n^*(1)$ | | |
|---|---|---|---|---|
| | | $\omega_1$ | $\omega_2$ | $\omega_3$ |
| 1 | 6 | 6 | 8 | 4 |
| 2 | 10 | 13 | 9 | 8 |

Note that there exists a unique risk neutral probability measure, because $Q = (1/3, 1/3, 1/3)$ is the unique solution of the following system of equations:

$$\begin{aligned} 6 &= 6Q(\omega_1) + 8Q(\omega_2) + 4Q(\omega_3) \\ 10 &= 13Q(\omega_1) + 9Q(\omega_2) + 8Q(\omega_3) \\ 1 &= Q(\omega_1) + Q(\omega_2) + Q(\omega_3) \end{aligned}$$

With the exponential utility function $u(w) = -\exp\{-w\}$, the marginal utility function is $u'(w) = \exp\{-w\}$. Hence the necessary conditions (2.4) are:

$$\begin{aligned} 0 = P(\omega_1) &\exp\{-(10/9)(v + 0H_1 + 3H_2)\}(10/9)(0) \\ &+ P(\omega_2)\exp\{-(10/9)(v + 2H_1 - H_2)\}(10/9)(2) \\ &\quad + P(\omega_3)\exp\{-(10/9)(v - 2H_1 - 2H_2)\}(10/9)(-2) \end{aligned}$$

$$\begin{aligned} 0 = P(\omega_1) &\exp\{-(10/9)(v + 0H_1 + 3H_2)\}(10/9)(3) \\ &+ P(\omega_2)\exp\{-(10/9)(v + 2H_1 - H_2)\}(10/9)(-1) \\ &\quad + P(\omega_3)\exp\{-(10/9)(v - 2H_1 - 2H_2)\}(10/9)(-2) \end{aligned}$$

Needless to say, these are not so easy to solve for $H_1$ and $H_2$.

**Exercise 2.1** Suppose $N = 1$, $K = 2$, $S_0 = 5$, $S_1(\omega_1) = 20/3$, and $S_1(\omega_2) = 40/9$. Solve (2.1) in the case of $r = 1/9$ and general scalar parameters for the initial wealth $v \geqslant 0$ and the probability $P(\omega_1) = p$ under the utility functions

(a) $u(w) = \ln w$
(b) $u(w) = -\exp(-w)$
(c) $u(w) = \gamma^{-1} w^{\gamma}$, where $-\infty < \gamma < 1$, $\gamma \neq 0$.



## 2.2 Risk Neutral Computational Approach

As seen in example 2.1, solving the optimal portfolio problem (2.1) can be computationally difficult. Fortunately, there is an alternative technique which involves the risk neutral probability measure and is much more efficient. The idea is based on the observation that the objective function $H \to Eu(V_1)$ in (2.1) can be viewed as the composition of two functions, as illustrated in figure 2.1. The first function $H \to V_1$ maps trading strategies into random variables which represent the time $t = 1$ value of the portfolio. The second function $V_1 \to Eu(V_1)$ maps random variables into numbers on the real line. Corresponding to this composition, the risk neutral computation technique involves a two-step process. First you identify the optimal random variable $V_1$, that is, the value of $V_1$ maximizing $Eu(V_1)$ over the subset of feasible random variables. Then you compute the trading strategy $H$ that generates this $V_1$, that is, you solve for the trading strategy that replicates the contingent claim $V_1$.

Step 2 is easy. This is exactly the same as was discussed in section 1.4 for computing the trading strategy which replicates an attainable contingent claim. If the subset of feasible random variables were chosen correctly for step 1, then the computed trading strategy which replicates $V_1$ will correspond to a portfolio having time $t = 0$ value equal to $v$. In other words, the attainable contingent claim $V_1$ will have time $t = 0$ price equal to $v$, the specified initial value of the portfolio.

Step 1 is a bit more challenging, but it just involves straightforward optimization theory. The key to success is to specify the subset of feasible random variables correctly and conveniently. If the model is complete, this subset is simply

**(2.8)** $$\mathbb{W}_v \equiv \{W \in \mathbb{R}^K : E_Q[W/B_1] = v\}$$

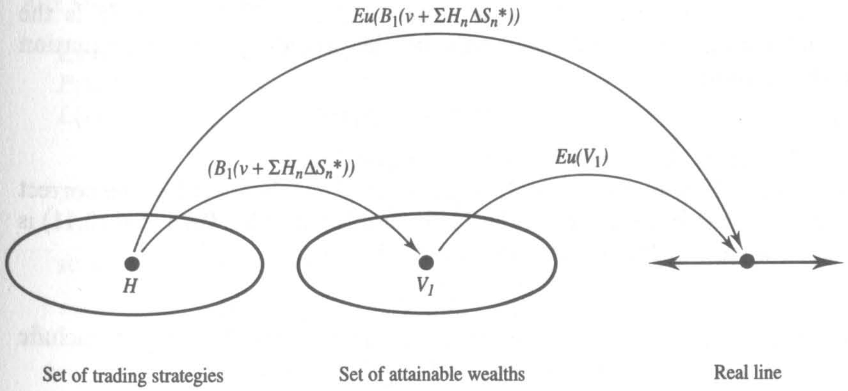


**Figure 2.1** The risk neutral computational approach

(the specification of $\mathbb{W}_v$ for incomplete models is more complex and will be discussed below). To see this, note that under any trading strategy $H$ with $V_0 = v$ one has $E_Q[V_1/B_1] = v$ by the risk neutral valuation principle. Conversely, for any contingent claim $W \in \mathbb{W}_v$ there exists, again by the risk neutral valuation principle, a trading strategy $H$ such that $V_0 = v$ and $V_1 = W$. In the context of optimal portfolio problems, the subset $\mathbb{W}_v$ (actually, an affine subspace) is called the *set of attainable wealths.*

The first step in the risk neutral computation technique is to solve the subproblem:

(2.9) $$\begin{array}{ll} \text{maximize} & Eu(W) \\ \text{subject to} & W \in \mathbb{W}_v \end{array}$$

When the model is complete, this problem can be conveniently solved with a Lagrange multiplier. In view of (2.8), problem (2.9) is equivalent to

(2.10) $$\text{maximize} \quad Eu(W) - \lambda E_Q[W/B_1]$$

where the Lagrange multiplier $\lambda$ is chosen so that the solution in (2.10) satisfies

(2.11) $$E_Q[W/B_1] = v$$

Introducing the state price density $L = Q/P$, the objective function in (2.10) can be rewritten as

$$\begin{aligned} Eu(W) - \lambda E[LW/B_1] &= E[u(W) - \lambda LW/B_1] \\ &= \sum_\omega P(\omega)[u(W(\omega)) - \lambda L(\omega)W(\omega)/B_1(\omega)] \end{aligned}$$

If $W$ maximizes this expression, then the necessary conditions must be satisfied, giving rise to one equation for each $\omega \in \Omega$:

$$u'(W(\omega)) = \lambda L(\omega)/B_1(\omega), \qquad \text{all } \omega \in \Omega$$

Note that this equation is exactly the same as the one in (2.6); in fact, since $W = V_1$ one can deduce that $\lambda$ is equal to $E[B_1 u'(W)]$, where $W$ is the optimal solution. To compute $W$ we solve the preceding displayed equation for $W(\omega)$ giving

(2.12) $$W(\omega) = \mathrm{I}\,[\lambda L(\omega)/B_1(\omega)]$$

where I denotes the *inverse function* corresponding to $u'$.

Hence (2.12) gives the optimal solution of (2.9) when $\lambda$ takes the correct value. But what is the correct value? It is simply the value such that (2.11) is satisfied when (2.12) is substituted for $W$, that is,

(2.13) $$E_Q[\mathrm{I}\,(\lambda L/B_1)/B_1] = v$$

The inverse function I is decreasing, and its range will normally include $(0, \infty)$, so normally a solution $\lambda$ to (2.11) will exist for $v > 0$.

**Example 2.2** Suppose $u(w) = -\exp(-w)$, so that $u'(w) = \exp(-w)$. Then $u'(w) = i$ if and only if $w = -\ln(i)$, so $I(i) = -\ln(i)$. Hence the optimal solution of (2.9) is of the form



$$W = -\ln(\lambda L/B_1) = -\ln(\lambda) - \ln(L/B_1)$$

and (2.13) becomes

$$v = -E_Q[B_1^{-1}\ln(\lambda L/B_1)] = -\ln(\lambda)E_Q B_1^{-1} - E_Q[\ln(L/B_1)/B_1]$$

Hence the correct value of $\lambda$ is given by

$$\lambda = \exp\left\{\frac{-v - E_Q[B_1^{-1}\ln(L/B_1)]}{E_Q B_1^{-1}}\right\}$$

so

$$W = \frac{v + E_Q[B_1^{-1}\ln(L/B_1)]}{E_Q B_1^{-1}} - \ln(L/B_1)$$

Substituting this into $-\exp(-W)$ gives

$$\begin{aligned} u(W) &= -\exp\left\{\frac{-v + \ln(L/B_1)E_Q B_1^{-1} - E_Q[B_1^{-1}\ln(L/B_1)]}{E_Q B_1^{-1}}\right\} \\ &= -\lambda L/B_1 \end{aligned}$$

so the optimal value of the objective function in (2.9) is

$$Eu(W) = -\lambda E[L/B_1] = -\lambda E_Q B_1^{-1}$$

This example illustrates a general pattern: general formulas for the optimal wealth $W$ and so forth are obtained which depend on the underlying securities market model only via the probability measures $P$ and $Q$. In other words, $P$ and $Q$ comprise what can be thought of as a sufficient statistic for the optimal portfolio subproblem (2.9). After deriving the formulas like those in example 2.2 for a particular utility function, one can quickly analyse any complete securities market model having the same utility function.

**Examples 2.1 and 2.2 (continued)** Suppose $P(\omega_1) = 1/2$ and $P(\omega_2) = P(\omega_3) = 1/4$ so the state price density $L$ is given by $L(\omega_1) = 2/3$ and $L(\omega_2) = L(\omega_3) = 4/3$. With $r = 1/9$ and $B_1 = 10/9$ we compute

$$E_Q[\ln(L/B_1)] = (1/3)\left[\ln\left(\frac{2}{3}\cdot\frac{9}{10}\right) + 2\ln\left(\frac{4}{3}\cdot\frac{9}{10}\right)\right] = -0.04873$$

so the optimal attainable wealth is

$$\begin{aligned} W &= v(1+r) + E_Q[\ln(L/B_1)] - \ln(L/B_1) \\ &= \begin{cases} v(10/9) + 0.46209, & \omega = \omega_1 \\ v(10/9) - 0.23105, & \omega = \omega_2, \omega_3 \end{cases} \end{aligned}$$

Note that $E_Q[W/B_1] = v$, as desired. Now $\lambda = \exp\{-(10/9)v + 0.04873\}$, so the optimal value of the objective function is

$$Eu(W) = -\lambda E_Q B_1^{-1} = -\frac{9}{10}\lambda$$

Notice that, consistent with (2.6), $\lambda = E[B_1 u'(V_1)]$. Also, having computed the optimal attainable wealth $W$, we can now easily compute the optimal trading strategy $H$ by solving $W/B_1 = v + G^*$. In state $\omega_1$ the discounted terminal wealth $W(\omega_1)/B_1$ equals $v + (9/10)(0.46209) = v + 0.41588$, whereas the initial wealth plus the discounted gain $v + G^*(\omega_1)$ equals $v + H_1(6-6) + H_2(13-10) = v + 3H_2$. Similarly, equations are obtained corresponding to states $\omega_2$ and $\omega_3$, yielding the following system of three equations:

$$\begin{aligned} \omega_1 &: \quad 0.41590 = \ \ 0H_1 + 3H_2 \\ \omega_2 &: -0.20795 = \ \ 2H_1 - \ \ H_2 \\ \omega_3 &: -0.20795 = -2H_1 - 2H_2 \end{aligned}$$

These equations are redundant and there exists a unique solution, which is $H_1 = -0.03466$ and $H_2 = 0.13863$. Using $v = V_0 = H_0 + 6H_1 + 10H_2$ to solve for $H_0$ yields $H_0 = v - 1.17834$. Finally, notice that, as desired, this trading strategy satisfies the necessary conditions (2.4) (see example 2.1 above).

---

**Exercise 2.2 (log utility)** Suppose $u(w) = \ln(w)$. Show that the inverse function $I(i) = i^{-1}$, the Lagrange multiplier $\lambda = v^{-1}$, the optimal attainable wealth is $W = vL^{-1}B_1$, and the optimal objective value is $\ln(v) - E[\ln(L/B_1)]$. Compute these expressions and solve for the optimal trading strategy in the case where $N = 1$, $K = 2$, $r = 1/9$, $S_0 = 5$, $S_1(\omega_1) = 20/3$, $S_1(\omega_2) = 40/9$, and $P(\omega_1) = 3/5$.

**Exercise 2.3 (isoelastic utility)** Suppose $u(w) = \gamma^{-1}w^\gamma$, where $-\infty < \gamma < 1$ and $\gamma \neq 0$. Show that the inverse function $I(i) = i^{-1/(1-\gamma)}$, the Lagrange multiplier

$$\lambda = v^{-(1-\gamma)}\{E[(L/B_1)^{-\gamma/(1-\gamma)}]\}^{(1-\gamma)}$$

the optimal attainable wealth

$$W = \frac{v(L/B_1)^{-1/(1-\gamma)}}{E[(L/B_1)^{-\gamma/(1-\gamma)}]}$$

and the optimal objective value $E[u(W)] = \lambda v/\gamma$. Compute these expressions and solve for the optimal trading strategy in the case where the underlying model is as in exercise 2.2.

---

## 2.3 Consumption Investment Problems

A *consumption process* $C = (C_0, C_1)$ consists of a non-negative scalar $C_0$ and a non-negative random variable $C_1$. A *consumption-investment plan* consists of a pair $(C, H)$, where $C$ is a consumption process and $H$ is a trading strategy. A consumption-investment plan is said to be *admissible* if (1) $C_0 + V_0 = v$, the money available at time $t = 0$, and (2) $C_1 = V_1$. We always assume $v \geqslant 0$.



The quantity $C_t$ should be interpreted as the amount consumed by the investor at time $t$. Since $C_0$ equals time zero consumption and since $V_0 = H_0 + \sum H_n S_n(0)$ is the amount invested at time $t = 0$, the amount of money $v$ available at time zero must be at least $C_0 + V_0$. Since $V_1 = H_0 B_1 + \sum H_n S_n(1)$ is the amount of money available at time $t = 1$, it must be that $C_1 \leqslant V_1$. Now a sensible investor who can consume only at times $t = 0$ and $t = 1$ would not leave money 'lying on the table' at either time, so this investor would probably not want to adopt a consumption-investment plan unless it is admissible.

A question that naturally arises is, given a consumption-investment plan $(C, H)$ and an initial amount of funds $v$, how do you check whether $(C, H)$ is admissible? Of course one way is to compute $V_t$ and then check whether both $C_0 + V_0 = v$ and $C_1 = V_1$. Notice that if $(C, H)$ is indeed admissible, then $C_1$ is an attainable contingent claim with

$$E_Q[C_1/B_1] = E_Q[V_1/B_1] = V_0$$

for every risk neutral probability measure $Q$, in which case

**(2.14)** $$E_Q[C_0 + C_1/B_1] = v$$

Now here is a harder question: given some $v \geqslant 0$ and some consumption process $C$, how do you know whether there exists some trading strategy $H$ such that $(C, H)$ is admissible? Well, if $C_1$ is an attainable contingent claim, then there exists some trading strategy $H$ such that $C_1 = V_1 = H_0 B_1 + \sum H_n S_n$. If, moreover, (2.14) is satisfied for some $Q$, then $C_0 + V_0 = v$, in which case $(C, H)$ is admissible. Notice that $E_Q[C_0 + C_1/B_1]$ is constant with respect to all risk neutral probability measures if and only if $C_1$ is attainable. Hence we can summarize all these findings as follows:

**(2.15)** Let the initial amount of money $v \geqslant 0$ and the consumption process $C$ be fixed. There exists a trading strategy $H$ such that the consumption-investment plan $(C, H)$ is admissible if and only if

$$C_0 + E_Q[C_1/B_1] = v$$

for every risk neutral probability measure $Q$.

**Example 2.1 (continued)** This model is complete with $Q = (1/3, 1/3, 1/3)$. In order for the consumption process $(C_0, C_1)$ to be part of an admissible consumption-investment plan we must have, of course, $0 \leqslant C_0 \leqslant v$ and $C_1 \geqslant 0$. In addition, we must have by (2.15)

$$v - C_0 = \frac{9}{10}E_Q C_1 = \frac{3}{10}[C_1(\omega_1) + C_1(\omega_2) + C_1(\omega_3)]$$

Suppose an investor starts with initial wealth $v$ and wants to choose an admissible consumption-investment plan so as to maximize the expected value of the utility of consumption at both times 0 and 1. Here the utility function $u : \mathbb{R}_+ \to \mathbb{R}$ is assumed to be concave, differentiable, and strictly increasing. Mathematically this problem is:

$$\text{(2.16)}\quad \begin{aligned} &\text{maximize} && u(C_0) + E[u(C_1)] \\ &\text{subject to} && C_0 + H_0 B_0 + \sum_{n=1}^{N} H_n S_n(0) = v \\ &&& C_1 - H_0 B_1 - \sum_{n=1}^{N} H_n S_n(1) = 0 \\ &&& C_0 \geqslant 0 \quad C_1 \geqslant 0 \quad H \in \mathbb{R}^{N+1} \end{aligned}$$

As with the optimal portfolio problem, this consumption-investment problem can be solved either with standard optimization theory or with a risk neutral computational approach. To illustrate the former, consider the following:

**Example 2.1 (continued)** Suppose $u(c) = \ln(c)$. Since $\ln(c) \to -\infty$ as $c \searrow 0$, we can drop the explicit non-negativity constraints in (2.16). With $P(\omega_1) = 1/2$ and $P(\omega_2) = P(\omega_3) = 1/4$ and $r = 1/9$, the optimization problem becomes

$$\text{maximize} \quad \ln(C_0) + \tfrac{1}{2}\ln(C_1(\omega_1)) + \tfrac{1}{4}\ln(C_1(\omega_2)) + \tfrac{1}{4}\ln(C_1(\omega_3))$$

$$\text{(2.17)}\quad \begin{aligned} \text{subject to} \quad C_0 &= v - H_0 - 6H_1 - 10H_2 \\ C_1(\omega_1) &= \frac{10}{9}H_0 + \frac{60}{9}H_1 + \frac{130}{9}H_2 \\ C_1(\omega_2) &= \frac{10}{9}H_0 + \frac{80}{9}H_1 + \frac{90}{9}H_2 \\ C_1(\omega_3) &= \frac{10}{9}H_0 + \frac{40}{9}H_1 + \frac{80}{9}H_2 \end{aligned}$$

This simplifies to become

$$\begin{aligned} \text{maximize} \quad & \ln(v - H_0 - 6H_1 - 10H_2) \\ & + \frac{1}{2}\ln\left(\frac{10}{9}H_0 + \frac{60}{9}H_1 + \frac{130}{9}H_2\right) \\ & + \frac{1}{4}\ln\left(\frac{10}{9}H_0 + \frac{80}{9}H_1 + \frac{90}{9}H_2\right) \\ & + \frac{1}{4}\ln\left(\frac{10}{9}H_0 + \frac{40}{9}H_1 + \frac{80}{9}H_2\right) \end{aligned}$$



Computing the partial derivatives with respect to the $H_n$ and then setting them equal to zero, we obtain the necessary conditions:

$$\text{(2.18)}\quad \begin{aligned} \frac{-1}{C_0} + \frac{1}{2}\cdot\frac{10}{9}\cdot\frac{1}{C_1(\omega_1)} + \frac{1}{4}\cdot\frac{10}{9}\cdot\frac{1}{C_1(\omega_2)} + \frac{1}{4}\cdot\frac{10}{9}\cdot\frac{1}{C_1(\omega_3)} &= 0 \\ \frac{-6}{C_0} + \frac{1}{2}\cdot\frac{60}{9}\cdot\frac{1}{C_1(\omega_1)} + \frac{1}{4}\cdot\frac{80}{9}\cdot\frac{1}{C_1(\omega_2)} + \frac{1}{4}\cdot\frac{40}{9}\cdot\frac{1}{C_1(\omega_3)} &= 0 \\ \frac{-10}{C_0} + \frac{1}{2}\cdot\frac{130}{9}\cdot\frac{1}{C_1(\omega_1)} + \frac{1}{4}\cdot\frac{90}{9}\cdot\frac{1}{C_1(\omega_2)} + \frac{1}{4}\cdot\frac{80}{9}\cdot\frac{1}{C_1(\omega_3)} &= 0 \end{aligned}$$

Here the four equations in (2.17) were used in order to be concise, so actually the three equations in (2.18) involve just the three unknowns $H_0$, $H_1$, and $H_2$. Hence, although it is not particularly easy to do so, these three equations can be solved for the optimal trading strategy $H$, and then finally the four equations in (2.17) can be used to obtain the optimal consumption process $C$.

A general consumption investment problem (2.16) can be solved in a manner similar to that just used for example 2.1. The $N+1$ first order necessary conditions are, in general:

$$\text{(2.19)}\quad u'(C_0) = E[B_1 u'(C_1)]$$

$$\text{(2.20)}\quad u'(C_0)S_n(0) = E[u'(C_1)S_n(1)], \qquad n = 1, \ldots, N$$

Using the constraints in (2.16) to substitute for $C_0$ and $C_1$ gives rise to $N+1$ equations which, in principle, can be solved for the $N+1$ unknowns $H_0, H_1, \ldots, H_N$. Substituting these values into the constraints of (2.16) gives the values of $C_0$ and $C_1$. This procedure yields a solution of (2.16) provided $C_0$ and $C_1$ are both non-negative (suitable assumptions about the utility function, such as $u'(c) \to \infty$ as $c \searrow 0$, will guarantee the success of this procedure).

For some utility functions it is possible that one or more of the non-negativity constraints will be binding, in which case the procedure just described would be unsuccessful. Standard methods can be used to cope with such complications, but they will not be described here.

Notice that equation (2.20) is similar to the condition that must be satisfied by a risk neutral probability measure. Not surprisingly, therefore, we have the following counterpart to (2.6):

**(2.21)** If $C$ is part of a solution to the optimal consumption investment problem (2.16) with $C_0 > 0$ and $C_1(\omega) > 0$ for all $\omega$, then

$$Q(\omega) = P(\omega)B_1(\omega)\frac{u'(C_1(\omega))}{u'(C_0)}$$

defines a risk neutral probability measure.

To see this, simply note that

$$E_Q[S_n(1)/B_1] = \sum Q(\omega)S_n(1,\omega)/B_1(\omega)$$
$$= \sum P(\omega)B_1(\omega)\left(\frac{u'(C_1(\omega))}{u'(C_0)}\right)\left(\frac{S_n(1,\omega)}{B_1(\omega)}\right)$$
$$= \frac{1}{u'(C_0)}E[u'(C_1)S_n(1)]$$

Meanwhile, if $C_0 > 0$ and $C_1(\omega) > 0$ for all $\omega \in \Omega$, then the first order necessary condition (2.20) must hold. In this case it follows that $E[u'(C_1)S_n(1)]/u'(C_0)$ and thus $E_Q[S_n(1)/B_1]$ equal $S_n(0)$. Finally, using (2.19) it is easy to show that $\sum Q(\omega) = 1$, and so, indeed, $Q$ as defined in (2.21) is a risk neutral probability measure.

We now turn to the risk neutral computational approach for solving the consumption-investment problem (2.16) in the case of a complete model. The idea is to first use principle (2.15) to rewrite (2.16) as follows:

**(2.22)**
$$\begin{aligned} &\text{maximize} && u(C_0) + E[u(C_1)] \\ &\text{subject to} && C_0 + E_Q[C_1/B_1] = v \\ & && C_0 \geqslant 0 \qquad C_1 \geqslant 0 \end{aligned}$$

The optimization problems (2.16) and (2.22) are essentially the same, because if the pair $(C, H)$ is feasible for (2.16), then $C$ is feasible for (2.22); conversely, if $C$ is feasible for (2.22), then there exists some $H$ such that $(C, H)$ is feasible for (2.16).

Notice the trading strategy $H$ does not appear at all in (2.22), so the first step with the risk neutral computational approach, solving (2.22), is much easier than solving (2.16). This leaves for the second and final step the computation of the trading strategy $H$ that generates the contingent claim $C_1$, where $C_1$ is time $t = 1$ consumption under the solution of subproblem (2.22). Analogous to the optimal portfolio problem, these two steps can be readily solved with standard methods.

To solve (2.22) with a Lagrange multiplier, one first analyses the unconstrained problem

**(2.23)**
$$\text{maximize} \quad u(C_0) + E[u(C_1)] - \lambda\{C_0 + E[C_1L/B_1]\}$$

(recall $E[C_1L/B_1] = E_Q[C_1/B_1]$). With suitable assumptions about the utility function $u$ to ensure the optimal solution of (2.22) will feature strictly positive consumption values, the following first order necessary conditions must be satisfied:

$$u'(C_0) = \lambda \quad \text{and} \quad u'(C_1(\omega)) = \lambda L/B_1$$

Hence

**(2.24)**
$$C_0 = I(\lambda) \quad \text{and} \quad C_1(\omega) = I(\lambda L/B_1)$$



where $I(\cdot)$ is the inverse of the marginal utility function $u'(\cdot)$. Of course, the Lagrange multiplier $\lambda$ must take the correct value, namely, the value such that the constraint in (2.22) is satisfied. This is

**(2.25)**
$$I(\lambda) + E_Q[I(\lambda L/B_1)/B_1] = v$$

As is the case with (2.13), the inverse function I is decreasing so this equation will normally have a solution $\lambda$. If the corresponding values of $C_0$ and $C_1$, as given by (2.24), are non-negative, then they must be an optimal solution of (2.22).

If this procedure yields a consumption value that is not non-negative, then a more complicated algorithm must be used to derive the solution of (2.22). Such algorithms are standard, although they will not be discussed here. Suffice it to say that in this case it is still much easier to solve (2.22) than (2.16).

**Example 2.3** Suppose $u(c) = \ln(c)$, so that $u'(c) = 1/c$ and the inverse function $I(i) = 1/i$. Equations (2.24) and (2.25) become

$$C_0 = 1/\lambda \quad \text{and} \quad C_1(\omega) = 1/(\lambda L/B_1)$$

and

$$\frac{1}{\lambda} + \frac{1}{\lambda}E_Q[L^{-1}] = \frac{1}{\lambda} + \frac{1}{\lambda}E[1] = \frac{2}{\lambda} = v$$

so $\lambda = 2/v$, $C_0 = v/2$, and $C_1(\omega) = vB_1(\omega)P(\omega)/[2Q(\omega)]$. Notice that these are non-negative as long as $v \geqslant 0$. Substituting these values gives the maximum value of the objective function in (2.22) to be $2\ln(v/2) + E[\ln(B_1/L)]$.

**Example 2.1 and 2.3 (continued)** With $L(\omega_1) = 2/3$, $L(\omega_2) = L(\omega_3) = 4/3$, and $r = 1/9$ as before, we have

$$C_1(\omega) = v\frac{5}{9}L^{-1} = \begin{cases} \frac{5}{6}v, & \omega = \omega_1 \\ \frac{5}{12}v, & \omega = \omega_2, \omega_3 \end{cases}$$

Note that the equation in (2.15) as well as the necessary conditions (2.18) are satisfied, as desired. We compute the optimal $H_1$ and $H_2$ by solving the system $C_1/B_1 = v/2 + G^*$, that is,

$$\frac{3}{4}v = \frac{1}{2}v + 0H_1 + 3H_2$$
$$\frac{3}{8}v = \frac{1}{2}v + 2H_1 - 1H_2$$
$$\frac{3}{8}v = \frac{1}{2}v - 2H_1 - 2H_2$$

Although there are three equations and two unknowns, the solution is unique: $H_1 = -v/48$ and $H_2 = v/12$. Since $v/2 = H_0 + 6H_1 + 10H_2$, it follows that $H_0 = -(5/24)v$.

In summary, principle (2.15) greatly simplifies the solution of the optimal consumption investment problem because it allows one to decompose the original problem into two simpler subproblems: in the first you solve for the optimal consumption process without worrying about the trading strategy, and in the second you derive the trading strategy that corresponds to the solution of the first subproblem.

The basic consumption investment problem (2.16) can be generalized in several directions. For example, the objective function can be written as

**(2.26)**
$$u(C_0) + \beta E[u(C_1)]$$

where the scalar $\beta$ satisfies $0 < \beta \leqslant 1$. The idea here is to model the time-value of when consumption occurs by regarding the specified parameter $\beta$ as a discount factor.

A second generalization of (2.16) is to allow the consumer to have income or endowment $\tilde{E}$ at time $t = 1$, where $\tilde{E}$ is a specified random variable. The optimization problem thus is:

**(2.27)**
$$\begin{aligned} &\text{maximize} && u(C_0) + E[u(C_1)] \\ &\text{subject to} && C_0 + H_0B_0 + \sum_{n=1}^{N} H_nS_n(0) = v \\ & && C_1 - H_0B_1 - \sum_{n=1}^{N} H_nS_n(1) = \tilde{E} \\ & && C_0 \geqslant 0 \qquad C_1 \geqslant 0 \qquad H \in \mathbb{R}^{N+1} \end{aligned}$$

The pair $(v, \tilde{E})$ is sometimes called the *endowment process* for the consumer.

**Exercise 2.4** Derive formulas for $\lambda$, $C_0$, and $C_1$ for the consumption investment problem in the case where the utility function is:

(a) $u(c) = -\exp\{-c\}$.
(b) $u(c) = \gamma^{-1}c^{\gamma}$, where $-\infty < \gamma < 1$ and $\gamma \neq 0$.

**Exercise 2.5** Show that if the objective function for the consumption investment problem is as in (2.26), then the equation in (2.21) should be generalized to:

$$Q(\omega) = \beta P(\omega) B_1(\omega) \frac{u'(C_1(\omega))}{u'(C_0)}$$

**Exercise 2.6** For the consumption investment problem (2.27) with an endowment, show that the equation in (2.15) generalizes to

$$C_0 + E_Q[(C_1 - \tilde{E})/B_1] = v$$



**Exercise 2.7** For example 1.4, suppose the initial wealth $v = 100$. Characterize the set of all consumption processes $C$ such that there is a trading strategy $H$ making the consumption-investment plan $(C, H)$ admissible.

## 2.4 Mean-Variance Portfolio Analysis

Throughout this section it will be assumed that the interest rate $r$ is deterministic, there are no arbitrage opportunities, and there exists some portfolio with $E[R] \neq r$. A classical problem in this case is to solve the mean-variance portfolio problem:

**(2.28)**
$$\begin{aligned} &\text{minimize} && \text{var}(R) \\ &\text{subject to} && E[R] = \rho \\ & && R \text{ is a portfolio return} \end{aligned}$$

where $\rho$ is a specified scalar. Notice that for each value of $\rho \geqslant r$, the feasible region in (2.28) is non-empty (to see this, just take a suitable linear combination of the riskless portfolio and a specific one with $E[R] \neq r$), and so the solution of (2.28) is well defined. In particular, the optimal value of the objective function equals zero if and only if $\rho = r$; otherwise, it will be a finite, positive number.

Recalling that the return for a portfolio can be expressed as

$$R = \frac{H_0}{V_0} r + \sum_{n=1}^{N} \left[\frac{H_nS_n(0)}{V_0}\right] R_n$$

it follows that $R$ is the return for a portfolio if and only if it can be written in the form $R = (1 - F_1 - \ldots - F_N)r + \sum F_nR_n$, where $F_n$ can be interpreted as the fraction of time $t = 0$ wealth that is invested in security $n$. Hence (2.28) can be rewritten as

**(2.29)**
$$\begin{aligned} &\text{minimize} && F\mathbb{C}F' \\ &\text{subject to} && (1 - F_1 - \ldots - F_N)r + \sum F_n\bar{R}_n = \rho \end{aligned}$$

where $\mathbb{C}$ is the $N \times N$ matrix of the covariances for the returns (its $ij$th element is $\text{cov}(R_i, R_j)$, $\bar{R}_n = ER_n$, and $F \equiv (F_1, \ldots, F_N)$. This is a quadratic programming problem, a well-known kind of problem in the area of optimization theory, and it has been the subject of extensive study by financial researchers and mathematical programmers.

In this section, problem (2.28) will be solved with a different approach, one that relies on the modern theory developed in section 2.2. Consider the following problem:

**(2.30)**
$$\begin{aligned} &\text{minimize} && \text{var}(V_1) \\ &\text{subject to} && E[V_1] = v(1 + \rho) \\ & && V_0 = v \end{aligned}$$

Here the constraints identify the set of all time $t = 1$ portfolio values that can be achieved starting with initial wealth $v$ and which have mean $v(1+\rho)$, where $\rho \geqslant r$ is as in (2.28) and $v > 0$ is a specified scalar. Notice that if $\hat{V}_1$ is a solution of (2.30), then $\hat{R} \equiv (\hat{V}_1 - v)/v$ satisfies the constraints in (2.28). Moreover, if $R$ is any other return that is feasible for (2.28), then $V_1 \equiv v(1+R)$ satisfies $E(V_1) = v(1+\rho)$, which means $V_1$ is feasible for (2.30). Hence

$$\operatorname{var}(\hat{R}) = \frac{1}{v^2}\operatorname{var}(\hat{V}_1) \leqslant \frac{1}{v^2}\operatorname{var}(V_1) = \operatorname{var}(R)$$

which means $\hat{R}$ is a solution of (2.28). Conversely, a solution of (2.28) gives just as easily a solution of (2.30), so really (2.28) and (2.30) are equivalent problems. In other words,

**(2.31)** The relationship $V_1 = v(1+R)$ establishes a one-to-one correspondence between feasible solution of (2.28) and (2.30).

You can see where we are headed: we want to reformulate (2.28) so we can apply the results of section 2.2, and (2.30) is a step in this direction. For the next step, consider how to solve (2.30) with a Lagrange multiplier. Introducing the scalar $\beta$, one is interested in minimizing the objective function $\operatorname{var}(V_1) - \beta E[V_1]$, subject to the constraint $V_0 = v$. But $\operatorname{var}(V_1) = E[V_1^2] - (EV_1)^2$, so, as will be verified later, this objective function can be written as $E[\frac{1}{2}V_1^2 - \beta V_1]$. In other words, with the Lagrange multiplier approach we are interested in solving

$$\text{(2.32)} \qquad \begin{aligned} &\text{maximize} \quad E[-\frac{1}{2}V_1^2 + \beta V_1] \\ &\text{subject to} \quad V_0 = v. \end{aligned}$$

Problem (2.32) is in the form studied in section 2.2, so applying the results there one concludes the optimal solution, denoted $\hat{V}$, is given by (this is left as an exercise)

$$\text{(2.33)} \qquad \hat{V} = \frac{\beta}{E_Q L}(E_Q L - L) + v(1+r)\frac{L}{E_Q L}$$

in which case $E[\hat{V}] = \beta(E_Q L - 1)/E_Q L + v(1+r)/E_Q L$ (recall $L = Q/P$ is the state price density). Now assuming $Q$ and $P$ are not identical, one has $E_Q L > 1$ (this is also left as an exercise). Hence $E[\hat{V}] = v(1+\rho)$ if and only if

$$\text{(2.34)} \qquad \beta = \frac{v[(1+\rho)E_Q L - (1+r)]}{E_Q L - 1}$$

Substituting this into (2.33) means that $\hat{V}$ is feasible for (2.30). If $V$ is any other random variable that is feasible for (2.30), then the facts that $E[\hat{V}] = E[V]$ and that $E[-\frac{1}{2}\hat{V}^2 + \beta\hat{V}] \geqslant E[-\frac{1}{2}V^2 + \beta V]$ imply $\operatorname{var}(\hat{V}) \leqslant \operatorname{var}(V)$. Hence $\hat{V}$ is an optimal solution of (2.30).

Conversely, suppose $\hat{V}$ is a solution of (2.30) with $\beta$ as in (2.34). If $V$ is any other random variable that is feasible for (2.30), then $E[-\frac{1}{2}\hat{V}^2 + \beta\hat{V}] \geqslant E[-\frac{1}{2}V^2 + \beta V]$. We saw above that any solution of (2.32) must be feasible



for (2.30), so we conclude that $\hat{V}$ must be an optimal solution of (2.32). The relationship between (2.30) and (2.32) is summarized as follows:

**(2.35)** Portfolio problems (2.30) and (2.32) are equivalent provided $\rho$ and $\beta$ are related according to (2.34).

Notice according to (2.34) that $\beta$ is a strictly increasing function of $\rho$ which equals $v(1+r)$ when $\rho = r$. Moreover, when $\rho = r$, the optimal solution (2.33) is $\hat{V} = v(1+r)$. This is a constant, as anticipated.

Problem (2.32) has the standard form of sections 2.1 and 2.2 if the investor's utility function is taken to be the quadratic function $u(w) = -w^2/2 + \beta w$. Although quadratic utility functions are dubious because they are not non-decreasing with respect to wealth (this function is concave but attains its maximum value at $w = \beta < \infty$), they are accepted for various applications. In particular, principle (2.31) implies there is a one-to-one correspondence between solutions of the quadratic utility portfolio problem (2.32) and solutions of the mean-variance portfolio problem (2.28).

This correspondence has a fundamental consequence. Looking at (2.33) we see that the solution of (2.28) must be an affine[1] function of the state price density. Indeed, substituting (2.34) in (2.33) we compute that the return $\hat{R}$ corresponding to $\hat{V}$ is

$$\text{(2.36)} \qquad \hat{R} = \frac{\rho E_Q L - r}{E_Q L - 1} - \frac{\rho - r}{E_Q L - 1}L$$

Hence we conclude (this also could have been worked out using (2.12) for the case of a quadratic utility function):

**(2.37)** The optimal solution $R$ of the mean-variance portfolio problem (2.28) is an affine function of the state price density $L$.

A further consequence is realized if you tie this together with result (1.35) concerning the relationship between the risk premium of an arbitrary portfolio and its beta with respect to a portfolio whose return is an affine function of the state price density. It is apparent that we have established the famous security market line result of the capital asset pricing model (CAPM) theory:

**(2.38)** *CAPM*: If $R'$ is a solution of the mean-variance portfolio problem (2.28) for $\rho \geqslant r$ and if $R$ is the return of an arbitrary portfolio, then

$$E[R] - r = \frac{\operatorname{cov}(R, R')}{\operatorname{var}(R')}(E[R'] - r)$$

This relationship is quite important, because in a world of mean-variance investors there is often a portfolio (for example, a stock index) which can be assumed to be a solution of (2.28) and whose mean return can be estimated, thereby giving via (2.38) estimates of the mean return of arbitrary portfolios.

Principle (2.37) has another important consequence. Fix some arbitrary $\hat{\rho} > r$ and consider the corresponding return $\hat{R}$ as given by (2.36). Think of $\hat{R}$

as corresponding to some portfolio or mutual fund that is available for investment. Suppose an investor puts the fraction $\lambda$ of his money in the riskless security and the balance $1-\lambda$ in this mutual fund, where $\lambda = (\hat{\rho}-\rho)/(\hat{\rho}-r)$ and $\rho \geqslant r$ ($\lambda < 0$ corresponds to borrowing at the riskless rate). Since this portfolio's return $R = \lambda r + (1-\lambda)\hat{R}$, one can do some tedious algebra to verify that, in fact, $R$ is given precisely by (2.36), only with $\rho$ instead of $\hat{\rho}$ on the right hand side. Hence to achieve any solution of the mean-variance portfolio problem (2.28) it is not necessary to trade the individual securities, provided there is a mutual fund available which corresponds to one solution. It is just a question of dividing up the invested funds between the riskless security and the mutual fund. Since all the risky securities are in the mutual fund, it must mean that the relative proportions invested in the risky securities (that is, the money invested in security $n$ divided by the money invested in the mutual fund) are constants with respect to $\rho$. This all can be summarized as follows:

**(2.39)** *Mutual fund principle*: Suppose you fix a portfolio whose return is a solution of the mean-variance portfolio problem (2.28) corresponding to some mean return $\hat{\rho} > r$. Then the solution of (2.28) can be achieved for any other mean return by a portfolio consisting of investments in just the riskless security and the fixed portfolio.

Hence a world of mean-variance investors is quite nice; it enjoys many nice properties. However, it should be stressed that many of these nice properties may disappear in the presence of investors whose decisions are not consistent with quadratic utility functions. For example, as seen in section 2.2, log utility investors will want to choose portfolios whose time $t = 1$ values are proportional to the *inverse* of the state price density. In this case the time $t = 1$ wealth cannot, in general, be expressed as an affine function of the state price density, and the security market line result will not hold with respect to any portfolio in which the log utility investor would desire to invest his or her money.

Indeed, it is almost a lucky accident that the CAPM security market line result holds for a reasonable class of utility functions. With most utility functions the return of the optimal portfolio will not be an affine function of the state price density and thus cannot play the role of $R'$ in (1.35). Result (1.35) is more fundamental in the sense that it applies to any single period securities market provided its general hypotheses are satisfied, whereas the CAPM result (2.38) is a special case or corollary. But the general version is not particularly useful unless you can identify $R'$ with an economically meaningful portfolio. In the mean-variance world you can do this, and that is why the CAPM result (2.38) is so important.

**Example 2.4** Suppose $N = 2$, $K = 3$, and the return processes $R_n$ and the probability measure $P$ are as indicated in table 2.1.



**Table 2.1** Data for example 2.4

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $R_1(\omega)$ | 0.2 | −0.2 | 0.05 |
| $R_2(\omega)$ | 0.15 | 0 | −0.1 |
| $P(\omega)$ | 1/3 | 1/3 | 1/3 |
| $R_1^*(\omega)$ | $\frac{0.2-r}{1+r}$ | $\frac{-0.2-r}{1+r}$ | $\frac{0.05-r}{1+r}$ |
| $R_2^*(\omega)$ | $\frac{0.15-r}{1+r}$ | $\frac{-r}{1+r}$ | $\frac{-0.1-r}{1+r}$ |
| $Q(\omega)$ | $0.258 + 4.52r$ | $0.355 - 1.3r$ | $0.387 - 3.22r$ |
| $L(\omega)$ | $0.774 + 13.56r$ | $1.065 - 3.9r$ | $1.161 - 9.66r$ |

Solving $E_Q R_n^* = 0$ for the risk neutral probability measure $Q$, it is apparent that this exists, provided $r < 0.387/3.22 = 0.12$. These quantities along with the resulting state price density $L$ are displayed in table 2.1.

To solve the classical mean-variance problem (2.29), we first compute $ER_1 = ER_2 = 1/60$, $\mathrm{var}(R_1) = 0.02722$, $\mathrm{var}(R_2) = 0.01056$, and $\mathrm{cov}(R_1, R_2) = 0.00805$. Assuming (for simplicity) from now on that $r = 0$, this leads to the solution $F_1 = 6.95\rho$ and $F_2 = 53.05\rho$. Hence, for example, if $\rho = 1$ percent, then 6.95 percent of the funds should be invested in the first risky security, 53.05 percent in the second risky security, and 40 percent in the riskless security. Moreover, the return of the resulting portfolio is

$$R = F_1 R_1 + F_2 R_2 = \begin{cases} 0.0936, & \omega = \omega_1 \\ -0.0140, & \omega = \omega_2 \\ -0.0496, & \omega = \omega_3 \end{cases}$$

Alternatively, using formulas (2.33) and (2.34) one computes $E_Q L = 1.027$, $\beta = v[1 + 38.04\rho]$, $\hat{V} = v[1 + 38.04\rho - 37.04\rho L]$, and $\hat{R} = 38.04\rho - 37.04\rho L$. To get $F_1$ and $F_2$, you solve $\hat{R} = F_1 R_1 + F_2 R_2$. Hence, for example, if $\rho = 1$ percent, then you get the values already presented above.

**Exercise 2.8** With $Q$ a risk neutral probability measure, $P$ the original probability measure, and $L$ the corresponding state price density, show that $E_Q L \geqslant 1$, with equality if and only if $Q = P$ (hint: minimize $\sum Q^2(\omega_k)/P(\omega_k)$ subject to suitable constraints).

**Exercise 2.9** Verify equation (2.33).

**Exercise 2.10** Verify the assertion made in connection with the mutual fund principle (2.39): $R = \lambda r + (1-\lambda)\hat{R}$ satisfies (2.36), where $\lambda = (\hat{\rho}-\rho)/(\hat{\rho}-r)$.

## 2.5 Portfolio Management with Short Sales Restrictions and Similar Constraints

The basic optimal portfolio problem studied in earlier sections may be inappropriate for many practical situations because important constraints are ignored. For example, the investor might be prohibited by stock exchange rules from selling stocks short or from financing the purchase of stocks by borrowing money. Consequently, it is important to be able to solve versions of problem (2.1) where constraints are imposed on the admissible trading strategies.

In actual situations it is usually more natural to express the constraints in terms of the fractions $F_n \equiv H_n S_n(0)/V_0$ of money invested in security $n, n = 1, \ldots, N$, rather than in terms of the number $H_n$ of shares invested in security $n$. For example, no short selling of security $n$ is $F_n \geqslant 0$, no borrowing from the bank account is $F_1 + \ldots + F_N \leqslant 1$, and a stipulation that no more than 4 percent of the wealth can be invested in security $n$ is $F_n \leqslant 0.04$. In general, therefore, the constraints of interest will be expressed by stipulating that $F \equiv (F_1, \ldots, F_N) \in \mathbb{K}$, where $\mathbb{K} \subset \mathbb{R}^N$ is a specified subset that is assumed to be closed and convex. In order to simplify the presentation later in this section, it will also be assumed that the strategy $0 \in \mathbb{K}$, that is, it is always feasible to invest all the funds in the bank account. For example, $\mathbb{K}$ is equal to $\{F \in \mathbb{R}^N : F_n \geqslant 0\}$, $\{F \in \mathbb{R}^N : F_1 + \ldots + F_N \leqslant 1\}$, and $\{F \in \mathbb{R}^N : F_n \leqslant 0.04\}$, respectively, in the three cases mentioned just above.

With the trading strategy expressed in the form $F$, it is convenient to express the time $t = 1$ wealth $V_1 = V_0(1 + R)$ in terms of $R$, the return process for the portfolio. In view of (1.32) this is

$$R = \left(1 - \sum_{n=1}^{N} F_n\right) r + \sum_{n=1}^{N} F_n R_n = r + \sum_{n=1}^{N} F_n (R_n - r)$$

where $R_n$ is the return process for security $n$. Hence the optimal portfolio problem can be written as

$$\text{(2.40)} \qquad \underset{F \in \mathbb{K}}{\text{maximize}} \quad Eu\Big(v\Big(1 + r + \sum_{n=1}^{N} F_n (R_n - r)\Big)\Big)$$

where $v = V_0$ is the initial value of the portfolio.

It is straightforward to solve this kind of problem with traditional methods, similar to those used in section 2.1 for the unconstrained problem. This is best explained by considering a simple example.

**Example 2.5** The security processes are the same as in example 2.4, but now $r = 0$, $P(\omega_1) = 0.26$, $P(\omega_2) = P(\omega_3) = 0.37$, and the utility function is the log function. Thus the objective function is

$$0.26 \ln[1 + 0.2F_1 + 0.15F_2] + 0.37 \ln[1 - 0.2F_1] + 0.37 \ln[1 + 0.05F_1 - 0.1F_2]$$



and the corresponding partial derivatives are

$$\frac{\partial}{\partial F_1} = \frac{0.052}{1 + 0.2F_1 + 0.15F_2} - \frac{0.074}{1 - 0.2F_1} + \frac{0.0185}{1 + 0.05F_1 - 0.1F_2}$$

and

$$\frac{\partial}{\partial F_2} = \frac{0.039}{1 + 0.2F_1 + 0.15F_2} - \frac{0.037}{1 + 0.05F_1 - 0.1F_2}$$

Hence the optimal solution for the unconstrained problem, obtained by setting these partials equal to zero, is $F_1 = -0.21333$ and $F_2 = 0.33467$.

Now suppose that short sales of the risky securities are prohibited, that is, $\mathbb{K} = \{F \in \mathbb{R}^2 : F_1 \geqslant 0 \text{ and } F_2 \geqslant 0\}$. In this case the unconstrained optimal solution is not feasible, and so we must do some more work. It is apparent that the new optimal solution must be on the boundary of $\mathbb{K}$ at a point where the directional derivative is normal to $\mathbb{K}$. In view of the unconstrained optimal solution, we conjecture the new optimal solution satisfies $F_1 = 0$ and $F_2 > 0$. We therefore look for a point satisfying $F_1 = 0$, $F_2 > 0$, $\partial/\partial F_2 = 0$, and $\partial/\partial F_1 < 0$. Using the above expressions for the partials, we readily compute the optimal solution for the constrained problem to be $F_1 = 0$ and $F_2 = 0.21164$. Note that the optimal attainable wealth is

$$W = v[1 + F_1 R_1 + F_2 R_2] = \begin{cases} 1.03175v & \omega = \omega_1 \\ v, & \omega = \omega_2 \\ 0.97884v, & \omega = \omega_3 \end{cases}$$

and the optimal objective value is

$$E \ln W = \ln v + 0.26 \ln(1.03175) + 0.37 \ln(0.97884) = \ln v + 0.00021$$

To summarize the traditional method for solving problem (2.40), first obtain the unconstrained optimal solution and check to see if it is feasible; if not, then look for a point on the boundary of $\mathbb{K}$ where the directional derivative is normal to $\mathbb{K}$, using the partial derivatives throughout. While this method was easy to apply in the case of example 2.5, the computations could become formidable in a case where there are many securities and/or there is a different kind of utility function. Furthermore, these computational difficulties are compounded when dealing with multiperiod models. We therefore are interested in an alternative method, a risk neutral computational approach.

The risk neutral computational approach for constrained portfolio problems is roughly as follows. For each value of a parameter $\kappa$ contained in a certain subset $\tilde{\mathbb{K}} \subset \mathbb{R}^N$ define a modified securities market denoted $\mathcal{M}_\kappa$ ($\kappa = 0$ corresponds to the original market). Then consider the unconstrained problem for the market $\mathcal{M}_\kappa$. In particular, using the formulas developed in section 2.2 focus on the optimal objective value, denoted $J_\kappa(v)$, for each $\kappa \in \tilde{\mathbb{K}}$. Then solve the *dual problem*:

$$\text{(2.41)} \qquad \underset{\kappa \in \tilde{\mathbb{K}}}{\text{minimize}} \quad J_\kappa(v)$$

If $\hat{\kappa}$ denotes the optimal solution of (2.41), then the optimal solution for the unconstrained problem in the market $\mathscr{M}_{\hat{\kappa}}$ will turn out to be the optimal solution for the constrained problem in the original market $\mathscr{M}_0$. Moreover, the corresponding optimal objective values will coincide.

Although the risk neutral computational approach still requires one to solve a constrained optimization problem with traditional methods, it turns out that solving (2.41) is often easier than solving (2.40).

Turning to some details, the set $\tilde{\mathbb{K}}$ is simply

$$\tilde{\mathbb{K}} \equiv \{\kappa \in \mathbb{R}^N : \delta(\kappa) < \infty\}$$

where the function $\delta : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\delta(\kappa) \equiv \sup_{F \in \mathbb{K}} (-F\kappa')$$

and $\kappa'$ denotes the transpose of the row vector $\kappa$. The function $\delta$ is convex and is called the *support function* of $-\mathbb{K}$. Notice that $\delta$ is non-negative, because $0 \in \mathbb{K}$. It will be assumed that $\mathbb{K}$ is such that $\delta$ is continuous on $\tilde{\mathbb{K}}$. The set $\tilde{\mathbb{K}}$, called the *effective domain* of $\delta$, is a convex cone that contains the point $\kappa = 0$. For example, if $\mathbb{K} = \{F \in \mathbb{R}^N : F_n \geqslant 0, n = 1, \ldots, N\}$, then

$$\text{(2.42)} \qquad \delta(\kappa) = \begin{cases} 0, & \kappa \in \mathbb{K} \\ \infty, & \kappa \notin \mathbb{K} \end{cases}$$

and $\tilde{\mathbb{K}} = \mathbb{K}$. For another example, if $\mathbb{K} = \{F \in \mathbb{R}^N : F_1 + \ldots + F_N \leqslant 1\}$, then

$$\text{(2.43)} \qquad \delta(\kappa) = \begin{cases} -\lambda, & \kappa_1 = \ldots = \kappa_N = \lambda \leqslant 0 \\ \infty, & \text{otherwise} \end{cases}$$

and $\tilde{\mathbb{K}} = \{\kappa \in \mathbb{R}^N : \kappa_1 = \ldots = \kappa_n \leqslant 0\}$.

To define the auxiliary market $\mathscr{M}_\kappa$ for each $\kappa \in \tilde{\mathbb{K}}$, we simply modify the return processes for the bank account and the risky securities according to

$$r \to r + \delta(\kappa)$$
$$R_n \to R_n + \delta(\kappa) + \kappa_n, \qquad n = 1, \ldots, N$$

In other words, in the market $\mathscr{M}_\kappa$ the bank's interest rate is replaced by the original rate plus the non-negative quantity $\delta(\kappa)$, and so forth. Notice that the case $\kappa = 0$ does indeed coincide with the original market.

In order to solve the unconstrained problem in the market $\mathscr{M}_\kappa$, we will need to use the corresponding risk neutral probability measure, which is denoted $Q_\kappa$. Throughout this section it will be assumed that the risk neutral probability measure $Q = Q_0$ for the original market exists and is unique. Since $Q_0$ is the unique probability measure satisfying

$$E_Q\left(\frac{R_n - r}{1 + r}\right) = 0, \qquad n = 1, \ldots, N$$



it follows for other $\kappa \in \tilde{\mathbb{K}}$ that $Q_\kappa$ must be a probability measure satisfying for $Q = Q_\kappa$

$$\text{(2.44)} \qquad E_Q\left(\frac{R_n + \kappa_n - r}{1 + r + \delta(\kappa)}\right) = 0, \qquad n = 1, \ldots, N$$

It is not clear whether $Q_\kappa$ will exist for all $\kappa \in \tilde{\mathbb{K}}$, but a unique $Q_\kappa$ will exist and the function $\kappa \to Q_\kappa$ will be continuous at least for all $\kappa \in \tilde{\mathbb{K}}$ in some open neighborhood of $\kappa = 0$, by the assumed continuity of $\delta(\cdot)$.

In the market $\mathscr{M}_\kappa$ the state price density and the bank account process are denoted respectively by

$$L_\kappa \equiv \frac{Q_\kappa}{P} \qquad \text{and} \qquad B_1^\kappa = 1 + r + \delta(\kappa)$$

Moreover, given any trading strategy $F$, the time $t = 1$ value of the portfolio in the market $\mathscr{M}_\kappa$ is given by

$$\begin{aligned} \text{(2.45)} \qquad V_1^\kappa &= v(1 + R^\kappa) = v\left[1 + r + \delta(\kappa) + \sum_{n=1}^{N} F_n(R_n + \kappa_n - r)\right] \\ &= v\left[1 + r + \sum_{n=1}^{N} F_n(R_n - r) + \delta(\kappa) + \sum_{n=1}^{N} F_n\kappa_n\right] \\ &= V_1^0 + v[\delta(\kappa) + F\kappa'] \end{aligned}$$

It is important to note that if $F \in \mathbb{K}$, then by the definition of $\delta(\cdot)$ one has $\delta(\kappa) + F\kappa' \geqslant 0$, in which case $V_1^\kappa \geqslant V_1^0$. On the other hand, if $F \notin \mathbb{K}$, then possibly $\delta(\kappa) + F\kappa' < 0$, in which case $V_1^\kappa < V_1^0$ and $Eu(V_1^\kappa) < Eu(V_1^0)$. This is the reason why it is possible to have the optimal objective values satisfy $J_\kappa(v) < J_0(v)$.

We are now in a position to illustrate how the risk neutral computational approach works.

**Example 2.5 (continued)** With $\mathbb{K} = \{F \in \mathbb{R}^2 : F_1 \geqslant 0, F_2 \geqslant 0\}$, we have $\tilde{\mathbb{K}} = \mathbb{K}$ and $\delta(\cdot)$ as in (2.42) above. With $r = \delta(\kappa) = 0$ for $\kappa \in \tilde{\mathbb{K}}$, (2.44) reduces to

$$\begin{aligned} 0.2Q_\kappa(\omega_1) - 0.2Q_\kappa(\omega_2) + 0.05Q_\kappa(\omega_3) &= -\kappa_1 \\ 0.15Q_\kappa(\omega_1) \qquad\qquad - 0.1Q_\kappa(\omega_3) &= -\kappa_2 \end{aligned}$$

Solving this system along with $E_{Q_\kappa}[1] = 1$ leads to

$$Q_\kappa(\omega) = \begin{cases} [8 - 40\kappa_1 - 100\kappa_2]/31, & \omega = \omega_1 \\ [11 + 100\kappa_1 - 60\kappa_2]/31, & \omega = \omega_2 \\ [12 - 60\kappa_1 + 160\kappa_2]/31, & \omega = \omega_3 \end{cases}$$

These probabilities are strictly positive, and thus $Q_\kappa$ is a valid risk neutral probability measure for the market $\mathscr{M}_\kappa$, as long as $\kappa \in \tilde{\mathbb{K}}$ and $40\kappa_1 + 100\kappa_2 < 8$.

The next step is to solve the unconstrained optimization problem for the market $\mathcal{M}_\kappa$. Letting $W_\kappa$ denote the corresponding optimal attainable wealth, we have by exercise 2.2 in section 2.2

(2.46) $$W_\kappa = vB_1^\kappa/L_\kappa = vP/Q_\kappa$$

Moreover, the optimal value of the objective function is

(2.47) $$\begin{aligned} J_\kappa(v) &= \ln(v) - E\ln(L_\kappa/B_1^\kappa) = \ln(v) + E\ln P - E\ln Q_\kappa \\ &= \ln(v) + [0.26\ln(0.26) + 2(0.37)\ln(0.37)] \\ &\quad - [0.26\ln(8 - 40\kappa_1 - 100\kappa_2) + 0.37\ln(11 - 100\kappa_1 - 60\kappa_2) \\ &\quad + 0.37\ln(12 - 60\kappa_1 + 160\kappa_2) - \ln(31)] \end{aligned}$$

The next step is to minimize $J_\kappa(v)$ with respect to $\kappa \in \tilde{\mathbb{K}}$. In view of the preceding expression, this is the same as solving

$$\begin{aligned} \text{maximize} \quad & 0.26\ln(8 - 40\kappa_1 - 100\kappa_2) + 0.37\ln(11 - 100\kappa_1 - 60\kappa_2) \\ & + 0.37\ln(12 - 60\kappa_1 + 160\kappa_2) \\ \text{subject to} \quad & \kappa_1 \geqslant 0, \quad \kappa_2 \geqslant 0 \end{aligned}$$

Using partial derivatives in the standard way, the optimal solution is computed to be $\hat{\kappa}_1 = 0.0047$ and $\hat{\kappa}_2 = 0$. Substituting these values in (2.46) and (2.47) yields the same values for the optimal attainable wealth and the optimal objective value as were obtained earlier with the traditional approach.

Why does this risk neutral approach work? The key is expression (2.45) for the attainable wealths in the market $\mathcal{M}_\kappa$ as well as the observations made immediately thereafter. By the same considerations, the optimal objective value for the constrained problem in the original market $\mathcal{M}_0$, which we denote by $J(v)$, must be less than or equal to the optimal objective value for the constrained problem in the market $\mathcal{M}_\kappa$ for any $\kappa \in \tilde{\mathbb{K}}$. And the latter, of course, must be less than or equal to the optimal objective value for the unconstrained problem in the market $\mathcal{M}_\kappa$. Hence we must have

(2.48) $$J(v) \leqslant J_\kappa(v), \quad \text{all} \quad \kappa \in \tilde{\mathbb{K}}$$

Apparently this inequality can be an equality for the $\kappa$ that minimizes the right hand side of (2.48), as explained in the following.

**(2.49)** Suppose for some $\hat{\kappa} \in \tilde{\mathbb{K}}$ that $F$, the optimal trading strategy for the unconstrained portfolio problem in the market $\mathcal{M}_{\hat{\kappa}}$, satisfies

(a) $F \in \mathbb{K}$

(b) $\delta(\hat{\kappa}) + F\hat{\kappa}' = 0$.

Then $F$ is optimal for the constrained problem in the original market $\mathcal{M}_0$, and $J(v) = J_{\hat{\kappa}}(v) \leqslant J_\kappa(v)$ for all $\kappa \in \tilde{\mathbb{K}}$.



To see this, note by (2.45) that (b) implies $W$, the attainable wealth under $F$ in the market $\mathcal{M}_{\hat{\kappa}}$, satisfies

$$W = v\left[1 + r + \sum_{n=1}^{N} F_n(R_n - r)\right]$$

which means that $W$ is also the attainable wealth under $F$ in the original market $\mathcal{M}_0$. Since $F$ is feasible for the constrained problem, it follows that $Eu(W) \leqslant J(v)$. But $Eu(W) = J_{\hat{\kappa}}(v)$, so by (2.48) we must have $Eu(W) = J(v) = J_{\hat{\kappa}}(v) \leqslant J_\kappa(v)$ for all $\kappa \in \tilde{\mathbb{K}}$.

In summary, the obvious candidate for the $\hat{\kappa}$ in (2.49) is the solution of the dual problem (2.41). Having computed this $\hat{\kappa}$, you then verify whether $F$, the optimal trading strategy for the unconstrained optimal portfolio problem in the market $\mathcal{M}_{\hat{\kappa}}$, satisfies $F \in \mathbb{K}$ and $\delta(\hat{\kappa}) + F\hat{\kappa}' = 0$ (there is no guarantee that both these conditions will be satisfied, but in a wide variety of cases they will both automatically hold). If so, then $F$ will be optimal for the constrained problem in the original market $\mathcal{M}_0$.

This section concludes with another example.

**Example 2.6** The security processes are the same as in examples 2.4 and 2.5, but now $r = 0$, $P(\omega_1) = P(\omega_2) = P(\omega_3) = 1/3$, the utility function is the log function, and borrowing funds from the bank is prohibited. Thus $\mathbb{K} = \{F \in \mathbb{R}^2 : F_1 + F_2 \leqslant 1\}$, $\tilde{\mathbb{K}} = \{\kappa \in \mathbb{R}^2 : \kappa_1 = \kappa_2 \leqslant 0\}$, and

$$\delta(\kappa) = \begin{cases} -\lambda, & \kappa_1 = \kappa_2 = \lambda \leqslant 0 \\ \infty, & \text{otherwise,} \end{cases}$$

and so for ease of exposition we will identify the vector $\kappa \in \tilde{\mathbb{K}}$ with the scalar $\lambda \leqslant 0$.

The interest rate for the bank account in the market $\mathcal{M}_\kappa$ will be $-\lambda$, but the return processes for the risky securities in the market $\mathcal{M}_\kappa$ will be the same as in the original market $\mathcal{M}_0$. Hence by (2.44) the risk neutral probability measure $Q_\kappa$ can be obtained by solving the system

$$\begin{aligned} 0.2Q_\kappa(\omega_1) - 0.2Q_\kappa(\omega_2) + 0.05Q_\kappa(\omega_3) &= -\lambda \\ 0.15Q_\kappa(\omega_1) \qquad\qquad - 0.1Q_\kappa(\omega_3) &= -\lambda \\ Q_\kappa(\omega_1) + Q_\kappa(\omega_2) + Q_\kappa(\omega_3) &= 1 \end{aligned}$$

This leads to

$$Q_\kappa(\omega) = \begin{cases} [8 - 140\lambda]/31, & \omega = \omega_1 \\ [11 + 40\lambda]/31, & \omega = \omega_2 \\ [12 + 100\lambda]/31, & \omega = \omega_3 \end{cases}$$

These probabilities are all strictly positive and thus $Q_\kappa$ is a legitimate risk neutral probability measure if $-0.12 < \lambda \leqslant 0$.

With $B_1^\kappa = (1-\lambda)$, the optimal objective value for the unconstrained problem in the market $\mathcal{M}_\kappa$ is given by

$$J_\kappa(v) = \ln(v) - E\ln(L_\kappa/B_1^\kappa) = \ln(v) - E\ln Q_\kappa + E\ln P + \ln(1-\lambda)$$

Hence dual problem (2.41) amounts to the same thing as

$$\begin{aligned} \underset{-0.12<\lambda\leqslant 0}{\text{maximize}} \quad & \frac{1}{3}\ln(8-140\lambda) + \frac{1}{3}\ln(11+40\lambda) \\ & + \frac{1}{3}\ln(12+100\lambda) - \ln(1-\lambda) \end{aligned}$$

Although this objective function is not concave on the real line, the solution of this constrained problem is easily found to be approximately $\lambda = -0.00711$, that is, $\hat{\kappa} = (-0.00711, -0.00711)$.

The next step is to compute the optimal trading strategy $F$ for the unconstrained problem in the market $\mathcal{M}_{\hat{\kappa}}$. The corresponding optimal attainable wealth is

$$W_{\hat{\kappa}} = vB_1^{\hat{\kappa}}/L_{\hat{\kappa}} = \frac{1.0071vP}{Q_{\hat{\kappa}}} = \begin{cases} 1.157v, & \omega = \omega_1 \\ 0.972v, & \omega = \omega_2 \\ 0.921v, & \omega = \omega_3 \end{cases}$$

so $F$ can be computed by solving (2.45), that is,

$$W_{\hat{\kappa}}(\omega) = v[1.00711 + F_1(R_1(\omega) - 0.00711) + F_2(R_2(\omega) - 0.00711)]$$

This yields $F = (0.14, 0.86)$. Clearly $F \in \mathbb{K}$ and $\delta(\hat{\kappa}) + F\hat{\kappa}' = 0$, so by (2.49) $F$ must be the optimal solution for the constrained problem. Substituting $\hat{\kappa} = -0.00711$ in the above expression for $J_\kappa(v)$ gives the optimal objective value equal to $\ln v + 0.01171$.

**Exercise 2.11** Solve example 2.6 assuming $P(\omega_1) = 0.5$, $P(\omega_2) = 0.3$, and $P(\omega_3) = 0.2$.

## 2.6 Optimal Portfolios in Incomplete Markets

Throughout this chapter up to this point it has been assumed that the model is complete, a crucial assumption for the risk neutral computational approach. Under this assumption the set of attainable wealths is easy to identify and characterize, and so after using convex optimization theory to identify the optimal attainable wealth, one is assured of finding the trading strategy which generates this wealth. In the case of incomplete markets the principles are the same, but more work must be done to properly identify the set of attainable wealths. Having done so, one computes the optimal attainable wealth and finally the trading strategy which generates this wealth, the same as before.



Of course, one could always revert to the standard approach, working with the first order necessary conditions and so forth, exactly the same as described in section 2.1. The standard approach is no more difficult with incomplete than with complete models, so the relative advantages of the risk neutral computational approach are diminished in the case of incomplete models. However, the risk neutral approach is still preferred when dealing with certain utility functions and, as will be seen later, when analyzing multiperiod models.

It turns out there is even a third approach for solving optimal portfolio problems in incomplete markets. This approach relies on the constrained optimization methods presented in the preceding section, and so it lends itself well to the introduction of short sales restrictions and/or similar constraints. The idea is to introduce fictitious securities to the market in such a way as to make the model complete, and then use constraints to prohibit any positions in these fictitious securities. This *fictitious security* approach will be described later in this section, after examining the risk neutral computational approach.

A key to the identification of the set of attainable wealths is principle (1.23), which says that a contingent claim (i.e., wealth) $W$ is attainable if and only if $E_Q[W/B_1]$ takes the same value for every risk neutral probability measure $Q \in \mathbb{M}$. Thus $\mathbb{W}_v$, the set of wealths that can be generated starting with initial capital $v$, is given by $\mathbb{W}_v = \{\mathbb{W} \in \mathbb{R}^K : E_Q[W/B_1] = v, \text{ all } Q \in \mathbb{M}\}$. But this characterization of $\mathbb{W}_v$ is not practical, because with incomplete models the set $\mathbb{M}$ of risk neutral probability measures contains an infinite number of elements. This difficulty is resolved with the help of relationship (1.17), which says that $\mathbb{M}$ is the intersection of a linear subspace and the set of strictly positive probability measures. In particular, there exists a finite number of independent vectors in $\bar{\mathbb{M}}$ (the closure of $\mathbb{M}$), say $Q(1), Q(2), \ldots, Q(J)$, such that each element of $\mathbb{M}$ can be expressed as a linear combination of these $J$ vectors (see the linear program (1.25)), where the weights, some of which can be zero or negative, add up to one. Hence $E_Q[W/B_1] = v$ for all $Q \in \mathbb{M}$ if and only if $E_{Q(j)}[W/B_1] = v$ for $j = 1, \ldots, J$, and so

$$\mathbb{W}_v = \{W \in \mathbb{R}^K : E_{Q(j)}[W/B_1] = v \quad \text{for} \quad j = 1, \ldots, J\}$$

It follows that the optimal portfolio problem (2.1) (or (2.9)) can be written as

$$\textbf{(2.50)} \qquad \begin{aligned} &\text{maximize} \quad Eu(W) \\ &\text{subject to} \quad E_{Q(j)}[W/B_1] = v, \quad j = 1, \ldots, J \end{aligned}$$

As in section 2.2, problem (2.50) can be solved by introducing $J$ Lagrange multipliers as well as $J$ corresponding state price densities $L_j \equiv Q(j)/P$:

$$\textbf{(2.51)} \qquad \text{maximize} \quad Eu(W) - \sum_{j=1}^{J} \lambda_j E[L_j W/B_1]$$

The first order necessary conditions, one for each $\omega \in \Omega$, are:

$$u'(W(\omega)) = \sum_{j=1}^{J} \lambda_j L_j(\omega)/B_1(\omega), \qquad \text{all} \quad \omega \in \Omega$$

or

$$W(\omega) = I\left[\sum_{j=1}^{J} \lambda_j L_j(\omega)/B_1(\omega)\right], \qquad \text{all} \quad \omega \in \Omega \tag{2.52}$$

where $I(\cdot)$ is the inverse function of $u'$. This gives the solution of (2.51) as a function of the $J$ Lagrange multipliers. Substituting this expression into the $J$ constraints of (2.50) enables one to solve for the values of the Lagrange multipliers which provide the solution of (2.50). In other words, substituting the values of the Lagrange multipliers satisfying

$$E[L_j I(\lambda_1 L_1/B_1 + \ldots + \lambda_J L_J/B_1)/B_1] = v, \quad j = 1, \ldots, J \tag{2.53}$$

into (2.52) provides the optimal solution of (2.50). From this solution, the optimal attainable wealth, one finally computes the optimal trading strategy in the usual way.

The system (2.53) will normally have a unique, non-negative solution, depending upon the properties of the utility function. If the utility function is strictly concave, then the solution of (2.50) will be unique. This computational procedure will now be illustrated with an example.

**Example 2.7** The securities model is the same as in example 1.2, namely, $K = 3$, $N = 1$, $r = 1/9$, $S_0 = 5$, and

| $\omega$ | $S_1(\omega)$ | $S_1^*(\omega)$ | $P(\omega)$ |
|---|---|---|---|
| $\omega_1$ | 20/3 | 6 | 1/3 |
| $\omega_2$ | 40/9 | 4 | 1/3 |
| $\omega_3$ | 30/9 | 3 | 1/3 |

In chapter 1 it was established that this model is incomplete with $\mathbb{M}$ consisting of all probability measures of the form

$$Q = (\theta, 2 - 3\theta, -1 + 2\theta), \qquad \text{where} \quad \frac{1}{2} < \theta < \frac{2}{3}$$

and with a contingent claim $X = (X_1, X_2, X_3)$ being attainable if and only if

$$X_1 - 3X_2 + 2X_3 = 0 \tag{2.54}$$

In this case a 'basis' for $\mathbb{M}$ can be obtained by taking any two distinct elements of $\mathbb{M}$. In fact, one can take the two endpoints, corresponding to $\theta = 1/2$ and $\theta = 2/3$, and this is what we will do:

$$Q(1) = (1/2, 1/2, 0) \quad Q(2) = (2/3, 0, 1/3)$$
$$L_1 = (3/2, 3/2, 0) \qquad L_2 = (2, 0, 1)$$



Taking $u(w) = \ln(w)$ one has $u'(w) = 1/w$ and $I(i) = 1/i$, so system (2.53) becomes, after a little algebra,

$$\frac{1}{3\lambda_1 + 4\lambda_2} + \frac{1}{3\lambda_1} = v$$
$$\frac{4}{9\lambda_1 + 12\lambda_2} + \frac{1}{3\lambda_2} = v$$

The unique, non-negative solution is found to be

$$\lambda_1 = 0.46482\, v^{-1} \qquad \lambda_2 = 0.53519\, v^{-1}$$

Substituting these values into (2.52) yields

$$W(\omega) = \frac{v}{0.46482(9/10)L_1(\omega) + 0.53519(9/10)L_2(\omega)}$$

$$= \begin{cases} 0.62860\ v, & \omega = \omega_1 \\ 1.59360\ v, & \omega = \omega_2 \\ 2.07611\ v, & \omega = \omega_3 \end{cases}$$

for the optimal attainable wealth. Notice, as verification, that $W$ satisfies equation (2.54). Solving $H_0 + 6H_1 = (9/10)(0.6286)v$ and $H_0 + 4H_1 = (9/10)(1.5936)v$ for the optimal trading strategy yields

$$H_0 = 3.17124\ v \quad \text{and} \quad H_1 = -0.43425\ v$$

The optimal objective value is $0.24409 + \ln v$.

In summary, the risk neutral computational approach for incomplete models is essentially the same as for complete models, but the computational difficulties are increased due to the need to first specify and then cope with the additional constraints in (2.50).

We now turn to an alternative computational approach that features fictitious securities. The idea is to add one or more securities to the model in such a way as to make it complete (without, of course, creating any arbitrage opportunities). Then one solves, using the methods of the preceding section, the optimal portfolio problem with the constraint that no position can be taken in any of the added, fictitious securities. Since this optimization problem is done for a complete market, the computations may be simpler than with the two alternative approaches, even with the constraints.

While this concept is simple, a key step is to properly specify the added, fictitious securities. A good way to do this is to work with the $K \times N$ matrix $A$ of chapter 1:

$$A = \begin{bmatrix} B_1(\omega_1) & S_1(1)(\omega_1) & \dots & S_N(1)(\omega_1) \\ B_1(\omega_2) & S_1(1)(\omega_2) & \dots & S_N(1)(\omega_2) \\ \vdots & \vdots & & \vdots \\ B_1(\omega_K) & S_1(1)(\omega_K) & \dots & S_N(1)(\omega_K) \end{bmatrix}$$

This matrix has rank less than $K$, since the market is incomplete. We need to add some fictitious securities, that is, column vectors of non-negative numbers to $A$, so that the rank of $A$ becomes equal to $K$. When selecting column vectors one must be careful to avoid adding arbitrage opportunities. A little linear algebra will ensure a successful result, as illustrated in the following example.

**Example 2.7 (continued)** It suffices to add one fictitious security, so that the matrix $A$ has the form

$$A = \begin{bmatrix} 10/9 & 60/9 & S_2(1)(\omega_1) \\ 10/9 & 40/9 & S_2(1)(\omega_2) \\ 10/9 & 30/9 & S_2(1)(\omega_3) \end{bmatrix}$$

Taking, for instance, $S_2(1) = (50/9, 20/9, 70/9)$, it is easy to verify that the matrix $A$ will have full rank 3. Since all the risk neutral probabilities in the original market satisfy $Q = (\theta, 2 - 3\theta, -1 + 2\theta)$ for $1/2 < \theta < 2/3$, it follows that the unique risk neutral probability measure in the new market must be of this form as well. Taking, for instance, $\theta = 7/13$ gives $Q = (7/13, 5/13, 1/13)$ as well as $S_2(0) = E_Q[(9/10)S_2(1)] = 4$.

It remains to solve the optimal portfolio problem in the new market with the constraint that positions in security #2 are prohibited. Taking the approach described in the preceding section, this means that $\mathbb{K} = \{F \in \mathbb{R}^2 : F_2 = 0\}$,

$$\delta(\kappa) = \sup_{F \in \mathbb{K}} (-F\kappa') = \sup_{F_1 \in \mathbb{R}} (-F_1\kappa_1) = \begin{cases} 0, & \kappa_1 = 0 \\ \infty & \text{otherwise} \end{cases}$$

and $\tilde{\mathbb{K}} = \{\kappa \in \mathbb{R}^2 : \kappa_1 = 0\}$. The return processes in the market $\mathcal{M}_\kappa$ are

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $R_1(\omega)$ | $1/3$ | $-1/9$ | $-1/3$ |
| $R_2(\omega)$ | $7/18 + \kappa_2$ | $-4/9 + \kappa_2$ | $17/18 + \kappa_2$ |

and the corresponding risk neutral probability measure is computed to be

$$Q_\kappa(\omega) = \begin{cases} 7/13 - (18/65)\kappa_2, & \omega = \omega_1 \\ 5/13 + (54/65)\kappa_2, & \omega = \omega_2 \\ 1/13 - (36/65)\kappa_2, & \omega = \omega_3 \end{cases}$$



Notice that these probabilities are strictly positive as long as $-25/54 < \kappa_2 < 5/36$. Since $B_1^\kappa = 10/9$ and we are still using log utility, the optimal objective value for the unconstrained problem in the market $\mathcal{M}_\kappa$ is $J_\kappa(v) = \ln v - \ln(9/10) - E\ln Q_\kappa + E\ln P$. The dual problem is therefore the same as maximizing $E\ln Q_\kappa$ over $\tilde{\mathbb{K}}$, that is, with respect to $\kappa_2$ over the interval $(-25/54, 5/36)$. Some simple calculus provides the optimal solution: $\hat{\kappa} = (0, -0.18321)$.

The corresponding optimal attainable wealth is

$$W_{\hat{\kappa}} = vB_1^{\hat{\kappa}}/L_{\hat{\kappa}} = \frac{10v}{27Q_{\hat{\kappa}}} = \begin{cases} 0.62860\ v & \omega = \omega_1 \\ 1.59360\ v, & \omega = \omega_2 \\ 2.07611\ v, & \omega = \omega_3 \end{cases}$$

which is seen to be generated by the trading strategy $F = (-2.17125, 0)$. Since $F \in \mathbb{K}$ and $\delta(\hat{\kappa}) + F\hat{\kappa}' = 0$, it follows that this is also the optimal solution for the constrained problem as well as for the original unconstrained problem in the incomplete market.

A virtue of the fictitious securities approach is that it readily lends itself to problems which have short sales restrictions or similar constraints on the real securities. One proceeds in exactly the same way, only choosing the constraint set $\mathbb{K}$ so as to capture the explicit constraints on the real securities as well as the prohibition from taking a position in the fictitious securities. A return to the same example will illustrate this.

**Example 2.7 (continued)** Suppose we prohibit short sales, so the solution obtained earlier is now infeasible. With the fictitious security the same as before, take $\mathbb{K} = \{F \in \mathbb{R}^2 : F_1 \geqslant 0, F_2 = 0\}$, so that

$$\delta(\kappa) = \sup_{F \in \mathbb{K}} (-F\kappa') = \sup_{F_1 \geqslant 0} (-F_1\kappa_1) = \begin{cases} 0, & \kappa_1 \geqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

and $\tilde{\mathbb{K}} = \{\kappa \in \mathbb{R}^2 : \kappa_1 \geqslant 0\}$. The return processes in the market $\mathcal{M}_\kappa$ are

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $R_1(\omega)$ | $1/3 + \kappa_1$ | $-1/9 + \kappa_1$ | $-1/3 + \kappa_1$ |
| $R_2(\omega)$ | $7/18 + \kappa_2$ | $-4/9 + \kappa_2$ | $17/18 + \kappa_2$ |

and the corresponding risk neutral probability measure is

$$Q_\kappa(\omega) = \begin{cases} 7/13 - (45/26)\kappa_1 - (18/65)\kappa_2, & \omega = \omega_1 \\ 5/13 + (9/13)\kappa_1 \ + (54/65)\kappa_2, & \omega = \omega_2 \\ 1/13 + (27/26)\kappa_1 - (36/65)\kappa_2, & \omega = \omega_3 \end{cases}$$

Note these probabilities are strictly positive on the triangular subset of $\mathbb{R}^2$ where $\kappa_2 < -(25/4)\kappa_1 + 35/18$, $\kappa_2 > -(5/6)\kappa_1 - 25/54$, and

$\kappa_2 < (15/8)\kappa_1 + 5/36$. The optimal objective value for the unconstrained problem in the market $\mathscr{M}_\kappa$ has the same form as before, so the dual problem comes down to maximizing $E \ln Q_\kappa$ on the intersection of this triangular subset with the half-plane where $\kappa_1 \geqslant 0$. Using the first order conditions, it is easy to verify that $\hat{\kappa} = (4/27, -5/27)$ is the optimal solution. Corresponding to this are $Q_{\hat{\kappa}} = (1/3, 1/3, 1/3)$ and $W_{\hat{\kappa}} = 10v/(27Q_{\hat{\kappa}}) = (10v/9)(1,1,1)$, both constants on $\Omega$. The trading strategy that generates $W_{\hat{\kappa}}$ is easily computed to be $F = (0,0)$, that is, invest all the money in the bank account. Clearly $F \in \mathbb{K}$ and $\delta(\hat{\kappa}) + F\hat{\kappa}' = 0$, so $F$ is also the optimal trading strategy for the original constrained optimal portfolio problem, which is what we suspected from the start. The optimal objective value is $\ln v + \ln(10/9)$.

**Exercise 2.12** Solve example 2.7 with $P(\omega_1) = 0.5$, $P(\omega_2) = 0.3$, and $P(\omega_3) = 0.2$, assuming

(a) short sales are allowed,
(b) short sales are prohibited.

## 2.7 Equilibrium Models

Until now, the specification of the security price processes $S_1, S_2, \ldots, S_N$, has been part of the data, external to the model. But it is important to understand prices, and so financial economists develop and study models where the price processes are internal, that is, endogenous.

An important category of models of this type is the class of *equilibrium* models. Sometimes the prices at both time $t = 0$ and time $t = 1$ are internal. Other times, and this is the kind of equilibrium model that will be looked at in this section, the prices at time $t = 1$ are specified and only the prices at time $t = 0$ are internal.

The data for the one-period equilibrium model will consist of the sample space $\Omega$, the probability measure $P$, the bank account process $B$, and the $N$ random variables $S_1(1), S_2(1), \ldots, S_N(1)$ representing the time $t = 1$ prices of the risky securities. In addition, there are $I$ investors (or traders or consumers), numbered $i = 1, 2, \ldots, I$. Corresponding to each trader is a utility function $u_i$ (differentiable, concave, strictly increasing) and an endowment process $(v_i, E_i)$.

Internal to the model are three kinds of variables: the time $t = 0$ security prices $S_1(0), S_2(0), \ldots, S_N(0)$; a consumption process $C^i = (C_0^i, C_1^i)$ for each investor; and a trading strategy $H^i = (H_0^i, H_1^i, \ldots H_N^i)$ for each investor. The *equilibrium solution concept* involves finding values of all these variables such that a set of internally consistent conditions is satisfied. In particular,



The variables $S_n(0)$, $n = 1, \ldots, N$, and $\{C^i, H^i\}$, $i = 1, \ldots, I$, are said to be an *equilibrium solution* if for each $i$ the consumption investment plan $(C^i, H^i)$ is optimal for investor $i$, that is, $(C^i, H^i)$ is a solution of

$$\textbf{(2.55)}\qquad \begin{aligned} &\text{maximize} && u_i(C_0^i) + E[u_i(C_1^i)] \\ &\text{subject to} && C_0^i + H_0^i B_0 + \sum_{n=1}^{N} H_n^i S_n(0) = v_i \\ &&& C_1^i - H_0^i B_1 - \sum_{n=1}^{n} H_n^i S_n(1) = E_i \\ &&& H^i \in \mathbb{R}^{N+1} \end{aligned}$$

and the security market clears, that is, the aggregate demand for each security is zero, that is

$$\textbf{(2.56)}\qquad \sum_{i=1}^{I} H_n^i = 0 \qquad \text{for} \quad n = 0, 1, \ldots, N$$

Note that (2.55) does not include any explicit constraints requiring the consumption to be non-negative; if negative consumption is a problem, then one could specify utility functions that would force the consumption to be non-negative. It is possible to add explicit non-negativity constraints, but doing so would make the analysis of the equilibrium model more complicated. The requirement (2.56) that aggregate demand be zero does not make much sense for securities such as stocks and bonds, but it does hold perfectly well for things like futures contracts. Alternatively, one can imagine that some individuals act as firms, raising capital by selling stocks and bonds, and investing in the technologies that produce returns. These individuals are short the securities while everybody else is long. In the aggregate, the net positions in the securities are zero, and real aggregate wealth is equal to the total investment in the fundamental technologies.

Since the traders have strictly increasing utility functions, if there exists a solution to the equilibrium problem, then by (2.21) there must exist a risk neutral probability measure, say $Q$. It follows, therefore, that the time $t = 0$ prices must satisfy

$$S_n(0) = E_Q[S_n(1)/B_1]$$

Hence if we can derive the equilibrium consumption processes, then everything else will fall into place: (2.21) and the preceding equation will provide time zero prices, and investor $i$'s trading strategy $H^i$ will be the one which generates the contingent claim $C_1^i - E_i$ (assuming $C_1^i - E_i$ is attainable for all $i$, which will be the case if the model is complete).

Unfortunately, it is rather difficult to compute equilibrium consumption processes. In fact, an equilibrium solution does not necessarily exist, in general, so we shall not attempt to compute one. We shall need to be content

with a study of the relationship between the equilibrium solution and something called Pareto efficiency.

But first notice that if you add up the time $t=0$ budget constraint in (2.55) across $i$ and rearrange terms you get

$$B_0 \sum_{i=1}^{I} H_0^i + \sum_{n=1}^{N} S_n(0) \sum_{i=1}^{I} H_n^i = \sum_{i=1}^{I} v_i - \sum_{i=1}^{I} C_0^i$$

In view of (2.56), if this is an equilibrium solution, then the left hand side equals zero, in which case the same can be said for the right hand side. One obtains a similar conclusion from the time $t=1$ budget constraint in (2.55), and so

**(2.57)** If the consumption processes $C^i$, $i=1,\ldots,I$, are part of an equilibrium solution, then

$$\sum_{i=1}^{I} C_0^i = \sum_{i=1}^{I} v_i \quad \text{and} \quad \sum_{i=1}^{I} C_1^i = \sum_{i=1}^{I} E_i$$

A collection of consumption processes satisfying these two equations is said to be *feasible*. In other words, the aggregate consumption equals the aggregate endowment, which is a kind of budget constraint.

The collection $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ of consumption processes is said to be *Pareto efficient* if they are feasible (as in (2.57)) and there is no other collection $\{C^1, C^2, \ldots, C^I\}$ of feasible consumption processes such that

$$\textbf{(2.58)} \qquad u_i(C_0^i) + Eu_i(C_1^i) \geqslant u_i(\hat{C}_0^1) + Eu_i(\hat{C}_1^i), \ i=1,\ldots,I$$

with this inequality being strict for at least one $i$.

The condition for Pareto efficiency says that there is no feasible collection of consumption processes such that all of the investors are just as happy as they would be under the feasible collection $\{\hat{C}^1, \hat{C}^2, \ldots \hat{C}^I\}$, with at least one being strictly happier. Hence one might conjecture that a necessary and sufficient condition for $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ to be part of an equilibrium solution is that it be Pareto efficient. While this is not exactly right, we do have the following:

**(2.59)** If the model is complete and $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ is part of an equilibrium solution, then $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ is Pareto efficient.

To see why this is true, suppose $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ is part of an equilibrium solution, but there exists a feasible collection $\{C^1, C^2, \ldots, C^I\}$ of consumption processes as in (2.58), with at least one inequality being strict. I will show this leads to a contradiction. Since the model is complete, for each investor $i$ there exists a trading strategy $H^i$ satisfying

$$\textbf{(2.60)} \qquad H_0^i B_1 + \sum_{n=1}^{N} H_n^i S_n(1) = C_1^i - E_i$$



In view of (2.56), this means that

$$0 = \sum_{i=1}^{I} \left[ H_0^i B_1 + \sum_{n=1}^{N} H_n^i S_n(1) \right] = \left( \sum_{i=1}^{I} H_0^i \right) B_1 + \sum_{n=1}^{N} \left( \sum_{i=1}^{I} H_n^i \right) S_n(1)$$

So defining a new trading strategy $\tilde{H}$ by

$$\tilde{H}_n = \sum_{i=1}^{I} H_n^i, \qquad n=0,1,\ldots,N$$

we see that the time $t=1$ value of the portfolio corresponding to $\tilde{H}$ is identical to zero. Since there are no arbitrage opportunities (recall that there must exist a risk neutral probability measure), by the law of one price the time $t=0$ value of this portfolio must be zero, that is,

$$\textbf{(2.61)} \qquad 0 = \tilde{H}_0 B_0 + \sum_{n=1}^{N} \tilde{H}_n S_n(0) = \left( \sum_{i=1}^{I} H_0^i \right) B_0 + \sum_{n=1}^{N} \left( \sum_{i=1}^{I} H_n^i \right) S_n(0)$$

This equation will be used in a moment. Meanwhile, define the scalars

$$\textbf{(2.62)} \qquad \psi_i \equiv C_0^i - v_i + H_0^i B_0 + \sum_{n=1}^{N} H_n^i S_n(0), \qquad i=1,\ldots,I$$

Think of $C_0^i - \psi_i$ as time $t=0$ consumption for investor $i$. The consumption process $\{C_0^i - \psi_i, C_1^i\}$ is attainable by (2.60) and (2.62). If $\psi_i < 0$, then investor $i$ strictly prefers the consumption process $\{C_0^i - \psi_i, C_1^i\}$ to the consumption process $\{C_0^i, C_1^i\}$. Moreover, $\{C_0^i - \psi_i, C_1^i\}$ satisfies the budget constraints in investor $i$'s optimization problem (2.55), and $\{C_0^i, C_1^i\}$ is preferred to $\{\hat{C}_0^i, \hat{C}_1^i\}$ by inequality (2.58). Hence $\psi_1 < 0$ would imply $\{C_0^i, C_1^i\}$ is strictly preferred to $\{\hat{C}_0^i, \hat{C}_1^i\}$, thereby contradicting the fact that the latter is an optimal solution of (2.55). It must be that $\psi_i \geqslant 0$ for all $i=1,\ldots,I$.

By almost the same logic, if inequality (2.58) for investor $i$ is strict, then $\psi_i = 0$ leads to a contradiction, and so this same $\psi_i$ must be strictly positive. Since at least one inequality in (2.58) is supposed to be strict, if we sum equation (2.62) across $i$, we see that both sides of the resulting equation must be strictly positive. Using (2.61), we see that the right hand side is simply

$$\sum_{i=1}^{I} C_0^i - \sum_{i=1}^{I} v_i > 0$$

But this contradicts the supposition that $\{C^1, C^2, \ldots, C^I\}$ satisfy the feasibility requirement (2.57); we conclude the collection $\{\hat{C}^1, \hat{C}^2, \ldots, \hat{C}^I\}$ must be Pareto efficient.

In summary, if the market is complete, then a necessary condition for a collection of consumption processes to be part of an equilibrium solution is that the collection be Pareto efficient. Hence to compute an equilibrium solution, a reasonable approach might be first to identify all the Pareto

efficient collections of consumption processes and then to search among these for one that is part of an equilibrium solution. Unfortunately, this approach is easier said than done, in general, and so this idea will not be pursued any further. Instead, we will study a variation of the equilibrium problem for which it is easier to compute equilibrium solutions.

Suppose we simplify the basic equilibrium problem by stipulating that $C_0^i = 0$ and $E_i = 0$ for every investor $i = 1, \ldots, I$ and ignoring the utility of time $t = 0$ consumption, so that (2.55) for each investor is simply a standard optimal portfolio problem. Furthermore, condition (2.56) for an equilibrium solution is replaced by

$$\text{(2.63)} \qquad \sum_{i=1}^{I} H_n^i = s_n, \qquad \text{for} \quad n = 1, \ldots, N$$

where $s_n > 0$ represents the total supply or number of units or shares of security $n$ that are present in the market. The equilibrium problem can now be thought of as a market where all the investors share common beliefs about the time $t = 1$ prices in each of the states, and the question is, 'What are the appropriate time $t = 0$ prices?' Or perhaps $N$ companies are making initial public offerings of their securities, they can assess the correct time $t = 1$ prices of their securities in each of the states, and they want to set the time $t = 0$ offering prices properly. In any event, the equilibrium solution will now consist of time $t = 0$ prices and a trading strategy for each investor such that the optimal portfolio problem (2.55) is satisfied for each investor and the market clearing condition (2.63) is satisfied.

To solve this kind of problem, a good approach is first to compute $H^i(S_0) = \{H_1^i(S_0), \ldots, H_N^i(S_0)\}$, the optimal solution of the portfolio problem (2.55) as a function of the specified time $t = 0$ prices $S_0 = \{S_1(0), \ldots, S_N(0)\}$. Thus $S_0 \rightarrow H_n^i(S_0)$ should be thought of as investor $i$'s demand 'curve' or *demand function* for security $n$. Knowing these demand functions for all $i$ and $n$, it remains to substitute them in (2.63) and solve for a time $t = 0$ price vector $S_0$ such that the market clearing condition (2.63) is satisfied.

We know something about the demand functions right away: they are finite at the point $S_0$ if and only if there exists a risk neutral probability measure at the point $S_0$. This is because if there is no risk neutral probability measure, then there is some arbitrage opportunity, in which case the investors would find it desirable to buy long or sell short an infinite quantity of one or more of the securities. It follows that the region where the demand functions are all finite, which will be denoted $\mathbb{S}$, is a subset of the $n$-dimensional interval

$$\mathop{\mathrm{X}}_{n=1}^{N} [\min\{S_n^*(1)(\omega) : \omega \in \Omega\}, \ \max\{S_n^*(1)(\omega) : \omega \in \Omega\}]$$

If $N = 1$, then this interval coincides with $\mathbb{S}$. With $N \geqslant 2$, $\mathbb{S}$ can either coincide with the $n$-dimensional interval or be a proper subset of its interior.



Consideration of this interval helps to organize the computation of the demand functions.

If the utility functions are suitably smooth with $u'(\cdot)$ taking all the nonnegative values on the real line, then the demand functions $S_0 \rightarrow H^i(S_0)$ are continuous on $\mathbb{S}$. This is because the first order necessary conditions (2.4) are satisfied for each investor $i$, namely,

$$\text{(2.64)} \qquad \begin{aligned} 0 = E[B_1 u'(v_i + H_1^i\{S_1^*(1) - S_1(0)\} + \ldots + H_n^i\{S_n^*(1) \\ - S_n(0)\})\{S_n^*(1) - S_n(0)\}], \qquad n = 1, \ldots, N \end{aligned}$$

(recall $S_n(0) = S_n^*(0)$). These equations are satisfied by some $H^i$ for each $S_0$ in $\mathbb{S}$, and with modest assumptions the solution $H^i(S_0)$ will vary in a continuous fashion with respect to $S_0$ by what is called the implicit function theorem.[2] Moreover, the absolute value of one or more components of the demand functions will become arbitrarily large as $S_0$ approaches the boundary of $\mathbb{S}$. For example, in the case $N = 1$ the function $H_1^i(S_0)$ becomes arbitrarily large (small) as $S_0$ approaches the lower (respectively, upper) endpoint of the interval $\mathbb{S}$.

As stated above, to compute an equilibrium solution the recommended approach is first to compute the demand functions and then to substitute them in the market clearing condition (2.63) to solve for $S_0$. But one should be warned that this recipe will usually entail some nasty calculations; in fact, this approach is not guaranteed to be successful. It is instructive to now look at an example.

**Example 2.8** Suppose $N = 2$, $K = 3$, $P = (1/3, 1/3, 1/3)$, $r =$ constant, and there are $I$ identical investors with $u_i(w) = \ln(w)$ and $v_i = v$. The time $t = 1$ discounted prices are:

| n | $S_n^*(1)$ | | |
|---|---|---|---|
| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
| 1 | 6 | 8 | 4 |
| 2 | 13 | 9 | 8 |

Recalling the matrix $A$ that was studied in chapter 1, it is apparent that if there is a risk neutral probability measure, then this market must be complete. To compute the risk neutral probability measure as a function of the still unknown time $t = 0$ prices, one solves the usual system of equations and obtains

$$Q(\omega) = \begin{cases} \dfrac{-28 - S_1(0) + 4S_2(0)}{18}, & \omega = \omega_1 \\ \dfrac{-4 + 5S_1(0) - 2S_2(0)}{18}, & \omega = \omega_2 \\ \dfrac{50 - 4S_1(0) - 2S_2(0)}{18}, & \omega = \omega_3 \end{cases}$$

The region where these three fractions are all strictly positive coincides with the region $\mathbb{S}$ where the demand functions are finite; some simple algebra reveals this to be the interior of the triangle with vertices at (4,8), (6,13), and (8,9). With log utility the optimal attainable wealth is of the form $W = vP(1+r)/Q$ (see exercise 2.2), so solving the system $H_0(1+r) + H_1S_1(\omega) + H_2S_2(\omega) = W(\omega)$, one obtains the demand functions

$$H_1^i(S_0) = \frac{-(1/3)v}{-28 - S_1(0) + 4S_2(0)} - \frac{(4/3)v}{50 - 4S_1(0) - 2S_2(0)} + \frac{(5/3)v}{-4 + 5S_1(0) - 2S_2(0)}$$

$$H_2^i(S_0) = \frac{(4/3)v}{-28 - S_1(0) + 4S_2(0)} - \frac{(2/3)v}{50 - 4S_1(0) - 2S_2(0)} - \frac{(2/3)v}{-4 + 5S_1(0) - 2S_2(0)}$$

Hence knowing the values $s_1$, $s_2$, $I$, and $v$, one can substitute these two functions in the two market clearing equations (2.63) and solve for the two unknown time $t = 0$ prices. For example, if there are $I = 2$ investors each having $v = \$6000$ to invest, and if there are available $s_1 = 4000$ and $s_2 = 2000$ shares of securities 1 and 2, respectively, then system (2.63) yields $S_1(0) = 5$ and $S_2(0) = 9$. Substituting these values back into the demand functions gives $H_1^i = 2000$ and $H_2^i = 1000$, numbers which are as anticipated, since with $I$ identical investors the equilibrium trading strategies will necessarily satisfy $H_n^i = s_n/I$. Note that $H_0 = v - H_1S_1(0) - H_2S_2(0) = -13{,}000$, so each investor will borrow \$13,000 in order to finance these transactions.

**Exercise 2.13** Suppose $N = 1$, $K = 2$, $r =$ constant, $S_1(\omega_1) = 6$, $S_1(\omega_2) = 4$, and $P(\omega_1) = 2/3$. There are $I$ identical investors, each with initial capital $v$ and with log utility preferences.

(a) Show that the risk neutral probability measure must be of the form $Q(\omega_1) = [S_0(1+r) - 4]/2$ with $\mathbb{S} = (4/(1+r), 6/(1+r))$.

(b) Show that the demand function is

$$H(S_0) = \frac{v(1+r)\{3S_0(1+r) - 16\}}{3\{4 - S_0(1+r)\}\{6 - S_0(1+r)\}}$$

and is strictly decreasing on $\mathbb{S}$.

(c) Derive a formula for the equilibrium price $S_0$ in terms of general parameters $I$, $v$, $s_1$, and $r$. What is the equilibrium price when $I = 3$, $v = s_1 = 1000$, and $r = 0$?



## NOTES

1 An *affine* function is equal to a constant plus a linear function.

2 For the situation here, suppose the $N$ partial derivatives of the right hand side of (2.64) with respect to each $H_n^i$ as well as to each $S_n(0)$ is a continuous function. Moreover, suppose the determinant of the $N \times N$ matrix of the partial derivatives of the right hand side of (2.64) with respect to each $H_n^i, n = 1, \ldots, N$ (this is called a *Jacobian*), is non-zero at some point $\hat{S}_0$ where (2.64) is satisfied. Then the implicit function theorem says there exist continuous functions $H_1^i(S_0), \ldots, H_N^i(S_0)$ such that, when substituted in (2.64), this equation is satisfied for all $S_0$ in some neighborhood of $\hat{S}_0$.

# 3 Multiperiod Securities Markets



## 3.1 Model Specifications, Filtrations, and Stochastic Processes

Multiperiod models of securities markets are much more realistic than single period models. In fact, they are extensively used for practical purposes in the financial industry.

The following elements of the basic, multiperiod model are specified as data:

- $T+1$ trading dates: $t = 0, 1, \ldots, T$.
- A finite sample space $\Omega$ with $K < \infty$ elements:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$$

- A probability measure $P$ on $\Omega$ with $P(\omega) > 0$ for all $\omega \in \Omega$.
- A *filtration* $\mathbb{F} = \{\mathscr{F}_t; t = 0, 1, \ldots, T\}$, which is a submodel describing how the information about the security prices is revealed to the investors.
- A *bank account* process $B = \{B_t; t = 0, 1, \ldots, T\}$, where $B$ is a stochastic process with $B_0 = 1$ and with $B_t(\omega) > 0$ for all $t$ and $\omega$. Here $B_t$ should be thought of as the time $t$ value of a savings account when \$1 is deposited at time 0. Usually $B$ is a non-decreasing process, and the (possibly random) quantity $r_t \equiv (B_t - B_{t-1})/B_{t-1} \geqslant 0$, $t = 1, \ldots, T$, should be thought of as the *interest rate* pertaining to the time interval $(t-1, t)$.



- $N$ *risky security* processes $S_n = \{S_n(t); t = 0, 1, \ldots T\}$, where $S_n$ is a non-negative stochastic process for each $n = 1, 2, \ldots N$. Here $S_n(t)$ should be thought of as the time $t$ price of risky security $n$, for example, the price of one share of common stock of a particular corporation.

Note the multiperiod securities market model has two new features not shared with single period models: the information submodel and stochastic process submodels of prices. These will now be described.

### 3.1.1 Information Structures

It is important to have a clear idea about how information concerning the security prices is revealed to the investors. This is done in terms of subsets of the sample space $\Omega$.

Consider that at time $t = 0$ every state $\omega \in \Omega$ is possible. Some states may be more likely than others, but none is ruled out. Meanwhile, at time $t = T$ (it will always be assumed) the investors learn the true state $\omega$ of the world (and thus the true value of every random variable). This is because as time evolves the investors will be able to deduce the true state by observing the information as it unfolds, since it will be assumed there exists a one-to-one correspondence between each possible sequence of information and each state.

What about the information at intermediate times, namely, when $0 < t < T$? How do we model the way the information evolves? Well, the new information observed over one time period enables the investors to rule out certain states as being impossible. Hence one can view the evolution of information as a random sequence $\{A_t\}$ of subsets of $\Omega$, where $A_0 = \Omega$, $A_T = \{\omega\}$ for some $\omega \in \Omega$ and $A_0 \supseteq A_1 \supseteq \ldots \supseteq A_{T-1} \supseteq A_T$. The investors know at time $t$ that for some subset $A_t$ the true state is some $\omega \in A_t$, but they are not sure which one it is. Some states $\omega \in A_t$ may be more likely than others, but none is ruled out. On the other hand, every state $\omega \in A_t^c$ (the complement of $A_t$) is ruled out by the investors at time $t$. The investors know the true state of the world is not outside $A_t$. Logically, one period later the relevant subset $A_{t+1}$ describing investor information must be, in turn, a subset of $A_t$. Thus the sequence $\{A_t\}$ of subsets that unfolds for the investors must satisfy $A_{t+1} \subseteq A_t$ for all $t$.

Notice there exist $K$ possible information sequences $\{A_t\}$ of subsets. At time $t = 0$ the investors are aware of all these sequences, but they do not know which one is going to unfold. Arbitrarily select one such sequence $\{\hat{A}_t\}$ and some time $s < T$, and consider the collection of all the sequences $\{A_t\}$ which coincide with $\{\hat{A}_t\}$ up through time $s$ along with $\{\hat{A}_t\}$ itself. In particular, consider all the time $s+1$ subsets $A_{s+1}$ from the sequences in this collection. If $\omega \in \hat{A}_s$, then there must exist at least one subset $A_{s+1}$ containing $\omega$ (if none of the sequences coinciding with $\{\hat{A}_t\}$ through time $s$ ends up in state $\omega$, then $\omega$ should not have been in $\hat{A}_s$ to begin with). Hence the union of all the subsets $A_{s+1}$ that can possibly follow $\hat{A}_s$ must be equal to $\hat{A}_s$. Moreover, this collection $\{A_{s+1}\}$ of subsets must be mutually exclusive

(if $\omega$, say, were contained in two distinct subsets, then there would exist two or more distinct sequences $\{A_t\}$ corresponding to state $\omega$, a contradiction). Hence the collection $\{A_{s+1}\}$ of subsets that can possibly follow $\hat{A}_s$ forms a *partition* of $\hat{A}_s$, that is, a collection of disjoint subsets whose union equals $\hat{A}_s$.

In particular, taking $s = 0$, we see that the collection $\{A_1\}$ of all possible time $t = 1$ subsets forms a partition of $\Omega$. This partition is denoted $\mathscr{P}_1$. Moreover, the collection $\{A_2\}$ of all possible time $t = 2$ subsets also forms a partition, denoted $\mathscr{P}_2$, of $\Omega$; it has the property that each $A \in \mathscr{P}_1$ is equal to the union of one or more of the elements of $\mathscr{P}_2$. It follows, therefore, that the information structure is fully described by a sequence $\mathscr{P}_0, \mathscr{P}_1, \ldots \mathscr{P}_T$ of partitions of $\Omega$, with $\mathscr{P}_0 \equiv \{\Omega\}$, $\mathscr{P}_T \equiv \{\{\omega_1\}, \ldots, \{\omega_K\}\}$, and satisfying the property that each $A \in \mathscr{P}_t$ is equal to the union of some elements in $\mathscr{P}_{t+1}$ for every $t < T$. This sequence $\{\mathscr{P}_t\}$ of partitions is uniquely constructed from the collection of possible information sequences $\{A_t\}$. Conversely, given a sequence $\{\mathscr{P}_t\}$ of partitions as above, there is a unique, corresponding collection of possible information sequences $\{A_t\}$.

There are several good ways to visualize the information structure. The sequence of partitions can, of course, be described with a sequence of pictures of the sample space, one picture for each point in time showing the corresponding partition. Alternatively, the sequence of partitions can be described with a network diagram known as a *tree*, where each node corresponds to one element $A_t$ of the time $t$ partition, and there is one arc going from this node to each node corresponding to some $A_{t+1} \subseteq A_t$. There will thus be one path from the time $t = 0$ node (i.e., $A_0 = \Omega$) to each $t = T$ node (i.e., $A_T = \omega$ for each state $\omega$), and each such path will indicate a possible information sequence $\{A_t\}$.

**Example 3.1** With $K = 8$ and $T = 3$, suppose the time $t = 1$ partition is

$$\mathscr{P}_1 = \{\{\omega_1, \omega_2, \omega_3, \omega_4\}, \quad \{\omega_5, \omega_6, \omega_7, \omega_8\}\}$$

Then for the time $t = 2$ partition we could take

$$\mathscr{P}_2 = \{\{\omega_1, \omega_2\}, \quad \{\omega_3, \omega_4\}, \quad \{\omega_5, \omega_6\}, \quad \{\omega_7, \omega_8\}\}$$

or

$$\mathscr{P}_2 = \{\{\omega_1\}, \quad \{\omega_2, \omega_3, \omega_4\}, \quad \{\omega_5, \omega_6, \omega_7, \omega_8\}\}$$

for example, but we could not take

$$\mathscr{P}_2 = \{\{\omega_1, \omega_2, \omega_3\}, \quad \{\omega_3, \omega_4\}, \quad \{\omega_5, \omega_6\}, \quad \{\omega_7, \omega_8\}\}$$

for example, because two of the subsets are not disjoint, nor could we take

$$\mathscr{P}_2 = \{\{\omega_1, \omega_2\}, \quad \{\omega_3, \omega_4, \omega_5\}, \quad \{\omega_6, \omega_7, \omega_8\}\}$$



because no union of any of these subsets equals $\{\omega_5, \omega_6, \omega_7, \omega_8\}$. Adopting the first suggestion for the time $t = 2$ partition, this example can be conveniently described by the sequence of pictures shown in figure 3.1 or by the tree diagram in figure 3.2.

In summary, the submodel of information structure can be organized as a sequence of partitions, with each successive partition becoming *finer*. Or it can be organized as a tree. There is still another way to specify the submodel.

A collection $\mathscr{F}$ of subsets of $\Omega$ is called an *algebra* on $\Omega$ if

(a) $\Omega \in \mathscr{F}$
(b) $F \in \mathscr{F} \Rightarrow F^c = \Omega \backslash F \in \mathscr{F}$
(c) $F$ and $G \in \mathscr{F} \Rightarrow F \cup G \in \mathscr{F}$.

Note that the empty set $\phi = \Omega^c$, so if $\mathscr{F}$ is an algebra, then it must contain the empty set. Note also that $F \cap G = (F^c \cup G^c)^c$, so if $\mathscr{F}$ is an algebra containing $F$ and $G$, then $\mathscr{F}$ must contain the intersection $F \cap G$ of $F$ and $G$. Hence an algebra on $\Omega$ is a family of subsets of $\Omega$ that is stable under finitely many set operations.

Given an algebra on $\Omega$, denoted $\mathscr{F}_t$, you can always find a unique collection $\{F_n\}$ of subsets $F_n$ such that

(a) each $F_n \in \mathscr{F}_t$,
(b) the subsets $\{F_n\}$ are disjoint, and
(c) the union of the subsets $\{F_n\}$ equals $\Omega$.

In other words, corresponding to the algebra $\mathscr{F}_t$ is a partition of $\Omega$, which is unique.

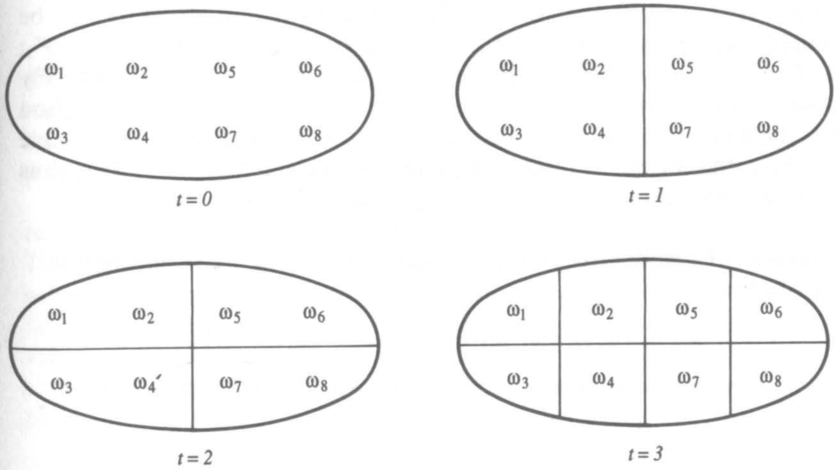


**Figure 3.1** Partitions corresponding to information submodel for example 3.1

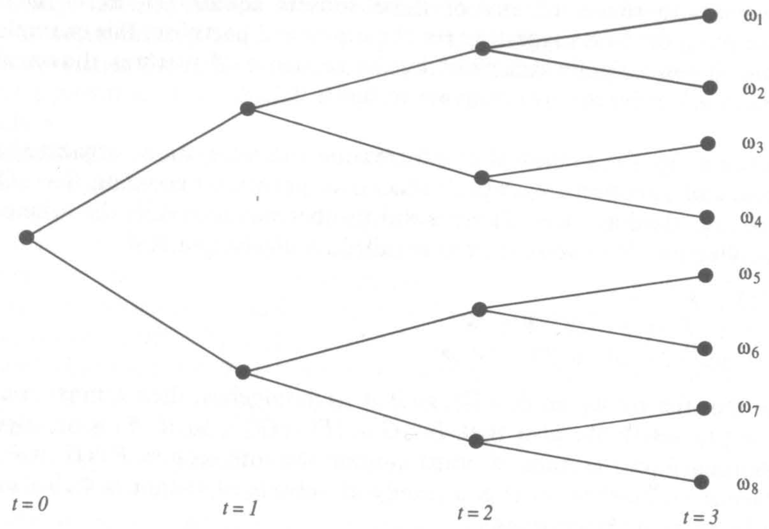


**Figure 3.2** Information tree submodel for example 3.1

Conversely, given a partition you can perform a variety of elementary set operations (taking complements, intersections, unions, and so forth), generating as many new subsets as possible. You thereby end up with an algebra, which is unique.

Hence there is a one-to-one correspondence between partitions of $\Omega$ and algebras on $\Omega$, and so the submodel of the information structure can be organized as a sequence $\{\mathscr{F}_t\}$ of algebras. We write $\mathbb{F} = \{\mathscr{F}_t; t = 0, 1, \ldots, T\}$ and call $\mathbb{F}$ a *filtration*. Note that $\mathscr{F}_0 = \{\phi, \Omega\}$ and $\mathscr{F}_T$ consists of all the subsets of $\Omega$. Since each subset in the time $t$ partition equals the union of some subsets in the time $(t+1)$ partition, we must have $\mathscr{F}_t \subseteq \mathscr{F}_{t+1}$, that is, each subset of $\mathscr{F}_t$ must be an element of $\mathscr{F}_{t+1}$. One thus can say that our filtration is a nested sequence of algebras.

**Example 3.1 (continued)** Corresponding to the time $t = 1$ partition is the algebra

$$\mathscr{F}_1 = \{\phi, \Omega, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_5, \omega_6, \omega_7, \omega_8\}\}$$

Corresponding to the time $t = 2$ partition we adopted is the algebra

$$\begin{aligned}\mathscr{F}_2 = \{&\phi, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6\}, \{\omega_7, \omega_8\}, \{\omega_1, \omega_2, \omega_3, \omega_4\},\\ &\{\omega_5, \omega_6, \omega_7, \omega_8\}, \{\omega_1, \omega_2, \omega_5, \omega_6\}, \{\omega_1, \omega_2, \omega_7, \omega_8\}, \{\omega_3, \omega_4, \omega_5, \omega_6\},\\ &\{\omega_3, \omega_4, \omega_7, \omega_8\}, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_7, \omega_8\},\\ &\{\omega_1, \omega_2, \omega_5, \omega_6, \omega_7, \omega_8\}, \{\omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}\}.\end{aligned}$$



### 3.1.2 Stochastic Process Models of Security Prices

A *stochastic process* $S_n$ is a real-valued function $S_n(t, \omega)$ of both $t$ and $\omega$. Hence the domain is $\{0, 1, \ldots, T\} \times \Omega$. For each fixed $\omega \in \Omega$, the function $t \to S_n(t, \omega)$ is called the *sample path*. For each fixed $t$, the function $\omega \to S_n(t, \omega)$ is a random variable.

For modeling purposes, we want our stochastic process model of the security prices to be consistent with the information structure. In particular, we want the information available to the investors at any point in time to include knowledge of the present and past security prices. This is accomplished by introducing the concept of measurability of random variables.

The random variable $X$ is said to be *measurable* with respect to the algebra $\mathscr{F}$ if the function $\omega \to X(\omega)$ is constant on any subset in the partition corresponding to $\mathscr{F}$. Equivalently, for every real number $x$, the subset $\{\omega \in \Omega : X(\omega) = x\}$ is an element of the algebra $\mathscr{F}$.

**Example 3.2** With $\mathscr{F}_1 = \{\phi, \Omega, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_5, \omega_6, \omega_7, \omega_8\}\}$ as in Example 3.1, suppose

$$X(\omega) = \begin{cases} 6, & \omega = \omega_1, \omega_2, \omega_3, \text{ or } \omega_4 \\ 8, & \omega = \omega_5, \omega_6, \omega_7, \text{ or } \omega_8 \end{cases}$$

and

$$Y(\omega) = \begin{cases} 1, & \omega = \omega_1, \omega_3, \omega_5, \text{ or } \omega_7 \\ 0, & \omega = \omega_2, \omega_4, \omega_6, \text{ or } \omega_8 \end{cases}$$

Then $X$ is measurable with respect to $\mathscr{F}_1$, but $Y$ is not.

A stochastic process $S_n = \{S_n(t); t = 0, 1, \ldots, T\}$ is said to be *adapted* to the filtration $\mathbb{F} = \{\mathscr{F}_t; t = 0, 1, \ldots, T\}$ if the random variable $S_n(t)$ is measurable with respect to $\mathscr{F}_t$ for every $t = 0, \ldots, T$. It will be assumed in all that follows that the price of the $n$th risky security is an adapted stochastic process $S_n$ for $n = 1, \ldots, N$, and the same for the bank account process $B$.

So how does the requirement that the stochastic processes be adapted ensure that each investor has full knowledge of the past and present prices? The investors know at time $t$ that the true state $\omega$ is contained in a particular subset in the time $t$ partition $\mathscr{P}_t$. The time $t$ price $S_n(t)$ of each security must be constant on this subset, so the investors can work out what the time $t$ values of the securities must actually be. Moreover, since the partitions form a nested sequence, the investors can infer the observed subsets in earlier partitions and thereby deduce the actual security prices at earlier times.

In summary, our securities market model will consist of security processes that are adapted to the filtration, so the investors will have full knowledge of the past and present prices. While the information and security submodels can be specified simultaneously, in practice the filtration is often specified only after first specifying the stochastic process submodel of the securities.

But starting with a specification of the stochastic processes, it is usually possible to specify two or more filtrations such that the security prices will be adapted. Some of these filtrations may be unacceptable, however, because they may be consistent with allowing the investors to look into the future. For example, if time $t+1$ prices are $\mathscr{F}_t$ measurable, then the investors know at time $t$ what the prices will be at time $t+1$. Nevertheless, there is always one filtration that corresponds to learning about the prices as time goes on, but learning nothing more. The derivation of this kind of filtration is illustrated in the following example.

**Example 3.3** Consider an investor who watches a security and knows that it is going to evolve as follows:

| $\omega_k$ | $t=0$ | $t=1$ | $t=2$ |
|---|---|---|---|
| $\omega_1$ | $S_0=5$ | $S_1=8$ | $S_2=9$ |
| $\omega_2$ | $S_0=5$ | $S_1=8$ | $S_2=6$ |
| $\omega_3$ | $S_0=5$ | $S_1=4$ | $S_2=6$ |
| $\omega_4$ | $S_0=5$ | $S_1=4$ | $S_2=3$ |

Here $N=1$, and we are using the convention that when $N=1$ the subscript can denote the time index instead of the identification of the risky security. Moreover, $T=2$ and $K=4$, so the stochastic process $S$ has been specified for every $(t,\omega)$.

Now at time $t=0$ all the investor observes is $S_0=5$; in other words, the investor does not have a clue about the true state, so $\mathscr{F}_0=\{\phi,\Omega\}$. But at time $t=1$ the investor observes either $S_1=8$ or $S_1=4$. In the former case the investor infers the true state must be either $\omega_1$ or $\omega_2$; in the latter case it must be either $\omega_3$ or $\omega_4$. Hence the relevant partition is $\{\omega_1,\omega_2\}\cup\{\omega_3,\omega_4\}$, and the corresponding algebra is

$$\mathscr{F}_1=\{\phi,\Omega,\{\omega_1,\omega_2\},\ \{\omega_3,\omega_4\}\}$$

At time $t=2$ the investor observes $S_2$ and thereby deduces the true state $\omega$ (the investor distinguishes $\omega_2$ from $\omega_3$ by remembering $S_1$). Hence the relevant partition is $\omega_1\cup\omega_2\cup\omega_3\cup\omega_4$ and $\mathscr{F}_2$ is the collection of all subsets of $\Omega$. The resulting information structure can be described as the tree in figure 3.3. Note that, indeed, the stochastic process $S$ is adapted to the filtration.

A filtration constructed in the manner illustrated in example 3.3 is said to be *generated* by the stochastic process. The resulting filtration is the coarsest one possible, that is, the various algebras have the fewest possible subsets such that the stochastic process under discussion is adapted.

But there is another useful way to construct the securities market model. You can start with a filtration submodel, based upon, perhaps, a variety of information reports. Then you add stochastic process models of security



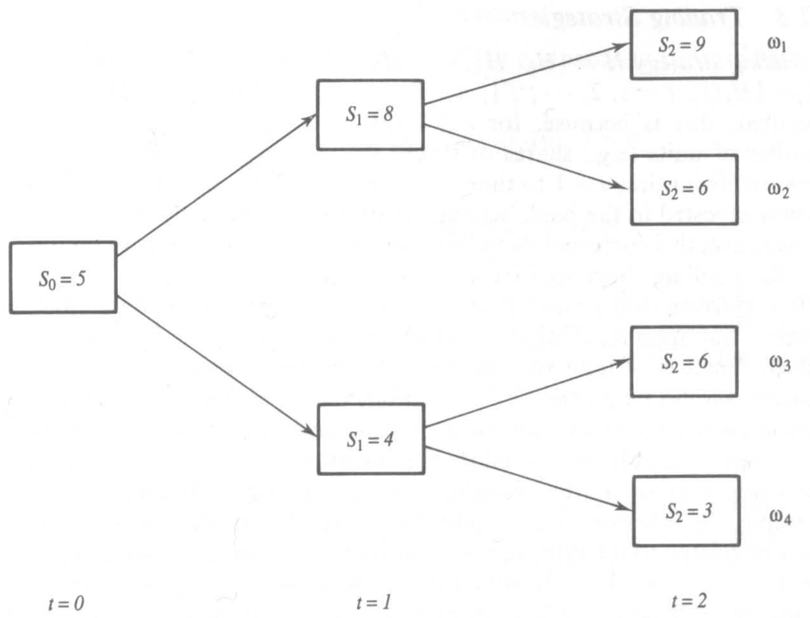


**Figure 3.3** Information structure and risky security for example 3.3

prices, making sure the processes are adapted to ensure appropriate investor knowledge of past and present prices. This is illustrated in the following example.

**Example 3.4** Suppose, with $K=4$, $N=1$, and $T=2$, that at time $t=1$ a marketing survey will be conducted that will be either favorable (corresponding to the subset $\{\omega_1,\ \omega_3\}$) or unfavorable $\{\omega_2,\ \omega_4\}$, respectively). Morever, in either case the risky security will possibly take one of two distinct values. Hence the relevant partition at time $t=1$ is $\omega_1\cup\omega_2\cup\omega_3\cup\omega_4$ and the corresponding algebra $\mathscr{F}_1$ is the collection of all subsets of $\Omega$. Note that the risky process defined in example 3.3 is adapted to this filtration and thus consistent with the information submodel. But since the investors can distinguish $\omega_1$ from $\omega_2$ as well as $\omega_3$ from $\omega_4$ at time $t=1$ by observing the marketing report, they learn more information than by only observing the price process. In fact, the investors can now look into the future, for they know at time $t=1$ what the prices will be at time $t=2$. The filtrations in examples 3.3 and 3.4 are different, although the price processes are the same. This modeling flexibility is one reason why we go to the trouble of having filtration submodels of the information structure.

### 3.1.3 Trading Strategies

A *trading strategy* $H = (H_0, H_1, \cdots, H_N)$ is a vector of stochastic processes $H_n = \{H_n(t);\ t = 1, 2, \cdots, T\}$, $n = 0, 1, \cdots, N$. Note that $H_n(0)$ is not specified; this is because, for $n \geqslant 1$, $H_n(t)$ should be interpreted as the number of units (e.g., shares of stock) that the investor owns (i.e., carries forward) from time $t-1$ to time $t$, whereas $H_0(t)B_{t-1}$ equals the amount of money invested in the bank account at time $t-1$. Note also that $H_n(t)$ can be negative; this corresponds to borrowing money from the bank (in the case $n = 0$) or selling short security $n$ (in the cases $n \geqslant 1$).

It may seem odd to model an investor's trading strategy as a stochastic process, but upon recalling that a stochastic process is little more than a real-valued function of time and the state, this begins to make sense. A trading strategy should be a rule (i.e., a function) that specifies the investor's position in each security at each point in time and in each state of the world. Moreover, this rule should allow the investor to choose a position in the securities based on all the available information, but it should not allow, for example, the investor to 'look into the future.' Hence the trading strategies must be related to the filtration submodel of the information structure in just the right way so that the investor can base the trading position on the available information, but nothing more. This is done by introducing the concept of predictibility.

A stochastic process $H_n$ is said to be *predictable* with respect to the filtration $\mathbb{F}$ if each random variable $H_n(t)$ is measurable with respect to $\mathscr{F}_{t-1}$ for all $t = 1, 2, \cdots, T$. Since $\mathscr{F}_{t-1} \subseteq \mathscr{F}_t$, this means that all predictable stochastic processes are *adapted*.

It will be assumed in all that follows that each component of a trading strategy $H$ is a predictable stochastic process. Since the trading position $H_n(t)$ established by the investor at time $t-1$ is constant on the subset that is observed in the time $t-1$ partition $\mathscr{P}_{t-1}$, the investor can take into account all of the information available at that time, but nothing more.

**Example 3.1 (continued)** Since $H_n(1) \in \mathscr{F}_0$, the position in security $n$ carried forward from time $t = 0$ to time $t = 1$ must be the same for all $\omega \in \Omega$. At time $t = 1$ the trader can adjust this position based on the information which becomes available, that is, on the observation as to whether the true state $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Consequently, the investor can choose one value for $H_n(2, \omega)$ if $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and a second value for $H_n(2, \omega)$ if $\omega \notin \{\omega_1, \omega_2, \omega_3, \omega_4\}$. In other words, the investor's new position at time $t = 1$ can be according to any rule $H_n(2)$ with $H_n(2) \in \mathscr{F}_1$. Finally, and in a similar fashion, at time $t = 2$ the investor learns the information corresponding to $\mathscr{F}_2$; the investor can take a new position according to any rule $H_n(3)$ with $H_n(3) \in \mathscr{F}_2$; and this rule will have the property that $H_n(3, \omega_1) = H_n(3, \omega_2), \cdots$, and $H_n(3, \omega_7) = H_n(3, \omega_8)$.



### 3.1.4 Value Processes and Gains Processes

The *value process* $V = \{V_t;\ t = 0, 1, \cdots, T\}$ is a stochastic process defined by

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

Hence $V_0$ is the initial value of the portfolio and, for $t \geqslant 1$, $V_t$ is the time-$t$ value of the portfolio before any transactions are made at that same time. Note that $V$ is an adapted stochastic process (if you know the subset in $\mathscr{P}_t$, then you know $H(t)$, $B_t$, and $S_n(t)$, in which case you know $V_t$).

Denote

$$\Delta S_n(t) \equiv S_n(t) - S_n(t-1)$$

for the change in the value of the stochastic process $S_n$ between times $t-1$ and $t$. Then $H_n(t)\Delta S_n(t)$ represents the one-period gain or loss due to the ownership of $H_n(t)$ units of security $n$ between times $t-1$ and $t$. Similarly,

$$\sum_{u=1}^{t} H_n(u)\ \Delta S_n(u)$$

represents the cumulative gain or loss through time $t$ due to the investment in security $n$. This sum is an example of what is called a (discrete time) *stochastic integral*, being the weighted sum of the values of one stochastic process $(H_n)$, where the weights are given by the one-period changes of another stochastic process . Finally,

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

defines the *gains process* and represents the cumulative gain or loss through time $t$ of the portfolio. Thus $G$ is the stochastic integral of the trading strategy with respect to the price process. Note that $G = \{G_t;\ t = 1, \ldots, T\}$ is an adapted stochastic process.

**Example 3.3 (continued)** Suppose $B_t = (1+r)^t$, where $r \geqslant 0$ is a constant. Then for the value process we have $V_0 = H_0(1) + 5H_1(1)$,

$$V_1 = \begin{cases} (1+r)H_0(1) + 8H_1(1), & \omega = \omega_1, \omega_2 \\ (1+r)H_0(1) + 4H_1(1), & \omega = \omega_3, \omega_4 \end{cases}$$

and

$$V_2 = \begin{cases} (1+r)^2 H_0(2) + 9H_1(2), & \omega = \omega_1, \\ (1+r)^2 H_0(2) + 6H_1(2), & \omega = \omega_2, \omega_3 \\ (1+r)^2 H_0(2) + 3H_1(2), & \omega = \omega_4 \end{cases}$$

The gains process is given by

$$G_1 = \begin{cases} rH_0(1) + 3H_1(1), & \omega = \omega_1, \omega_2 \\ rH_0(1) - H_1(1), & \omega = \omega_3, \omega_4 \end{cases}$$

and

$$G_2 = \begin{cases} rH_0(1) + 3H_1(1) + r(1+r)H_0(2) + H_1(2), & \omega = \omega_1 \\ rH_0(1) + 3H_1(1) + r(1+r)H_0(2) - 2H_1(2), & \omega = \omega_2 \\ rH_0(1) - H_1(1) + r(1+r)H_0(2) + 2H_1(2), & \omega = \omega_3 \\ rH_0(1) - H_1(1) + r(1+r)H_0(2) - H_1(2), & \omega = \omega_4 \end{cases}$$

### 3.1.5 *Self-Financing Trading Strategies*

As mentioned earlier, for $t \geqslant 1$ the quantity $V_t$ represents the time $t$ value of the portfolio just before any transactions (that is, any changes of ownership positions) take place at that time. Meanwhile,

$$H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t), \quad t \geqslant 1$$

represents the time $t$ value of the portfolio just after any time $t$ transactions, that is, just before the portfolio is carried forward to time $t+1$. In general, these two portfolio values can be different, which means that at time $t$ some money is either added to or withdrawn from the portfolio. However, for many applications money cannot be added to or withdrawn from the portfolio at times other than $t = 0$ and $t = T$, and so this leads to the concept of self-financing trading strategies.

A trading strategy $H$ is said to be *self- financing* if

$$\text{(3.1)} \qquad V_t = H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t), \quad t = 1, \ldots, T-1.$$

In other words, the time $t$ values of the portfolio just before and just after any time $t$ transactions are equal. Intuitively, if no money is added to or withdrawn from the portfolio between times $t = 0$ and $t = T$, then any change in the portfolio's value must be due to a gain or loss in the investments. Note this concept is not relevant to single period models. Moreover, one can show by some simple bookkeeping calculations that:

**(3.2)** A trading strategy $H$ is self-financing if and only if

$$V_t = V_0 + G_t, \quad t = 1, 2, \ldots, T.$$

**Example 3.3 (continued)** For the trading strategy $H$ to be self-financing, one must have at time $t = 1$

$$V_1 = (1+r)H_0(1) + 8H_1(1) = (1+r)H_0(2) + 8H_1(2)$$



in states $\omega_1$ and $\omega_2$ and

$$V_1 = (1+r)H_0(1) + 4H_1(1) = (1+r)H_0(2) + 4H_1(2)$$

in states $\omega_3$ and $\omega_4$. Equivalently, using $V_t = V_0 + G_t$ for $t = 1$ and $t = 2$ gives $V_1 = V_2 - [G_2 - G_1]$. Computing this for $\omega_1$ yields

$$\begin{aligned} V_1 &= (1+r)^2 H_0(2) + 9H_1(2) - [r(1+r)H_0(2) + H_1(2)] \\ &= (1+r)H_0(2) + 8H_1(2) \end{aligned}$$

which is the same as the self-financing equation (3.1). Similarly for $\omega_2$. For $\omega_3$ one computes

$$\begin{aligned} V_1 &= (1+r)^2 H_0(2) + 6H_1(2) - [r(1+r)H_0(2) + 2H_1(2)] \\ &= (1+r)H_0(2) + 4H_1(2) \end{aligned}$$

which is also the same as earlier.

### 3.1.6 *Discounted Prices*

It is convenient to introduce discounted versions of some of the price processes that have been introduced above. For much of the financial theory that will be developed, what matters is the behavior of the security prices relative to each other, rather than their absolute behavior. Hence we will be interested in normalized versions of the security prices, obtained by dividing the prices of the various securities by the price of one of them. For this purpose it is convenient to select the bank account as the divisor, that is, as the *numeraire*.

The *discounted price process* $S_n^* = \{S_n^*(t); t = 0, 1, \ldots, T\}$ is defined by

$$S_n^*(t) \equiv S_n(t)/B_t, \quad t = 0, 1, \ldots, T; \; n = 1, 2, \ldots, N$$

The *discounted value process* $V^* = \{V_t^*; t = 0, 1, \ldots, T\}$ is defined by

$$V_t^* \equiv \begin{cases} H_0(1) + \sum_{n=1}^{N} H_n(1)S_n^*(0), & t = 0 \\ H_0(t) + \sum_{n=1}^{N} H_n(t)S_n^*(t), & t = 1, \ldots, T \end{cases}$$

Finally, the *discounted gains process* $G^* = \{G_t^*; t = 1, 2, \ldots, T\}$ is defined by

$$G_t^* \equiv \sum_{n=1}^{N} \sum_{u=1}^{t} H_n(u)\, \Delta S_n^*(u), \quad t = 1, \ldots, T$$

where the notation $\Delta S_n^*(u)$ means $S_n^*(u) - S_n^*(u-1)$, as should be guessed from the earlier definition of the (undiscounted) gains process. All of these are adapted, stochastic processes.

By carrying out some bookkeeping calculations it is straightforward to verify that

(3.3) $$V_t^* = V_t/B_t, \ t = 0, 1, \ldots, T$$

and that

**(3.4)** A trading strategy $H$ is self-financing if and only if

$$V_t^* = V_0^* + G_t^*, \text{ for t} = 1, 2, \ldots, \text{T}$$

Exercise 3.1 Verify (3.2).

Exercise 3.2 Verify (3.3).

Exercise 3.3 Verify (3.4).

## 3.2 Return and Dividend Processes

Given a price process $S_n, n = 1, \ldots, N$, suppose one defines a new process $R_n = \{R_n(t);\ t = 0, 1, \ldots, T\}$ by setting $R_n(0) = 0$ and, for all $t = 1, \ldots, T$,

(3.5) $$\Delta R_n(t) \equiv \begin{cases} \Delta S_n(t)/S_n(t-1), & S_n(t-1) > 0 \\ 0, & S_n(t-1) = 0 \end{cases}$$

This process $R_n$ is called the *return process* corresponding to the price process $S_n$. The return process $R_0$ is defined in terms of the bank account process $B$ in a similar manner, giving $\Delta R_0(t) = r_t$. These and other kinds of return processes are often useful for making various kinds of calculations.

Note that $\Delta R_n(t) \geqslant -1$, because the prices are non-negative. Moreover, $\Delta R_n(t) > -1$ for all $t$ if and only if the price process $S_n$ is strictly positive.

The equation defining $R_n$ is the same as

(3.6) $$\Delta S_n(t) = S_n(t-1)\Delta R_n(t), \ t = 1, \ldots, T$$

which, in turn, is the same as

(3.7) $$S_n(t) = S_n(0) + \sum_{u=1}^{t} S_n(u-1)\ \Delta R_n(u), \ t = 1, \ldots, T$$

Still another equivalent equation is

(3.8) $$S_n(t) = S_n(0)\prod_{u=1}^{t}(1 + \Delta R_n(u)), \ t = 1, \ldots, T$$

These last two equations show that starting with a return process $R_n$ satisfying $\Delta R_n > -1$ together with an initial price $S_n(0)$, one can define a strictly positive price process . Hence there is a one-to-one correspondence between positive price processes and pairs consisting of a positive initial price together with a return process having jumps bigger than minus one. This is a useful fact, because it is often easier to set up a securities market model by



first specifying the return processes rather than by directly specifying the price processes.

### 3.2.1 Returns for Discounted Price Processes

The return processes corresponding to value processes, discounted price processes, and so forth can be defined in exactly the same way. Since $S_n^*(t) = S_n(t)/B_t$ for $t = 1, \ldots, T$, one may wonder how $R_n^*$, which denotes the return process corresponding to $S_n^*$, relates to $R_n$, the return process corresponding to the undiscounted price process. To find out, we can compute

$$\begin{aligned}\Delta S_n^*(t) &= S_n^*(t) - S_n^*(t-1) = S_n(t)/B_t - S_n^*(t-1) \\ &= \frac{S_n(t-1)[1 + \Delta R_n(t)]}{B_{t-1}[1 + \Delta R_0(t)]} - S_n^*(t-1) \\ &= S_n^*(t-1)\left[\frac{\Delta R_n(t) - \Delta R_0(t)}{1 + \Delta R_0(t)}\right]\end{aligned}$$

Since $\Delta S_n^*(t) = S_n^*(t-1)\Delta R_n^*(t)$ by definition, this implies

$$\Delta R_n^*(t) = \frac{\Delta R_n(t) - \Delta R_0(t)}{1 + \Delta R_0(t)}$$

This is consistent with

$$\begin{aligned}S_n^*(t) &= S_n^*(0)\prod_{u=1}^{t}(1 + \Delta R_n^*(u)) \\ &= S_n(0)\prod_{u=1}^{t}\left[\frac{1 + \Delta R_n(u)}{1 + \Delta R_0(u)}\right] = S_n(t)/B_t\end{aligned}$$

### 3.2.2 Returns for the Value and Gains Processes

Since $H_n(t)\Delta S_n(t) = H_n(t)S_n(t-1)\Delta R_n(t)$, it follows that the gains process satisfies

$$\begin{aligned}G_t &= \sum_{u=1}^{t} H_0(u)\ B_{u-1}\Delta R_0(u) + \sum_{n=1}^{N}\sum_{u=1}^{t} H_n(u)\ S_n(u-1)\ \Delta R_n(u) \\ &= \sum_{u=1}^{t} M_0(u)\ \Delta R_o(u) + \sum_{n=1}^{N}\sum_{u=1}^{t} M_n(u)\ \Delta R_n(u)\end{aligned}$$

where the quantity

$$M_n(t) \equiv \begin{cases} H_0(t)\ B_{t-1}, & \text{n} = 0 \\ H_n(t)\ S_n(t-1), & \text{n} = 1,2,\ldots,\text{N} \end{cases}$$

can be interpreted as the money invested in security $n$ beginning at time $t-1$. In other words, $M \equiv \{M_0, M_1, \ldots, M_N\}$ is an alternative way to

specify the trading strategy, and the preceding expression for $G$ says that the gains process is equal to the stochastic integral of the trading strategy $M$ with respect to the return process of the securities. Note that $M_n = \{M_n(t); t = 1, 2, \ldots, T\}$ is a predictable stochastic process.

Next, consider the return process, denoted $R$, corresponding to the value process $V$. Since

$$\begin{aligned} V_t &= V_{t-1} + H_0(t)\Delta B_t + \sum_{n=1}^{N} H_n(t)\Delta S_n(t) \\ &= V_{t-1} + M_0(t)\Delta R_0(t) + \sum_{n=1}^{N} M_n(t)\Delta R_n(t) \end{aligned}$$

it follows that

$$\begin{aligned} \Delta R(t) &= [V_t - V_{t-1}]/V_{t-1} \\ &= \frac{M_0(t)}{V_{t-1}}\Delta R_0(t) + \sum_{n=1}^{N}\left[\frac{M_n(t)}{V_{t-1}}\right]\Delta R_n(t) \\ &= \sum_{n=0}^{N} F_n(t)\Delta R_n(t) \end{aligned}$$

where

$$F_n(t) \equiv M_n(t)/V_{t-1}, \; n = 0, 1, \ldots, N$$

represents the fraction of the investor's wealth invested in security $n$ at time $t-1$ and about to be carried forward to time $t$. The equation for $R$ expresses the return process for the value process in terms of the return processes for the individual securities. Note that $F_n = \{F_n(t); t = 1, 2, \ldots, T\}$ is a predictable stochastic process. The quantity $F_n(t)$ can be negative for some $n$, $t$, and $\omega$, but one always has $F_0(t) = 1 - F_1(t) - \ldots - F_N(t)$. Hence $F \equiv \{F_1, \ldots, F_N\}$ is still another form of the trading strategy.

In summary, the trading strategy can be expressed in three ways: as the number of units, $H_n$, invested in security $n$; as the amount of money, $M_n$, invested in security $n$; or as the fraction of wealth, $F_n$, invested in security $n$. In the latter case, if you also know the return process for each security, then you have a convenient, alternative expression for the value process, namely,

$$V_t = V_0 \prod_{u=1}^{t}[1 + \Delta R(u)] = V_0 \prod_{u=1}^{t}[1 + \sum_{n=0}^{N} F_n(u)\Delta R_n(u)]$$

So, starting with a trading strategy in the fractional form $F = \{F_1, \ldots, F_N\}$ together with the individual return processes $\{R_n\}$ and the initial value $V_0$, one can compute $V_t$ as well as the trading strategy in the monetary form $M = \{M_0, M_1, \ldots, M_N\}$. Knowing, in addition, the initial prices $B_0$ and $S_n(0)$, one can compute the price processes as well as, finally, the trading strategy in unit form $H = \{H_0, H_1, \ldots, H_N\}$.



With $R^* = \{R^*(t);\ t = 0, 1, \ldots, T\}$ denoting the return process corresponding to the discounted value process $V^*$, it follows from the above results that

$$\Delta R^*(t) = \frac{\Delta R(t) - \Delta R_0(t)}{1 + \Delta R_0(t)} \tag{3.9}$$

Hence

$$V_t^* = V_0^* \prod_{u=1}^{t}[1 + \Delta R^*(u)] = V_0 \prod_{u=1}^{t}\left[\frac{1 + \Delta R(u)}{1 + \Delta R_0(u)}\right]$$

which is consistent with the fact that $V^* = V/B$.

### 3.2.3 Dividend Processes

Various kinds of securities, such as dividend-paying stocks, issue cash payments to the owners on a periodic basis. Up to this point, this feature has been ignored. It is of no interest for the single period model, because $S_n(1)$ represents the time $t = 1$ value of one unit of the security for the investor who made the purchase at time $t = 0$, and how this is divided up between a cash dividend and the time $t = 1$ value of a stock certificate, say, is of no consequence. For multiperiod models, however, it is often important to explicitly model any dividend payments. For example, the investor holding a stock over several periods may receive a cash dividend, and it is necessary to carefully model whether the investor reinvests the cash in the same stock, deposits the cash in the bank account, or uses the cash in another way.

There are two ways to incorporate dividend payments: implicitly and explicitly. With the implicit approach, $S_n(t)$ should be interpreted as the value of the investment where one unit of the security is purchased at time $t = 0$ and held indefinitely, and any dividends received are reinvested in the same security. For example, if a \$1 dividend is received at time $t = 1$, then the *ex-dividend* price at that time is $S_n(1) - 1$, which means the \$1 dividend was used to purchase $(S_n(1) - 1)^{-1}$ additional units of the security. But as time evolves further the bookkeeping becomes rather messy, as one tries to keep track of the true security price, the true position in the security, and so forth. Nevertheless, this implicit approach is sometimes convenient for addressing issues where the return process is what matters, because two securities having the same return process are (at least for some purposes) equivalent, even though one pays dividends and the other does not. In other words, the implicit approach is really one where you can work exclusively with return processes rather than a price process; in effect, each dividend-paying security is replaced by a security that does not pay any dividends but has exactly the same return process.

To see what the return process is for a dividend-paying security, and to describe the explicit approach, we call $D_n = \{D_n(t) : t = 0, \ldots, T\}$ the *dividend process* for security $n$, $n = 1, \ldots, N$, where $D_n(0) = 0$ and $\Delta D_n(t)$ represents the dividend per security unit paid at time $t$. Thus $D_n(t)$ represents

the cumulative dividend payments associated with one unit of the security. Moreover, $S_n(t)$ represents the *ex-dividend* price of the security, that is, the price after any time $t$ dividend payment. It will always be assumed that the dividend process is an adapted process. For securities paying dividends, the dividend processes should be specified as part of the data.

Now an investor owning one unit of security $n$ at time $t-1$ will earn a profit of $\Delta S_n(t) + \Delta D_n(t)$ over the ensuing period, so the corresponding one-period return is (assuming $S_n(t-1) > 0$)

$$\Delta R_n(t) = \frac{\Delta S_n(t) + \Delta D_n(t)}{S_n(t-1)}, \quad t = 1, \dots, T; \; n = 1, \dots, N$$

Thus knowing the price and dividend processes for a security, one can deduce the security's return process (of course, $R_n(0) = 0$), but not conversely. For a given return process, it is clear that there can exist an infinite number of price-dividend process pairs all having this same return process, with one of these pairs satisfying $D_n = 0$.

The discounted return process $R_n^*$ for a dividend-paying security is defined by taking $R_n^*(0) = 0$ and

$$\Delta R_n^*(t) = \frac{\Delta S_n^*(t) + \Delta D_n(t)/B_t}{S_n^*(t-1)}, \quad t = 1, \dots, T; \; n = 1, \dots, N$$

It is not difficult to verify that the earlier expression $\Delta R_n^*(t) = [\Delta R_n(t) - \Delta R_0(t)]/[1 + \Delta R_0(t)]$, derived for the case of no dividends, still holds.

In all that follows, a dividend-paying security will have its dividends modeled in this explicit fashion. Thus if no dividend process is specified, then it should be assumed the security does not pay any dividends.

---

**Exercise 3.4** Show that in example 3.3 one has $R_1(1,\omega_1) = R_1(1,\omega_2) = 0.6$, $R_1(1,\omega_3) = R_1(1,\omega_4) = -0.2$, $R_1(2,\omega_1) = 0.725$, $R_1(2,\omega_2) = 0.35$, $R_1(2,\omega_3) = 0.3$ and $R_1(2,\omega_4) = -0.45$. What is the return process $R_n^*$ corresponding to $S_1^*$ in the case where the interest rate is the constant $r > 0$?

**Exercise 3.5** Show that (3.5), (3.6), (3.7), and (3.8) are all equivalent when $S_n$ is strictly positive. What if $S_n$ can be zero?

**Exercise 3.6** Verify relationship (3.9) in two different ways.

---

## 3.3 Conditional Expectation and Martingales

Just as with single period models, the multiperiod securities market model will have no arbitrage opportunities if and only if there exists a risk neutral probability measure. However, in the multiperiod situation the risk neutral probability measures are defined in terms of things called martingales, and these, in turn, are defined with conditional expectations. The



purpose of this section will therefore be to introduce these two concepts from the world of probability theory.

In elementary probability theory, where, as we are assuming, the sample space $\Omega$ is finite, the conditional expectation of the discrete random variable $Y$ given the event $A$ is denoted $E[Y|A]$ and defined in terms of the conditional probability distribution $P\{Y = y|A\}$ by

$$E[Y|A] = \sum_y y \, P\{Y = y|A\}$$

Since $P\{Y = y|A\} = P\{Y = y, A\}/P\{A\}$ by Bayes's Law, it follows that

$$E[Y|A] = \sum_y y \, P\{Y(\omega) = y, A\}/P\{A\} = \sum_{\omega \in A} Y(\omega) \, P\{\omega\}/P\{A\}$$

Hence in example 3.3, for instance, where $P\{\omega\} = 1/4$, for all $\omega \in \Omega$, one has $P\{S_2 = 9|S_1 = 8\} = P\{S_2 = 6|S_1 = 8\} = (1/4)/(1/4 + 1/4) = 1/2$, in which case $E[S_2|S_1 = 8] = 7.5$. Similarly, $E[S_2|S_1 = 4] = 4.5$.

When working with stochastic processes defined on a filtered probability space, it is often convenient to use $E[Y|\mathscr{F}]$ as a summary of all the conditional expectations of the form $E[Y|A]$ as the event $A$ runs through the algebra $\mathscr{F}$. The idea is that $E[Y|\mathscr{F}]$ is defined by

$$E[Y|\mathscr{F}] \, 1_A = E[Y|A], \text{ all } A \in \mathscr{P}$$

where $\mathscr{P}$ is the partition of $\Omega$ that corresponds to $\mathscr{F}$. Thus $E[Y|\mathscr{F}]$ will be a random variable that is measurable with respect to $\mathscr{F}$. In the case of example 3.3, for instance,

$$E[S_2|\mathscr{F}_1] = \begin{cases} 7.5, & \omega_1 \text{ and } \omega_2 \\ 4.5, & \omega_3 \text{ and } \omega_4 \end{cases}$$

Since $E[Y|\mathscr{F}]$ is a perfectly good random variable, we can compute its expectation:

$$\begin{aligned} E\left[E[Y|\mathscr{F}]\right] &= E\Big[\sum_{A \in \mathscr{P}} E[Y|A] \, 1_A\Big] = \sum_{A \in \mathscr{P}} P\{A\} \, E[Y|A] \\ &= \sum_{A \in \mathscr{P}} P\{A\} \sum_{\omega \in A} Y(\omega) \, P\{\omega\}/P\{A\} \\ &= \sum_{A \in \mathscr{P}} \sum_{\omega \in A} Y(\omega) \, P\{\omega\} = EY \end{aligned}$$

A slight generalization of this is the following.

(3.10) $\quad$ If $\mathscr{F}_1 \subset \mathscr{F}_2$, then $\mathrm{E}\left[\mathrm{E}[\mathrm{Y}|\mathscr{F}_2] \mid \mathscr{F}_1\right] = \mathrm{E}[\mathrm{Y}|\mathscr{F}_1]$

In example 3.3, for instance, $E[E[S_2|\mathscr{F}_1]] = 7.5/2 + 4.5/2 = 6 = ES_2$.

If the random variable $X \in \mathscr{F}$, then one can write

(3.11) $$X = \sum_{A \in \mathscr{P}} x_A \, 1_A,$$

where $x_A$ is a scalar and $\mathscr{P}$ is the partition corresponding to $\mathscr{F}$. Hence

$$E[XY|\mathscr{F}] = \sum_{A\in\mathscr{P}} E[XY|A]\, 1_A = \sum_{A\in\mathscr{P}} E[x_A Y|A]\, 1_A$$
$$= \sum_{A\in\mathscr{P}} x_A E[Y|A]\, 1_A = X\, E[Y|\mathscr{F}]$$

In a similar fashion one can verify the following generalization.

**(3.12)** Given random variables $X_1, X_2, Y_1$, and $Y_2$ with $X_1, X_2 \in \mathscr{F}$, one has $E[X_1 Y_1 + X_2 Y_2|\mathscr{F}] = X_1 E[Y_1|\mathscr{F}] + X_2 E[Y_2|\mathscr{F}]$.

If $Y$ is a constant, then clearly $E[Y|\mathscr{F}] = Y$. Taking $Y = 1$ and using (3.12), it follows that

**(3.13)** If $X \in \mathscr{F}$ then $E[X|\mathscr{F}] = X$

In the case of example 3.3, for instance, $E[S_1 S_2|\mathscr{F}_1] = S_1 E[S_2|\mathscr{F}_1]$ and $E[S_1|\mathscr{F}_1] = S_1$.

Taking $A \in \mathscr{F}$ implies $1_A \in \mathscr{F}$, so $E[Y1_A|\mathscr{F}] = 1_A E[Y|\mathscr{F}]$ by (3.12). Hence by (3.10) one has

$$E[1_A E[Y|\mathscr{F}]] = E[Y1_A], \text{ all } A \in \mathscr{F}$$

It turns out this equation provides an alternative definition of $E[Y|\mathscr{F}]$, one that generalizes to probability spaces where $\Omega$ is not finite. In particular, suppose $X \in \mathscr{F}$ satisfies

**(3.14)** $E[1_A X] = E[Y1_A]$, all $A \in \mathscr{F}$

Taking $X$ as in (3.11), it follows that $E[1_A X] = x_A P\{A\}$ when $A \in \mathscr{P}$, the partition corresponding to $\mathscr{F}$. Meanwhile, taking the same $A \in \mathscr{P}$ one has

$$E[Y1_A] = \sum_{\omega\in A} Y(\omega)P\{\omega\} = P\{A\}\sum_{\omega\in A} Y(\omega)P\{\omega\}/P\{A\} = P\{A\}E[Y|A]$$

Hence (3.14) implies

$$x_A = E[Y|A], \text{ all } A \in \mathscr{P}$$

which means $X = E[Y|\mathscr{F}]$. This characterization of $E[Y|\mathscr{F}]$ is summarized in the following.

**(3.15)** Given an arbitrary random variable $Y$, the conditional expectation $E[Y|\mathscr{F}]$ is the unique random variable such that

(*a*) $E[Y|\mathscr{F}] \in \mathscr{F}$

(*b*) $E[E[Y|\mathscr{F}]1_A] = E[Y1_A]$, all $A \in \mathscr{F}$

We now turn to the topic of martingales. We are given a filtered probability space together with an adapted stochastic process $Z = \{Z_t;\ t = 0, 1, \ldots, T\}$. The process $Z$ is said to be a *martingale* if

$$E[Z_{t+s}|\mathscr{F}_t] = Z_t, \text{ all s,t} \geqslant 0$$



**Example 3.4** Consider a coin with P(heads) $\equiv p$, where $0 < p < 1$. Let

$N_t \equiv$ number of heads after $t$ independent coin flips,
$Z_t \equiv N_t - pt$, and
$\mathscr{F}_t \equiv$ algebra corresponding to the observations of the first $t$ coin flips.

It is easy to see that $E[N_t] = pt$. Moreover, $Z$ is a martingale, because

$$\begin{aligned} E[Z_{t+s}|\mathscr{F}_t] &= E[N_{t+s} - p(t+s)|\mathscr{F}_t] \\ &= E[N_{t+s} - N_t + N_t|\mathscr{F}_t] - p(t+s) \\ &= E[N_{t+s} - N_t|\mathscr{F}_t] + E[N_t|\mathscr{F}_t] - pt - ps \\ &= E[N_s] + N_t - pt - ps \\ &= N_t - pt = Z_t \end{aligned}$$

Here we used the self-evident fact that coin flips $t+1, t+2, \ldots, t+s$ are independent of the first $t$ flips, in which case the expected number of heads observed during flips $t+1, t+2, \ldots, t+s$, conditioned on the observations of the first $t$ flips, is equal to the expected number of heads observed during $s$ flips.

Martingales are often used as models of fair gambling games, where $Z_t$ represents the gambler's stake after $t$ plays of the game.

I conclude by mentioning two kinds of processes that are closely related to martingales. An adapted stochastic process $Z = \{Z_t; t = 0, 1, \ldots, T\}$ is said to be a *supermartingale* if

$$E[Z_{t+s}|\mathscr{F}_t] \leqslant Z_t \text{ , all s,t} \geqslant 0$$

Thus a supermartingale resembles a martingale, except that the conditional expectation of the future value can be less than as well as equal to the current value. All martingales are supermartingales, but not vice versa.

Finally, an adapted stochastic process $Z = \{Z_t; t = 0, 1, \ldots, T\}$ is said to be a *submartingale* if

$$E[Z_{t+s}|\mathscr{F}_t] \geqslant Z_t \text{ , all s,t} \geqslant 0$$

Thus $Z$ is a submartingale if and only if $-Z$ is a supermartingale. Also, $Z$ is a martingale if and only if it is both a submartingale and a supermartingale.

**Exercise 3.7** Verify (3.10).

**Exercise 3.8** Verify (3.12).

**Exercise 3.9** With $X = \{X_t; t = 0, 1, \ldots, T\}$ an adapted stochastic process, show that the following are equivalent:

(a) $X$ is a martingale.
(b) $X_t = E[X_T|\mathscr{F}_t]$, $t = 0, 1, \ldots, T-1$.
(c) $E[\Delta X_{t+1}|\mathscr{F}_t] = 0$, $t = 0, 1, \ldots, T-1$.

## 3.4 Economic Considerations

I now return to our securities market model and develop a relationship that is analogous to the one for single period models: there are no arbitrage opportunities if and only if there exists a risk neutral probability measure. This and most other economic concepts developed for single period models hold in a rather similar fashion for multiperiod models; really only the details are different.

An *arbitrage opportunity* in the case of a multiperiod securities market is some trading strategy $H$ such that

(a) $V_0 = 0$,
(b) $V_T \geqslant 0$,
(c) $EV_T > 0$, and
(d) $H$ is self-financing.

As with single period models, the existence of an arbitrage opportunity is not consistent with economic equilibrium. The presence of a possibility of turning zero dollars into a positive amount of dollars without any risk of losing money would beckon market forces that would disrupt the underlying structure of security prices.

In view of (3.3), it is immediate that

**(3.16)** The self-financing trading strategy $H$ is an arbitrage opportunity if and only if

(a) $V_0^* = 0$,
(b) $V_T^* \geqslant 0$, and
(c) $EV_T^* > 0$.

And thanks to (3.4), we also have that

**(3.17)** The self-financing trading stragegy $H$ is an arbitrage opportunity if and only if

(a) $G_T^* \geqslant 0$, and
(b) $EG_T^* > 0$.

**Example 3.3 (continued)** If $B_t = 1$ for $t = 0$, 1, and 2, then there are no arbitrage opportunities. If the investor has any position at all in the risky asset at time $t = 1$, then there is always the possibility that the price will move in a losing direction. If any position is taken in the risky asset at time $t = 0$, then there is the possibility of 'being in the red' at time $t = 1$ with no guaranteed way of recovering in the next period.

On the other hand, suppose $B_t = (1 + r)^t$ with the scalar $r \geqslant 12.5$ per cent. Consider the trading strategy where you start with zero dollars and do nothing at time $t = 0$ or at time $t = 1$ if $S_1 = 4$, but if $S_1 = 8$ then at time $t = 1$ you sell short one share of the risky asset (i.e., $H_1(2) = -1$) and invest the \$8 proceeds in the bank account (i.e., $H_0(2) = 8/(1 + r)$). Then at time $t = 2$ the value of the portfolio is



$$V_2 = \begin{cases} (1+r)^2 H_0(2) + 9H_1(2) = 8(1+r) - 9 \geqslant 0, & \omega = \omega_1 \\ (1+r)^2 H_0(2) + 6H_1(2) = 8(1+r) - 6 \geqslant 0, & \omega = \omega_2 \end{cases}$$

Hence this trading strategy is an arbitrage opportunity.

It turns out that, as with single period models, there are no arbitrage opportunities if and only if there exists a risk neutral probability measure. But while risk neutral probabilities are defined in terms of ordinary expectations for single period models, they are defined in terms of martingales for multiperiod models.

A *risk neutral probability measure* (also called a *martingale measure*) is a probability measure $Q$ such that

1 $Q(\omega) > 0$ for all $\omega \in \Omega$, and
2 The discounted price process $S_n^*$ is a martingale under $Q$ for every $n = 1, 2, \ldots, N$.

In other words, in view of the definition of martingales, a risk neutral probability measure $Q$ must satisfy

$$E_Q[S_n^*(t+s)|\mathscr{F}_t] = S_n^*(t), \quad t, s \geqslant 0$$

that is,

$$E_Q[B_t S_n(t+s)/B_{t+s}|\mathscr{F}_t] = S_n(t), \quad t, s \geqslant 0 \tag{3.18}$$

**Example 3.3 (continued)** Suppose $B_t = (1 + r)^t$ where $r \geqslant 0$ is a constant. We want to compute a martingale measure, if there is one. To do this, we can use (3.18) for different values of $s$ and $t$, giving the following equations:

$$\begin{aligned} t = 0,\ s = 1: &\quad 5(1+r) = 8[Q(\omega_1) + Q(\omega_2)] + 4[Q(\omega_3) + Q(\omega_4)] \\ t = 0,\ s = 2: &\quad 5(1+r)^2 = 9Q(\omega_1) + 6Q(\omega_2) + 6Q(\omega_3) + 3Q(\omega_4) \\ t = 1,\ s = 1: &\quad 8(1+r) = [9Q(\omega_1) + 6Q(\omega_2)]/[Q(\omega_1) + Q(\omega_2)] \\ t = 1,\ s = 1: &\quad 4(1+r) = [6Q(\omega_3) + 3Q(\omega_4)]/[Q(\omega_3) + Q(\omega_4)] \end{aligned}$$

Taking any three of these equations together with the equation $Q(\omega_1) + \ldots + Q(\omega_4) = 1$ allows one to solve for the four unknowns:

$$Q(\omega_1) = \left(\frac{1+5r}{4}\right)\left(\frac{2+8r}{3}\right) \quad Q(\omega_2) = \left(\frac{1+5r}{4}\right)\left(\frac{1-8r}{3}\right)$$
$$Q(\omega_3) = \left(\frac{3-5r}{4}\right)\left(\frac{1+4r}{3}\right) \quad Q(\omega_4) = \left(\frac{3-5r}{4}\right)\left(\frac{2-4r}{3}\right)$$

Note these are all strictly positive, and thus we have a valid probability measure, if $0 \leqslant r < 1/8$. On the other hand, if $r \geqslant 1/8$ then $Q(\omega_2)$ is not strictly positive, in which case there is no martingale measure.

If $r < 1/8$, then you cannot find any situation (that is, any time and state) where the discounted price next period can be strictly higher

than the current discounted price unless there is a chance the discounted price next period is strictly lower. Nor can you find any situation where the discounted price next period can be strictly lower than the current discounted price unless there is a chance the discounted price next period is strictly higher. Hence in every situation there is a risk that a non-zero position in the risky security will lose money over the next period, and so there are no arbitrage opportunities.

On the other hand, if $r \geqslant 1/8$, then if $S_1 = 8$ at time $t = 1$ (which means the state is $\omega$ or $\omega_2$), then $S_2^*(\omega_1) \leqslant S_1^*(\omega_1)$ and $S_2^*(\omega_2) < S_1^*(\omega_2)$. Here is a situation where the discounted price next period can be strictly lower than the present price without there being any risk that the discounted price next period can be strictly higher. Of course, now there is an arbitrage opportunity, as was described earlier.

We now come to the principal result of this section.

**(3.19)** There are no arbitrage opportunities if and only if there exists a martingale measure $Q$.

A proof of this is similar to our proof of the analogous result for single period models. In order to explain one direction, namely why the existence of a martingale measure implies there are no arbitrage opportunities, we will first provide a useful result that is in general terms:

**(3.20)** If $Z$ is a martingale and $H$ is a predictable process, then

$$G_t \equiv \sum_{u=1}^{t} H_u \, \Delta Z_u$$

is also a martingale.

This follows from some straightforward calculations using some properties of conditional expectations. Let $s, t \geqslant 0$ be arbitrary. Then

$$\begin{aligned}
E[G_{t+s}|\mathscr{F}_t] &= E[G_{t+s} - G_t + G_t|\mathscr{F}_t] \\
&= E[H_{t+1}\Delta Z_{t+1} + \ldots + H_{t+s}\Delta Z_{t+s}|\mathscr{F}_t] + G_t \\
&= E[E[H_{t+1}\Delta Z_{t+1}|\mathscr{F}_t] \mid \mathscr{F}_t] \\
&\quad + E[E[H_{t+2}\Delta Z_{t+2}|\mathscr{F}_{t+1}] \mid \mathscr{F}_t] + \ldots \\
&\quad + E[E[H_{t+s}\Delta Z_{t+s}|\mathscr{F}_{t+s-1}] \mid \mathscr{F}_t] + G_t \\
&= E[H_{t+1}E[\Delta Z_{t+1}|\mathscr{F}_t] \mid \mathscr{F}_t] \\
&\quad + E[H_{t+2}E[\Delta Z_{t+2}|\mathscr{F}_{t+1}] \mid \mathscr{F}_t] + \ldots \\
&\quad + E[H_{t+s}E[\Delta Z_{t+s}|\mathscr{F}_{t+s-1} \mid \mathscr{F}_t] + G_t \\
&= E[H_{t+1} \cdot 0|\mathscr{F}_t] + \ldots + E[H_{t+s} \cdot 0|\mathscr{F}_t] + G_t \\
&= G_t
\end{aligned}$$



where the next to last equality follows from the fact that $Z$ is a martingale. Hence $G$ is a martingale too.

It follows directly from (3.4) and (3.20) that we have the following result, which is of considerable practical importance:

**(3.21)** If $Q$ is a martingale measure and $H$ is a self-financing trading strategy, then $V^*$, the discounted value process corresponding to $H$, is a martingale under $Q$.

We can use (3.21) to quickly show that the existence of a martingale measure $Q$ implies there cannot be any arbitrage opportunities. Suppose $H$ is an arbitrary self-financing trading strategy with $V_T^* \geqslant 0$ and $EV_T^* > 0$. This implies $E_Q V_T^* > 0$. Since by (3.21) $V^*$ is a martingale under $Q$, it follows that $V_0^* = E_Q V_T^* > 0$. Hence by (3.16) $H$ cannot be an arbitrage opportunity, nor can any other trading strategies be arbitrage opportunities, by the arbitrary choice of $H$.

One can show the converse of (3.19) in several ways, such as by using an extension of the separating hyperplane theorem argument that was used for the case of single period models. However, it is much easier to build on what we already know for single period models, namely, the single-period result analogous to (3.19). In particular, knowing there are no arbitrage opportunities for the multiperiod model, one can construct one-period conditional probabilities that are compatible with risk neutrality. The martingale measure $Q$ can then be computed from these conditional probabilities by multiplying them together in accordance with the information structure of the multiperiod model. In other words, the martingale measure for the multiperiod model is constructed by 'pasting together' various single period models.

To be more precise, there is one underlying single period model corresponding to each non-terminal node of the tree structure of the information submodel, that is, to each $A \in \mathscr{P}_t$ (the minimal partition corresponding to $\mathscr{F}_t$) for each $t < T$. The single period 'time 0' discounted price of risky security $n$ is $S_n^*(t, \omega), \omega \in A$, which is constant on $A$. The corresponding single period 'sample space' consists of one state for each cell $A' \subseteq A$ that is a member of the partition $\mathscr{P}_{t+1}$ (that is, one state for each branch coming out of the node in the tree structure of the information submodel). Finally, the 'time 1' discounted prices for this single period model are given by the values of $S_n^*(t+1, \omega)$ for $\omega \in A$.

If any underlying single period model has an arbitrage opportunity in the single period sense, then the multiperiod model must have an arbitrage opportunity in the multiperiod sense. To see this, suppose there exists an arbitrage opportunity $\hat{H}$ for the single period model corresponding to some $A \in \mathscr{P}_t$ for $t < T$. This means that the discounted gain $\hat{H}_1 \Delta S_1^*(t+1) + \ldots + \hat{H}_N \Delta S_N^*(t+1)$ is non-negative and not identical to zero on the event $A$. Now construct a multiperiod trading strategy $H$ by taking

$$H_n(s,\omega)=\begin{cases}\hat{H}_n, & s=t+1, \omega\in A, n=1,\dots,N\\ -\hat{H}_1S_1^*(t)-\ldots-\hat{H}_NS_N^*(t), & s=t+1, \omega\in A, n=0\\ \hat{H}_1\Delta S_1^*(t+1)+\ldots+\hat{H}_N\Delta S_N^*(t+1), & s>t+1, \omega\in A, n=0\\ 0, & \text{otherwise}\end{cases}$$

Thus, as can be verified with a little work, $H$ is the self-financing trading strategy which starts with zero money and does nothing unless the event $A$ occurs at time $t$, in which case at time $t$ the position $\hat{H}_n$ is taken in the $n$th risky security, while the position in the bank account is chosen in a self-financing manner. Subsequently, no position is taken in any of the risky securities; any non-zero value of the portfolio is reflected by a position in the bank account. If $\hat{H}$ is an arbitrage opportunity for the single period model, then this subsequent position in the bank account will, in fact, be non-negative for all $\omega\in\Omega$ and strictly positive for at least one $\omega\in A$. In particular, under $H$ one will have $V_0^*=0$, $V_T^*\geqslant 0$, and $V_T^*(\omega)>0$ for at least one $\omega\in\Omega$, that is, $H$ will be an arbitrage opportunity.

In other words, we see that

(3.22) If the multiperiod model does not have any arbitrage opportunities, then none of the underlying single period models has any arbitrage opportunities in the single period sense.

Consequently, in view of what we know for single period models, corresponding to each underlying single period model is a risk neutral probability measure. For example, corresponding to each $A\in\mathscr{P}_t$ for $t<T$ is a probability measure, denoted $Q(t, A)$, on the single period sample space. This probability measure gives positive mass to each cell $A'\subseteq A$ in the partition $\mathscr{P}_{t+1}$, it sums to one over such cells, and it satisfies $E_{Q(t,A)}\Delta S_n^*(t+1)=0$ for $n=1,\ldots,N$.

Notice that $Q(t, A)$ gives rise to a probability for each branch in the information tree that emerges from the node corresponding to $(t, A)$. These probabilities should be thought of as conditional risk neutral probabilities, given the event $A$ at time $t$. Hence starting with a collection of risk neutral probability measures $Q(t, A)$ for all $A\in\mathscr{F}_t$ and $t<T$, one can construct a probability measure $Q$ for the whole multiperiod model by proceeding in an obvious manner: $Q(\omega)$ is set equal to the product of the conditional probabilities along the path from the node at $t=0$ to the node corresponding to $(T,\omega)$. Clearly $\sum_{\omega\in\Omega}Q(\omega)=1$. Moreover, $Q(\omega)>0$ for every $\omega\in\Omega$, because all the conditional risk neutral probabilities are strictly positive.

It remains to explain why the probability measure $Q$ that has been constructed is actually a martingale measure. Since $E_{Q(t,A)}\Delta S_n^*(t+1)=0$ for $n=1,\ldots,N, A\in\mathscr{P}_t$ and $t<T$, it follows that

(3.23) $$E_Q[\Delta S_n^*(t+1)|\mathscr{F}_t]=0 \text{ for n} = 1,\ldots,\text{N and t} < \text{T}$$



Now take arbitrary $s, t, \geqslant 0$ and $n$:

$$\begin{aligned}E_Q[S_n^*(t+s)|\mathscr{F}_t] &= E_Q[\Delta S_n^*(t+s)+\ldots+\Delta S_n^*(t+1)+S_n^*(t)|\mathscr{F}_t]\\ &= E_Q[E_Q[\Delta S_n^*(t+s)|\mathscr{F}_{t+s-1}]\,|\mathscr{F}_t]+\ldots\\ &\quad+E_Q[E_Q[\Delta S_n^*(t+1)|\mathscr{F}_t]\,|\mathscr{F}_t\,]+S_n^*(t)\\ &= E_Q[0|\mathscr{F}_t]+\ldots+E_Q[0|\mathscr{F}_t]+S_n^*(t)\\ &= S_n^*(t)\end{aligned}$$

where the next to last equality follows from (3.23) (note this calculation demonstrates that, in general, an expression like (3.23) is equivalent to that used in the definition of a martingale). Hence $S_n^*$ is a martingale under $Q$, and so $Q$ is a risk neutral probability measure.

The preceding explanation of the fundamental principle (3.19) may be a bit abstract, but it becomes transparent if you look at the picture of an information tree for a simple multiperiod model.

**Example 3.3 (continued)** Suppose $B_t=(1+r)^t$ with $r$ a constant, as before. Looking at the node corresponding to $t=0$ (which has two branches emerging, corresponding to $\{\omega_1,\omega_2\}$ and $\{\omega_3,\omega_4\}$, respectively), one sees that the conditional probability measure $Q(0,\Omega)$ can be obtained by solving $5=p8/(1+r)+(1-p)4/(1+r)$. Thus $p$, the conditional probability associated with the $\{\omega_1,\omega_2\}$ branch, is $(1+5r)/4$, in which case the conditional probability associated with the other branch is $(3-5r)/4$. In a similar fashion, one analyzes the $(1,\{\omega_1,\omega_2\})$ node and finally the $(1,\{\omega_3,\omega_4\})$ node to obtain $(2+8r)/3$, $(1-8r)/3$, $(1+4r)/3$, and $(2-4r)/3$ for the conditional probabilities associated with the branches leading into the $(2,\omega_1)$, $(2,\omega_2)$, $(2,\omega_3)$, and $(2,\omega_4)$ nodes, respectively. Notice that all of these conditional probabilities are strictly positive when $0\leqslant r<1/8$, in accordance with our earlier observation that arbitrage opportunities will exist when $r\geqslant 1/8$. Moreover, notice these conditional probabilities are unique, that is, no other choices will yield risk neutral probabilities for the underlying single period models. Hence multiplying the conditional probabilities along the four paths leading to the four states in $\Omega$, we quickly obtain the same martingale measure $Q$ that was derived earlier in a different way. Indeed, we now recognize the earlier expressions for $Q(\omega)$ as being simply the products of the appropriate conditional probabilities.

The martingale measures can be defined in terms of return processes instead of price processes. We saw above that a strictly positive probability measure $Q$ is a martingale measure if and only if (3.23) is satisfied. Since $\Delta S_n^*(t+1)=S_n^*(t)\Delta R_n^*(t+1)$, (3.23) is true if and only if $S_n^*(t)E_Q[\Delta R_n^*(t+1)\,|\mathscr{F}_t]=0$ for all $n$ and $t<T$. But $S_n^*(t)=0$ implies $\Delta R_n^*(t+1)=0$ by the definition of the return process, so this last statement is true if and only if:

(3.24) $$E_Q[\Delta R_n^*(t+1)|\mathscr{F}_t]=0, \text{all } n \text{ and t} < T$$

that is, if and only if $R_n^*$ is a martingale under $Q$ for all $n$. In summary, we have

**(3.25)** The strictly positive probability measure $Q$ is a martingale measure if and only if $R_n^*$ is a martingale under $Q$ for $n = 1, \ldots, N$.

The corresponding requirement in terms of the (undiscounted) return process $R_n$ is not so nice. It is not difficult to verify that

$$\Delta S_n^*(t+1) = S_n^*(t)\left(\frac{\Delta R_n(t+1) - \Delta R_0(t+1)}{1 + \Delta R_0(t+1)}\right)$$

so if all the price processes are strictly positive, then (3.23) is the same as

$$\textbf{(3.26)} \qquad E_Q\left[\frac{\Delta R_n(t+1) - \Delta R_0(t+1)}{1 + \Delta R_0(t+1)} \,|\mathscr{F}_t\right] = 0, \text{ all } n \text{ and } t < T$$

Of course, we could have known immediately that this is equivalent to (3.24), in view of the expression in section 3.2 for $\Delta R_n^*$ in terms of $\Delta R_n$.

Now suppose some of the securities pay a dividend. To check whether there are any arbitrage opportunities, what really matters are the return processes of the securities, so principle (3.25) remains true. In other words, there are no arbitrage opportunities if and only if (3.24) (or (3.26)) holds, where now the return processes are defined in terms of dividend processes as in section 3.2: $\Delta R_n(t+1) = [\Delta S_n(t+1) + \Delta D_n(t+1)]/S_n(t)$ and $\Delta R_n^*(t+1) = [\Delta S_n^*(t+1) + \Delta D_n(t+1)/B_{t+1}]/S_n^*(t)$. Thus (3.24) can be rewritten as

$$\textbf{(3.27)} \qquad E_Q[S_n^*(t+1) + \Delta D_n(t+1)/B_{t+1}|\mathscr{F}_t] = S_n^*(t), \text{ all n and t} < \text{T}$$

This makes sense: if the investor purchases one unit of security $n$ at time $t$, then the expected discounted value of this investment next period is equal to the discounted value of the time $t$ position.

Using (3.27) and a fundamental property of conditional expectations, it is easy to see that

$$\begin{aligned} E_Q[S_n^*(t+2) &+ \Delta D_n(t+2)/B_{t+2} + \Delta D_n(t+1)/B_{t+1}|\mathscr{F}_t] \\ &= E_Q[E_Q[S_n^*(t+2) + \Delta D_n(t+2)/B_{t+2}|\mathscr{F}_{t+1}] \\ &\quad + \Delta D_n(t+1)/B_{t+1}|\mathscr{F}_t] \\ &= E_Q[S_n^*(t+1) + \Delta D_n(t+1)/B_{t+1}|\mathscr{F}_t] \\ &= S_n^*(t) \end{aligned}$$

By a generalization of this argument, we therefore have:

**(3.28)** If $Q$ is a risk neutral probability measure, then for each risky security and every $t$, $s \geqslant 0$

$$S_n^*(t) = E_Q[\Delta D_n(t+1)/B_{t+1} + \ldots + \Delta D_n(t+s)/B_{t+s} + S_n^*(t+s)\,|\mathscr{F}_t]$$

Thus the time $t$ discounted price equals the conditional expected value of the discounted dividend payments up through time $t+s$ plus the time $t+s$ discounted price.



Turning to another topic which is analogous to a notion introduced for single period models, a *linear pricing measure* is a non-negative vector $\pi = (\pi_1, \ldots, \pi_k)$ such that for every self-financing trading strategy $H$ you have

$$V_0 = \sum_\omega \pi(\omega)\, V_T^*(\omega)$$

If $Q$ is a risk neutral probability measure, then clearly it is also a linear pricing measure. Conversely, any strictly positive linear pricing measure $\pi$ must be a risk neutral probability measure. To see this, first take any trading strategy with $H_1 = \ldots = H_N = 0$ to conclude $\pi_1 + \ldots + \pi_k = 1$. Next, fix arbitrary $n, t < T$, and some event $A \in \mathscr{F}_t = 1$, and consider the self-financing trading strategy which starts at time $t = 0$ with \$1 in the bank account and does no transactions unless event $A$ occurs at time $t$, in which case all the money is transferred into a long position in security $n$ for one period, after which it is immediately transferred back into the bank account, where it remains until time $T$. The bank account equals $B_t$ at time $t$, so with security $n$ having value $S_n(t)1_A$ (here $1_A$ denotes the indicator function of the event $A$, that is, $1_A(\omega) = 1$ if $\omega \in A$, whereas $1_A(\omega) = 0$ if $\omega \notin A$), this strategy entails a time $t$ purchase of $B_t/S_n(t) = 1/S_n^*(t)$ units of security $n$ if event $A$ occurs. Since all the money is transferred back to the bank account at time $t+1$, the discounted gain under this trading strategy will be

$$G_T^* = (1_A/S_n^*(t))\; \Delta S_n^*(t+1) = 1_A\, \Delta R_n^*(t+1)$$

Now the trading strategy is self-financing, so (3.4) holds. Thus if $\pi$ is a linear pricing measure, it follows that

$$\sum_\omega \pi(\omega) G_T^*(\omega) = 0$$

This is true for every $A \in \mathscr{F}_t$, so taking $Q(\omega) = \pi(\omega)$ we conclude by (3.24), (3.25), and our expression for $G_T^*$ that $Q$ is a risk neutral probability measure.

In summary, a vector $\pi$ is a linear pricing measure if and only if it is a probability measure on $\Omega$ under which all the discounted price processes are martingales. This is the multiperiod generalization of principle (1.8).

As with single period models, the *law of one price* holds for a multiperiod model if there do not exist two trading strategies, say $\hat{H}$ and $\tilde{H}$, such that $\hat{V}_T(\omega) = \tilde{V}_T(\omega)$ for all $\omega \in \Omega$ but $\hat{V}_0 \neq \tilde{V}_0$. Clearly the existence of a linear pricing measure implies that the law of one price will hold.

Denote

$$\begin{aligned} \mathbb{W} &= \{X \in \mathbb{R}^K : X = G^* \text{ for some trading strategy H}\} \\ \mathbb{W}^\perp &= \{Y \in \mathbb{R}^K : X \cdot Y = 0 \text{ for all } \mathrm{X} \in \mathbb{W}\} \\ \mathbb{A} &= \{X \in \mathbb{R}^K : X \geqslant 0,\ X \neq 0\} \\ \mathbb{P} &= \{X \in \mathbb{R}^K : X_1 + \ldots + X_K = 1,\ X \geqslant 0\} \text{ and} \\ \mathbb{P}^+ &= \{X \in \mathbb{P}\ : X_1 > 0, \ldots, X_K > 0\} \end{aligned}$$

Then just as with single period models, $\mathbb{P} \cap \mathbb{W}^{\perp}$ is the set of all the linear pricing measures and $\mathbb{M} \equiv \mathbb{P}^{+} \cap \mathbb{W}^{\perp}$ is the set of all risk neutral probability measures. Moreover, fundamental principle (3.19) is the same thing as saying $\mathbb{W} \cap \mathbb{A} = \emptyset$ if and only if $\mathbb{M} \neq \emptyset$. Notice that $\mathbb{M}$ is a convex set whose closure equals $\mathbb{P} \cap \mathbb{W}^{\perp}$.

**Exercise 3.10** Consider a 2-period problem with $\Omega = \{\omega_1, \ldots, \omega_5\}$, $r = 0$, and one risky security:

| $\omega$ | $S_0(\omega)$ | $S_1(\omega)$ | $S_2(\omega)$ |
|---|---|---|---|
| $\omega_1$ | 6 | 5 | 3 |
| $\omega_2$ | 6 | 5 | 4 |
| $\omega_3$ | 6 | 5 | 8 |
| $\omega_4$ | 6 | 7 | 6 |
| $\omega_5$ | 6 | 7 | 8 |

The filtration is the one generated by this risky security. Show that the set of all the martingale measures is

$$\mathbb{M} = \{Q \in \mathbb{R}^5 : Q_1 = q/2, Q_2 = (3 - 5q)/8,\ Q_3 = (1 + q)/8,$$
$$Q_4 = Q_5 = 1/4, 0 < q < 3/5\}$$

Show that $\mathbb{P} \cap \mathbb{W}^{\perp}$, the set of all the linear pricing measures, is equal to $\mathbb{M} \cup \{(0, 3/8, 1/8, 1/4, 1/4)\} \cup \{3/10, 0, 1/5, 1/4, 1/4)\}$.

**Exercise 3.11** Let $Q$ be a probability measure, such as a risk neutral probability measure, that is equivalent to $P$, set $X_T(\omega) = Q(\omega)/P(\omega)$, and let $X_t = E[X_T|\mathscr{F}_t]$ for $t = 0, 1, \ldots, T-1$. Show that $X$ is strictly positive with $X_0 = 1$. Let $\{Y_t;\ t = 0, 1, \ldots, T\}$ be a stochastic process. Show that $Y$ is a martingale under $Q$ if and only if the process $\{X_t Y_t;\ t = 0, 1, \ldots, T\}$ is a martingale under $P$. (Hint: use the abstract definition of conditional expectations).

**Exercise 3.12** Use exercise 3.11 to show that if there exists a martingale measure, then there must exist a strictly positive, adapted, real-valued process $Z = \{Z_t;\ t = 0, 1, \ldots, T\}$ satisfying $Z_0 = 1$ and such that $B_t Z_t$, $S_1(t) Z_t, \ldots$, and $S_N(t) Z_t$ are all martingales under $P$. Conversely, show that if there exists a process $Z$ as indicated, then there must exist a martingale measure $Q$. Moreover, show how to compute $Q$ from a specified $Z$. (Note: such a process $Z$ is called a *state-price deflator*.)

## 3.5 The Binomial Model

The 'binomial model' is a simple yet very important model for the price of a single risky security. It is commonly used by practitioners, for example, to determine the price of various kinds of stock options.



Each period there are two possibilities: the security price either goes up by the factor $u$ $(u > 1)$ or it goes down by the factor $d$ $(0 < d < 1)$. The probability of an up move during a period is equal to the parameter $p$, and the moves over time are independent of each other. Hence the binomial model is related to the process $N_t$ (introduced in example 3.4) representing the number of heads after $t$ independent coin flips. The process $N_t$, in turn, is based upon what is called a *Bernoulli process*.

The stochastic process $\{X_t; t = 1, 2, \ldots\}$ is said to be a *Bernoulli process* with parameter $p$ if the random variables $X_1, X_2, \ldots$ are independent and $P(X_t = 1) = 1 - P(X_t = 0) = p$ for all $t$. Hence one should think of a sequence of coin flips where the event $\{X_t = 1\}$ means that the outcome of flip number $t$ is a 'head.' The underlying sample space $\Omega$ consists of all the sequences of the form

$$\omega = (0, 1, 0, 0, 1, 1, \ldots)$$

with each $\omega \in \Omega$ providing, in an obvious manner, a record of a possible sequence of flips.

Strictly speaking, the Bernoulli process features an infinite number of coinflips, so the vector $\omega$ has an infinite number of components and the sample space $\Omega$ has an infinite number of states. However, our securities market model features just a finite number $T$ of periods, so for our purposes it suffices to consider a 'modified' Bernoulli process corresponding to only $T$ coin flips. Now each state $\omega$ will have $T$ components, each being either a 0 or a 1. There are $2^T$ vectors like this, and the sample space for our modified Bernoulli process will have exactly one of each. Whether standard or modified, $X_t(\omega)$ will take the value 1 or 0 as the $t^{th}$ component of $\omega \in \Omega$ is 1 or 0, respectively. Moreover, $\mathscr{F}_t$ will be the algebra corresponding to the observations of the first $t$ coin flips, that is, $\mathscr{P}_t$ will be the partition consisting of $2^t$ cells, one for each possible sequence of $t$ coin flips. And the probability measure is given by $P(\omega) = p^n(1-p)^{T-n}$, where $\omega \in \Omega$ is any state corresponding to $n$ 'heads' and $T$-$n$ 'tails.'

The process $\{N_t; t = 1, 2, \ldots\}$ is defined in terms of the Bernoulli process (or the modified Bernoulli process) by setting

$$N_t(\omega) = X_1(\omega) + \ldots + X_t(\omega)$$

Hence the random variable $N_t$ should be thought of as the number of heads in the first $t$ coin flips (or as the number of up moves by the security during the first $t$ periods). Since $E[X_t] = p$ and $\mathrm{var}(X_t) = p(1-p)$, it follows that, for any $t$,

$$E[N_t] = tp$$

and

$$\mathrm{var}\ (N_t) = tp(1-p)$$

Moreover, it is not difficult to show that

**(3.29)** For all $t = 1, 2, \ldots$

$$P(N_t = n) = \binom{t}{n} p^n (1-p)^{t-n},\ n = 0, 1, \ldots, t.$$

This is called the *binomial probability distribution*, and

$$\binom{t}{n} = \frac{t!}{n!(t-n)!}$$

is called a *binomial coefficient*.

We are now ready to define the binomial security price model. This model features four parameters: $p$, $d$, $u$, and $S_0$, where $0 < p < 1, 0 < d < 1 < u$, and, of course, $S_0 > 0$. The time $t$ price of the security is given simply by

$$S_t = S_0\, u^{N_t}\, d^{t-N_t}, \quad t = 1, 2, \ldots, T$$

Hence, as advertised, each period there are two possibilities: either with probability $p$ the coin flip is heads and the price goes up by the factor $u$, or with probability $1-p$ it is tails and the price goes down by the factor $d$. Moreover, in view of (3.29), the probability distribution of $S_t$ is given by

$$P(S_t = S_0 u^n d^{t-n}) = \binom{t}{n} p^n (1-p)^{t-n}, \quad n = 0, 1, \ldots, t \tag{3.30}$$

With $2^T$ elements in the underlying sample space, the information tree terminates with $2^T$ nodes, as illustrated in figure 3.4. In particular, there are $2^T$ possible sample paths for the security price process. However, it is convenient to use a more compact diagram to illustrate the various possible sample paths. The event $\{S_t(\omega) = S_0 u^n d^{t-n}\}$ occurs if and only if exactly $n$ out of the first $t$ moves are 'up' moves; the order of these $t$ moves does not matter. For example, $\{S_2(\omega) = S_0 ud\}$ if either the first move is an 'up' and

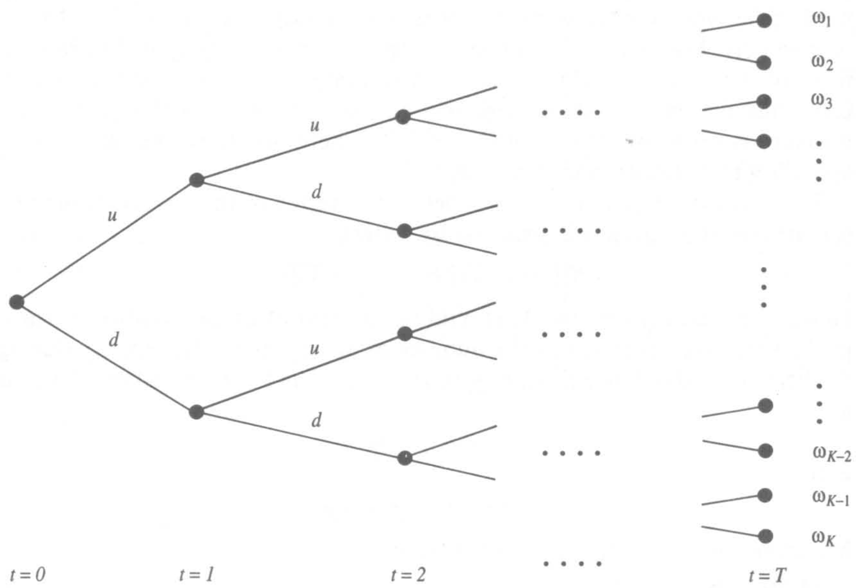


**Figure 3.4** Information tree for the binomial model ($K = 2^T$)



the second is a 'down,' or vice versa. As you can see from (3.30), at time $t$ the price process $S_t$ can take one of only $t+1$ possible values, although there are $2^t$ possible sample paths of length $t$. Figure 3.5 shows a network where there is one node corresponding to each event of the form $\{S_t(\omega) = S_0 u^n d^{t-n}\}$. Note that the number of ways this event can occur is equal to the number of paths to this event from the beginning node, and this is equal to $\binom{t}{n}$. While the kind of network shown in figure 3.5 is convenient for many purposes, it should not be confused with the information structure network as in figure 3.4.

A desirable feature of the binomial security price model is that its return process is given simply by

$$\Delta R_1(t) = u^{X_t}\, d^{1-X_t} - 1, \quad t = 1, 2, \ldots, T \tag{3.31}$$

In other words, either $\Delta R_1 = u - 1$ with probability $p$ or $\Delta R_1 = d - 1$ with probability $1-p$. In particular, the value of the return process is independent of the current price of the security, a feature that is often desirable when modeling securities such as common stocks.

What about the martingale measure? Assuming the interest rate is constant so that $\Delta R_0(t) = r$ for all $t$, by (3.26) and (3.31) we must have

$$q\left[\frac{u-1-r}{1+r}\right] + (1-q)\left[\frac{d-1-r}{1+r}\right] = 0$$

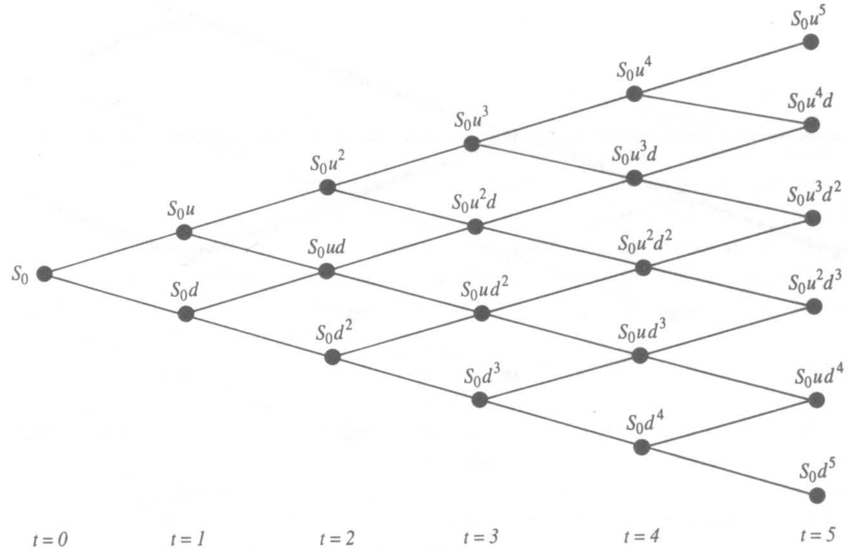


**Figure 3.5** Lattice showing price process for the binomial model ($T = 5$)

where $q$ is the conditional probability the next move is an 'up' move given the information $\mathscr{F}_t$ at any time $t$. Hence

$$q = \frac{1+r-d}{u-d}$$

for all $\mathscr{F}_t$ and $t$. Since we need $q < 1$, we realize that there will exist a martingale measure if and only if $u > 1+r$. In this case the martingale measure is given by

$$Q(\omega) = q^n(1-q)^{T-n}$$

where $\omega \in \Omega$ is any state corresponding to $n$ 'ups' and $T-n$ 'downs.' It follows that the probability distribution of $S_t$ under the risk neutral probability measure is given for all $t$ by

$$Q(S_t = S_o u^n d^{t-n}) = \binom{t}{n} q^n (1-q)^{t-n}, \; n = 0, 1, \ldots, t \tag{3.32}$$

The binomial model lends itself to some useful computations, such as the probability distribution for the maximum value achieved by the security process during the $T$ periods. We will derive this for the special case where $d = u^{-1}$, thereby leading to the simplification $S_t = S_o u^{2N_t - t}$. Define

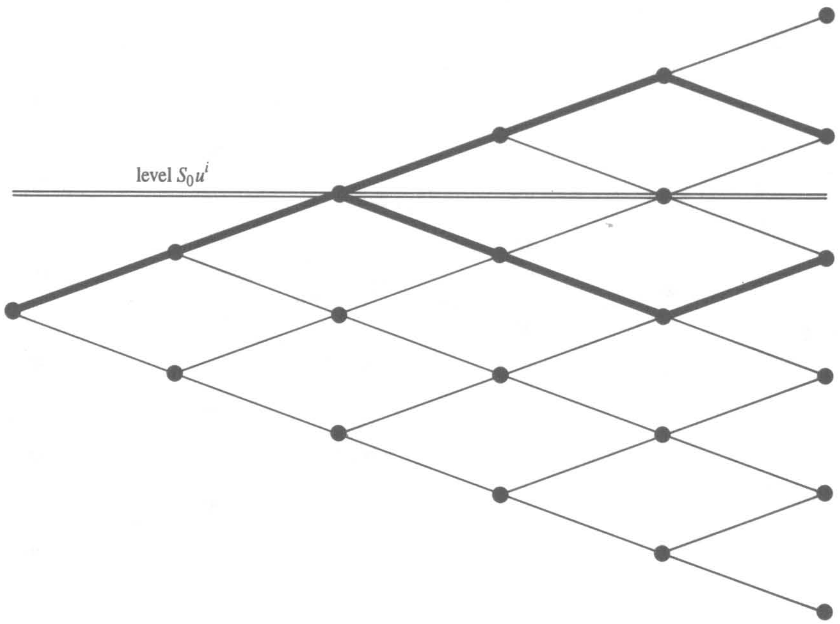


**Figure 3.6** The reflection principle



$Y_T = \max\{S_t : t = 0, 1, \ldots, T\}$, and note this random variable takes the $T+1$ values $S_0$, $S_0u, \ldots S_0u^T$. Our aim is to compute $P\{Y_T \geqslant S_0u^i\}$ for $i = 1, 2, \ldots, T$.

Fix $i$ and notice that $S_t \geqslant S_0u^i$ if and only if $2N_t - t \geqslant i$, so $P\{Y_T \geqslant S_0u^i\}$ is the same as $P\{2N_t - t \geqslant i$ for some $t\}$. We shall compute this latter probability with something called the *reflection principle*, as illustrated in figure 3.6. The idea is to define the *first passage time* $\tau_i \equiv \min\{t : 2N_t - t = i\}$, where $\tau = \infty$ if $2N_t - t < i$ for all $t \leqslant T$, and consider all the sample paths for which $\tau_i \leqslant T$. There are three mutually exclusive events. If $i$ equals one of the values $T$, $T-2$, $T-4, \ldots$, then it is possible to have $2N_T - T = i$, in which case, of course, $\tau_i \leqslant T$. Secondly, you can have $\tau_i < T$ and $2N_T - T > i$. Thirdly, you can have $\tau_i < T$ and $2N_T - T < i$. Hence

$$\begin{aligned} P\{Y_T \geqslant S_0u^i\} &= P\{2N_t - t \geqslant i \text{ for some t}\} \\ &= P\{\text{event } 1\} + P\{\textit{event } 2\} + P\{\textit{event } 3\} \end{aligned}$$

This first probability is

$$P\,\{\text{event } 1\} = P\{N_T = (T+i)/2\} = \binom{T}{\frac{T+i}{2}} p^{(T+i)/2}(1-p)^{(T-i)/2}$$

when $T+i$ is an even number, whereas $P\{\text{event } 1\} = 0$, otherwise. The second probability is easy, because if $2N_T - T > i$, then automatically $\tau_i < T$. Thus

$$P\{\text{event } 2\} = P\{N_T > (T+i)/2\} = \sum_{n=n^*}^{T} \binom{T}{n} p^n (1-p)^{T-n}$$

where $n^*$ denotes the smallest integer strictly greater than $(T+i)/2$ (this sum is taken to be zero if $n^* > T$).

Computing the third probability is more difficult, and this is where we use the reflection principle. The idea is that each sample path in event 2 is paired with a unique sample path in event 3, as illustrated in figure 3.6. The sample paths coincide up to time $\tau_i$, and then each is the mirror image of the other across the level $i$. Hence the number of sample paths in the two events are equal, although the probabilities of the two events are not equal unless $p = 1/2$.

To finish the computation of $P\{\text{event } 3\}$ we need to do some bookkeeping. Consider an arbitrary sample path from event 2, and suppose it is such that $N_T = n (\geqslant n^*)$. This sample path occurs with probability $p^n(1-p)^{T-n}$ and there are $\binom{T}{n}$ sample paths with $N_T = n$. Now upon looking at figure 3.6 it becomes apparent that the 'partner' of this sample path terminates with $N_T = T + i - n$, a symmetric distance below the level $(T+i)/2$. The probability of this 'partner' sample path is $p^{T+i-n}(1-p)^{n-i}$. Since there are $\binom{T}{n}$ sample paths in event 3 with $N_T = T + i - n$, it follows that

$$P\{\{\textit{event } 3\} \cap \{N_T = T + i - n\}\} = \binom{T}{n} p^{T+i-n}(1-p)^{n-i}$$

in which case

$$P\{event\ 3\} = \sum_{n=n^*}^{T} \binom{T}{n} p^{T+i-n}(1-p)^{n-i}$$

Hence we finally have

$$(3.33)\quad P\{Y_T \geqslant S_0 u^i\} = \binom{T}{\frac{T+i}{2}} p^{(T+i)/2}(1-p)^{(T-i)/2} + \sum_{n=n^*}^{T} \binom{T}{n} [p^n(1-p)^{T-n} + p^{T+i-n}(1-p)^{n-i}]$$

where the first term on the right hand side is zero when $T+i$ is an odd number and where $n^*$ denotes the smallest integer strictly greater than $(T+i)/2$.

We now have, in principle, the probability distribution for the maximum security price during the $T$ periods. More generally, these formulas can also be used for the maximum security price during the first $t$ periods when $t < T$. Since the event $\{Y_t \geqslant S_0 u^i\}$ is the same as the event $\{T_i \leqslant t\}$, we also have the probability distribution for the first passage time to security price level $S_0 u^i$.

Similar formulas can be derived for the probability distributions of the minimum security price and the first passage times to price levels below $S_0$.

---

**Exercise 3.13** Derive (3.29).

**Exercise 3.14** For the case $T = 4$ and $d = 1/u$, compute the probability distributions of $Y_T$ and $\tau_2$.

---

## 3.6 Markov Models

This section will introduce a class of stochastic processes that share what is called the 'Markov property:' the future is independent of the past, given the present values of the process. Markov processes are important models of security prices, because they are often realistic representations of true prices and yet the Markov property leads to simplified computations.

Throughout this section the filtration $\mathbb{F} = \{\mathcal{F}_t;\ t = 0, 1, \ldots, T\}$ is generated by a stochastic process $X = \{X_t;\ t = 0, 1, \ldots, T\}$. This process takes values in some finite set $E$, called the *state space*. If $X_t = j \in E$, we shall say 'the process is in state $j$ at time $t$.' The most common situation is for the state to be a scalar, but frequently it is more convenient for the state to be a vector. As usual, there is a sample space $\Omega$ and a probability measure $P$ on it, and the information $\mathcal{F}_t$ should be thought of as the history of the present and past values of the process $X$.



The stochastic process $X$ is said to be a *Markov chain* if

$$P\{X_{t+1} = j | \mathcal{F}_t\} = P\{X_{t+1} = j | X_t\}$$

for all $j \in E$ and all $t$. By elementary probability calculations, it follows that

$$(3.34)\quad P\{X_{t+s} = j | \mathcal{F}_t\} = P\{X_{t+s} = j | X_t\}, \text{ all } s \geqslant 1$$

Thus a Markov chain is a stochastic process where the only information useful for predicting future values is the current state; in other words, given the history of the process, the past values can be ignored as long as you know the present state.

The Markov chain $X$ is said to be *stationary* or *time-homogeneous* if the conditional probabilities $P\{X_{t+1} = j | \mathcal{F}_t\}$ do not depend on time $t$. In this case it is convenient to define the *transition probabilities*

$$P(i,j) \equiv P\{X_{t+1} = j | X_t = i\},\ i, j \in E$$

and to organize them into a *transition matrix*

$$P \equiv \left[P(i,j)\right]$$

Note this is a square matrix with the number of rows equal to the number of elements in the state space $E$. Moreover, the sum of the elements in each row of $P$ equals one. It should be clear from the context whether $P$ denotes the probability measure, a transition probability, or the transition matrix.

If the Markov chain is not stationary, then one can still talk about the transition probabilities, only now they depend on time: $P_t(i,j) \equiv P\{X_{t+1} = j | X_t = i\}$. In this more general case there is a distinct transition matrix for each point in time:

$$P_t \equiv \left[P_t(i,j)\right]$$

**Example 3.5** The process $N = \{N_t;\ t = 1, \ldots, T\}$ studied in section 3.5 and representing the number of heads in $t$ flips of a coin that lands 'head' with probability $p$ is an example of a stationary Markov chain. For the state space it is convenient to take $\{0, 1, \ldots, T\}$, in which case the transition matrix is

$$P = \begin{pmatrix} 1-p & p & 0 & \ldots & 0 \\ 0 & 1-p & p & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

Since this Markov chain starts in state 0, it can only reach state $T$ at the final time period $T$, and so the transition probabilities from state $T$ do not really matter, except that they must add up to one. We arbitrarily took $P(T, T) = 1$, meaning that the Markov chain would remain in state $T$ forever, even if it were to keep operating after time $T$.

A useful property of Markov chains is provided by the following:

**(3.35)** If $Y = f(X_t, X_{t+1}, \ldots, X_T)$ for some function $f$, then

$$E[Y|\mathscr{F}_t] = E[Y|X_t]$$

This says the future and the past are conditionally independent, given the present.

Now suppose we have a securities market model where the discounted price process is a Markov chain. This means that $P\{S^*_{t+1} = j|\mathscr{F}_t\} = P\{S^*_{t+1} = j|S^*_t\}$ for all $j \in E$ and $t$. It is natural to wonder whether anything special can be said about the martingale measure, if there is one. In particular, one should ask whether the discounted security prices are Markov chains under the martingale measure. This is a non-trivial question, because a stochastic process may lose its Markov property when you change from one probability measure to an equivalent one. For example, you can make the third coin flip depend on the first coin flip just by changing the probability measure associated with the heads counting process $N_t$ in example 3.5.

It turns out, however, that our question can be answered in the affirmative:

**(3.36)** If there are no arbitrage opportunities, if the discounted price process $S^*$ is a Markov chain under $P$, and if the filtration $\mathbb{F}$ is the one generated by $S^*$, then there exists a martingale measure $Q$ under which $S^*$ is a Markov chain.

To see this, suppose the Markov chain (under $P$) is stationary, and consider the construction of the martingale measure as developed in section 3.4. In particular, consider the conditional probabilities associated with the 'single period model' associated with an individual node of the information tree. If the corresponding, current state of the discounted price process $S^*$ is $s$, then the conditional, risk neutral probabilities must be, of course, such that the expected value of $S^*$ at the end of the period is equal to $s$. But there is a one-to-one correspondence between each branch leading out of this node and each transition probability $P(s, j)$ from state $s$ that is strictly positive. In other words, $P(s,j) > 0$ if and only if $j$ is a possible state for $S^*$ at the end of the period. Hence choosing the risk neutral conditional probabilities for this node amounts to suitably choosing new transition probabilities $Q(s, j)$ from state $s$, making sure $Q(s, j) > 0$ if and only if $P(s, j) > 0$ for all $j$. By hypothesis, we know this choice can be made.

Now there may be many other nodes in the information tree where the corresponding, current state of $S^*$ is $s$, but the same situation will exist. The single period models will all be identical, and so you can choose the same set of risk neutral transition probabilities $Q(s, j)$ for all these nodes.

In a similar manner, one chooses the risk neutral transition probabilities for all the other nodes in the information tree, making sure



the probabilities are the same for all the nodes that share the same value for the current state of the discounted price process. Hence not only does this lead to the risk neutral probability measure $Q$ as was derived in section 3.4, but these transition probabilities form a Markov transition matrix under which $S^*$ is a Markov chain. Hence (3.36) is true when $S^*$ is a stationary Markov chain under $P$, and this argument can be easily extended to cover the case where $S^*$ is a Markov chain that is not stationary.

It follows from (3.35) and (3.36) that $E_Q[S^*_n(t+s)|\mathscr{F}_t] = E_Q[S^*_n(t+s)|S^*_t]$. Hence the standard relationship for $Q$ to be a martingale measure can be written as

$$S^*_n(t) = E_Q[S^*_n(t+s)|S^*_t], \qquad \text{all } n,\ t, \text{ and } s$$

If the original price process $S_t$ is a Markov chain under $P$ and the bank account $B_t$ is deterministic, then the discounted price process $S^*_t = S_t/B_t$ will also be a Markov chain. However, if the bank account process is stochastic, even a Markov chain, then the discounted price process will not necessarily be a Markov chain. In this case all is not lost; it might still be possible to take advantage of the Markov properties of the model.

One way to proceed when $B_t$ is stochastic is to set up the securities market model with an underlying stochastic process $X$ that is a stationary Markov chain under $P$ and with the filtration being generated by this process. For example, $X_t$ could be a vector of the current prices of all the relevant securities. In doing this, one would normally fix the initial state $X_0 = i_0$ and then let the sample space $\Omega$ be the set of all the possible sample paths of $X$. Then if $\omega \in \Omega$ corresponds to the sample path $(i_0, i_1, i_2, \ldots, i_T)$, one would take the probability measure to be such that $P\{\omega\} = P(i_0, i_1)P(i_1, i_2) \ldots P(i_{T-1}, i_T)$.

Next, suppose for each $t$ that $f_t$ is a real valued function on the state space $E$ such that the bank account process satisfies $B_t = f_t(X_t)$. If the risky security price processes are defined in a similar manner, then the discounted price processes can also be expressed as functions of the underlying Markov chain $X$, being the ratio of two such functions. Now none of these price process is necessarily a Markov chain; for example, the function $f$ may give rise to the same price for two distinct states in $E$. However, we may re-do the argument for (3.36) and recover a useful result.

Again, we look at the single period model associated with each node of the information tree, only now associated with each node is the current state of the Markov chain $X$. We construct the conditional probabilities $Q(i, j)$ and the martingale measure $Q$ as before, only now $X$ will turn out to be a Markov chain under $Q$. Hence by (3.35) we will have for the resulting $Q$ that $E_Q[S^*_n(t+s)|\mathscr{F}_t] = E_Q[S^*_n(t+s)|X_t]$, in which case the standard relationship for $Q$ to be a martingale measure can be written as

$$S^*_n(t) = E_Q[S^*_n(t+s)|X_t], \qquad \text{all, } n,\ t, \text{ and } s$$

**Example 3.5 (continued)** Taking $f_t(x) = S_0 u^x d^{t-x}$ we have for the binomial security price model of section 3.5 $S_t = f_t(N_t)$, where the

head counting process $N_t$ is a Markov chain, as discussed earlier. Actually, $S_t$ is also a stationary Markov chain, even though $f_t$ depends on $t$, because the future changes in the price process depend only on the future 'coin flips,' which are independent of time. If $s$ is the state where $S_t = s$, then the transition probability $P(s,j) = p$ if $j = \text{su}$, $P(s,j) = 1 - p$ if $j = sd$, and $P(s,j) = 0$, otherwise. For general parameters $u$ and $d$ the state space is rather messy, as it can contain up to $(T+1)(T+2)/2$ distinct values, that is, the number of nodes in the network such as figure 3.4. However, in the important special case where $d = u^{-1}$, the state space has only $2T + 1$ distinct values, as can be seen by studying figure 3.5.

To demonstrate the preceding equations which the martingale measure $Q$ must satisfy, we have $S_t^* = S_0 u^{N_t} d^{t-N_t}/(1+r)^t$ and

$$\begin{aligned} E_Q[S_{t+1}^* | S_t^*] &= E_Q[S_0 u^{X_{t+1}+N_t} d^{t+1-X_{t+1}-N_t}/(1+r)^{t+1} | S_t^*] \\ &= (S_t^*/(1+r)) E_Q[u^{X_{t+1}} d^{1-X_{t+1}} | S_t^*] \\ &= (S_t^*/(1+r)) E_Q[u^{X_{t+1}} d^{1-X_{t+1}}] \\ &= (S_t^*/(1+r)) \left[ \frac{1+r-d}{u-d} u + \frac{u-1-r}{u-d} d \right] = S_t^* \end{aligned}$$

Similarly, one verifies $E_Q[S_{t+1}^* | N_t] = S_t^*$.

A virtue of Markov chains is that it is straightforward to compute conditional probability distributions for the state of the Markov chain at a specified number of periods in the future. For example,

$$\begin{aligned} P\{X_2 = j | X_o = i\} &= \sum_{k \in E} P\{X_2 = j,\ X_1 = k | X_0 = i\} \\ &= \sum_{k \in E} P\{X_2 = j |\ X_1 = k,\ X_0 = i\} P\{X_1 = k | X_0 = i\} \\ &= \sum_{k \in E} P\{X_2 = j | X_1 = k\} P\{X_1 = k | X_0 = i\} \\ &= \sum_{k \in E} P_1(k,j) P_0(i,k) \end{aligned}$$

We recognize this last expression as the result of matrix multiplication. In other words, $P\{X_{t+2} = j | X_t = i\}$ will be equal to the $ij$th element of the matrix product $P_t P_{t+1}$. This pattern extends to any number of periods: $P\{X_{t+s} = j | X_t = i\}$ will be equal to the $ij$th element of the matrix product $P_t P_{t+1} \ldots P_{t+s-1}$. Of course, in the stationary case one simply has $P_t P_{t+1} \ldots P_{t+s-1} = P^s$.

**Example 3.6** Consider the binomial security process model with $d = u^{-1}$; the price process $S_t$ is a Markov chain with state space



$E = \{S_0 u^{-T}, S_0 u^{-T+1}, \ldots, S_0 u^{-1}, S_0, S_0 u, \ldots, S_0 u^{T-1}, S_0 u^T\}$. But suppose the conditional probability of an 'up' move varies with the state. In particular, suppose $P\{S_{t+1} = S_0 u^{i-1} | S_t = S_0 u^i\} = 1 - p_i$ and $P\{S_{t+1} = S_0 u^{i+1} | S_t = S_0 u^i\} = p_i$ for $2T - 1$ parameters $p_i, i = -T+1, \ldots, -1, 0, 1, \ldots, T-1$. (For example, the model builder could give price level $S_0$ a measure of stability by setting $p_i > 1/2$ for $i < 0$ and $p_i < 1/2$ for $i > 0$). Thus the transition matrix is

$$\begin{bmatrix} 1 & 0 & 0 & & & & & & & & \\ 1-p_{-T+1} & 0 & p_{-T+1} & & & & & & & & \\ \cdot & \cdot & \cdot & & & & & & & & \\ & & 1-p_{-1} & 0 & p_{-1} & & & & & & \\ & & & 1-p_0 & 0 & p_0 & & & & & \\ & & & & 1-p_1 & 0 & p_1 & & & & \\ & & & & & & & \cdot & \cdot & \cdot & \\ & & & & & & & & 1-p_{T-1} & 0 & p_{T-1} \\ & & & & & & & & 0 & 0 & 1 \end{bmatrix}$$

and, for example, $P\{S_3 = S_0 u\} = (1 - p_0) p_{-1} p_0 + p_0 (1 - p_1) p_0 + p_0\, p_1 (1 - p_2)$.

**Exercise 3.15** Derive (3.35).

**Exercise 3.16** Derive (3.34). (Hint: Use (3.35))

**Exercise 3.17** Prove by mathematical induction for a stationary Markov chain $X$ with transition matrix $P$ that $P\{X_{t+s} = j | X_t + i\}$ is equal to the $ij$th element of the $s$- fold matrix product of $P$.

# 4 Options, Futures, and Other Derivatives



## 4.1 Contingent Claims

A *contingent claim* is a random variable $X$ that represents the time $T$ payoff from a 'seller' to a 'buyer.' This definition is essentially the same as for single period models, and it turns out that the basic ideas are very similar as well. However, much more can be said, primarily because the multiperiod setting leads to a number of rich examples, many of which see practical use in the financial industry.

In most instances the random variable $X$ can be taken to be some function of an underlying security price, and so contingent claims are examples of what are called *derivative securities.* With a single period model, contingent claims are about the only kind of derivative security you can think of. But with several periods to work with it is possible to consider other kinds of derivative securities, that is, securities whose values depend on underlying securities but which cannot be modeled as a time $T$ payoff $X$. Other kinds of derivatives will be discussed later in this chapter.

As with single period models, a contingent claim is like a contract or agreement between two parties. Since one party (the seller) promises to pay the other party (the buyer) the amount $X$ at time $T$, the buyer will normally pay some money to the seller when they make their agreement, say at time $t < T$. The fundamental question to be addressed is: what is the appropriate value for this time $t$ payment? In other words, what is the time $t$ value of this contingent claim?

In general, the time $T$ payoff $X$ can be strictly negative for some states of the world $\omega \in \Omega$. This amounts to a payment by the buyer to the seller. In contrast, of considerable importance is the case where the buyer has the option at time $T$ to proceed with the payment at that time. In this case, called



a *European option*, the payment will naturally take place if and only if it is positive. Consequently, European options are the same things as non-negative contingent claims.

**Example 4.1** Consider the simple model with $T = 2$ and $K = 4$ that was introduced as example 3.3 in chapter 3. The single risky security is as follows:

| $\omega_k$ | $t=0$ | $t=1$ | $t=2$ |
|---|---|---|---|
| $\omega_1$ | $S_0 = 5$ | $S_1 = 8$ | $S_2 = 9$ |
| $\omega_2$ | $S_0 = 5$ | $S_1 = 8$ | $S_2 = 6$ |
| $\omega_3$ | $S_0 = 5$ | $S_1 = 4$ | $S_2 = 6$ |
| $\omega_4$ | $S_0 = 5$ | $S_1 = 4$ | $S_2 = 3$ |

If $X = S_2 - 5$, the net profit at time $T = 2$ for purchasing one unit of the security for the price 5, then the payoff is negative in state $\omega_4$. But if this is a European option, then one should take $X = \max\{S_2 - 5, 0\}$, so $X(\omega)$ takes the values 4, 1, 1, and 0 in states $\omega_1$, $\omega_2$, $\omega_3$, and $\omega_4$, respectively. This European option is called a *call option* with exercise price 5. Similarly, $X = \max\{e - S_2, 0\}$ is a *put option* with exercise price $e$, that is, $X$ is the option to sell one unit of the security for the price $e$ at time $T = 2$.

**Example 4.2** Suppose for the same securities market model as in example 4.1 that $X = \max\{[S_0 + S_1 + S_2]/3 - 5, 0\}$, so that $X(\omega)$ takes the values 7/3, 4/3, 0, and 0 in states $\omega_1$, $\omega_2$, $\omega_3$, and $\omega_4$, respectively. Now $X$ is an example of what is called an *Asian* or *averaging* option. Options like this are used to hedge against rising prices for parties who need to buy fixed quantities of the security every period. Unlike put and call options, where the value of $X$ depends only on the final value of the security, here the value of $X$ depends on the whole history of the security.

Throughout this chapter it will be assumed that the securities market model is economically reasonable, that is, there exists a risk neutral probability measure $Q$. A contingent claim is said to be *marketable* or *attainable* if there exists a self-financing trading strategy such that $V_T(\omega) = X(\omega)$ for all $\omega \in \Omega$. The corresponding portfolio or trading strategy $H$ is said to *replicate* or *generate* $X$. In the case of single period models, the time $t = 0$ value of a contingent claim was seen to be the expectation under a risk neutral probability measure of the discounted value of the claim. In the multiperiod case this conclusion generalizes slightly to:

**(4.1)** *Risk neutral valuation principle*: The time $t$ value of a marketable contingent claim $X$ is equal to $V_t$, the time $t$ value of the portfolio which replicates $X$. Moreover,

$$V_t^* = V_t/B_t = E_Q[X/B_T|\mathscr{F}_t], \quad t = 0, 1, \ldots, T$$

for all risk neutral probability measures $Q$.

The first statement in (4.1) is a consequence of the law of one price. Indeed (as will be discussed more fully below), if $X$ could be purchased or sold at time $t$ for an amount other than $V_t$, then one could exploit an arbitrage opportunity by taking suitable positions in this contingent claim and the other securities. Since $V_T = X$ if and only if $V_T^* = X/B_T$, the equation in (4.1) follows from the fact that $V^*$ is a martingale under $Q$ (see (3.21)).

**Example 4.1 (continued)** Suppose $r = 0$ (see example 3.3) so $Q = (1/6, 1/12, 1/4, 1/2)$. If the call option with exercise price $e = 5$ is attainable, then its time $t = 0$ value must be

$$V_0 = E_Q[X] = \frac{1}{6}(4) + \frac{1}{12}(1) + \frac{1}{4}(1) + \frac{1}{2}(0) = 1$$

In states $\omega_1$ and $\omega_2$ we have

$$V_1 = E_Q[X|S_1 = 8] = \frac{2}{3}(4) + \frac{1}{3}(1) = 3$$

while in states $\omega_3$ and $\omega_4$ we have

$$V_1 = E_Q[X|S_1 = 4] = \frac{1}{3}(1) + \frac{2}{3}(0) = \frac{1}{3}$$

Similarly, the put with exercise price $e = 5$ pays off the amounts 0, 0, 0, and 2 in states $\omega_1$ to $\omega_4$, respectively, so $V_0 = 1$, $V_1 = 0$ in states $\omega_1$ and $\omega_2$, and $V_1 = 4/3$ in states $\omega_3$ and $\omega_4$.

It is important to be able to compute the trading strategy which generates a particular contingent claim. For one thing, this verifies that the contingent claim is indeed attainable. Moreover, even if you know the contingent claim is attainable, you may want to use the replicating trading strategy, perhaps to hedge a position in the contingent claim.

There are several good methods for computing replicating trading strategies. For the first method you already know the value process $V$ for the replicating portfolio, and so you solve for the trading strategy $H$ using the linear equations (one for each state ) in the definition of the value process

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

and keeping in mind that $H$ is predictable. This is illustrated in the following example:

**Example 4.1 (continued)** For $t = 2$ we have

$$\begin{aligned}
V_2(\omega_1) &= 4 = H_0(2)(\omega_1)1 + H_1(2)(\omega_1)9 \\
V_2(\omega_2) &= 1 = H_0(2)(\omega_2)1 + H_1(2)(\omega_2)6 \\
V_2(\omega_3) &= 1 = H_0(2)(\omega_3)1 + H_1(2)(\omega_3)6 \\
V_2(\omega_4) &= 0 = H_0(2)(\omega_4)1 + H_1(2)(\omega_4)3
\end{aligned}$$



Since $H$ is predictable we also have

$$\begin{aligned}
H_0(2)(\omega_1)1 &= H_0(2)(\omega_2) \quad H_0(2)(\omega_3)1 = H_0(2)(\omega_4) \\
H_1(2)(\omega_1)1 &= H_1(2)(\omega_2) \quad H_1(2)(\omega_3)1 = H_1(2)(\omega_4)
\end{aligned}$$

Solving these equations yields $H_0(2) = -5$ and $H_1(2) = 1$ in states $\omega_1$ and $\omega_2$, whereas $H_0(2) = -1$ and $H_1(2) = 1/3$ in states $\omega_3$ and $\omega_4$. Meanwhile, for $t = 1$ we have

$$\begin{aligned}
V_1(\omega) &= 3 = H_0(1)(\omega)1 + H_1(1)(\omega)8, \quad \omega = \omega_1, \omega_2 \\
V_1(\omega) &= \frac{1}{3} = H_0(1)(\omega)1 + H_1(1)(\omega)4, \quad \omega = \omega_3, \omega_4 \\
H_0(1)(\omega_1) &= H_0(1)(\omega_2) = H_0(1)(\omega_3) = H_0(1)(\omega_4) \\
H_1(1)(\omega_1) &= H_1(1)(\omega_2) = H_1(1)(\omega_3) = H_1(1)(\omega_4)
\end{aligned}$$

Solving these gives $H_0(1) = -7/3$ and $H_1(1) = 2/3$ for all $\omega$.

For the second method for computing a replicating trading strategy, suppose all you know is $X$. You then work backwards in time, deriving $V$ and $H$ simultaneously. Since $V_T = X$, you first solve

$$X = H_0(T)B_T + \sum_{n=1}^{N} H_n(T)S_n(T)$$

for $H(T)$ as in the first method. Since $H$ is self-financing, it follows that

$$V_{T-1} = H_0(T)B_{T-1} + \sum_{n=1}^{N} H_n(T)S_n(T-1)$$

so now you know $V_{T-1}$.

Next you solve

$$V_{T-1} = H_0(T-1)B_{T-1} + \sum_{n=1}^{N} H_n(T-1)S_n(T-1)$$

for $H(T-1)$ as in the first method, and then you compute $V_{T-2}$. You repeat this cycle working backwards in time until, finally, you compute $V_0$.

**Example 4.1 (continued)** Using this second method we first compute $H_0(2) = -5$ and $H_1(2) = 1$ for states $\omega_1$ and $\omega_2$, so $V_1 = -5(1)+1(8) = 3$ in these same two states. Similarly, $H_0(2) = -1$ and $H_1(2) = 1/3$ in states $\omega_3$ and $\omega_4$, so $V_1 = -1(1) + (1/3)(4) = 1/3$ in these same two states. Next we use these values of $V_1$ to compute $H_0(1) = -7/3$ and $H_1(1) = 2/3$ for all $\omega$. Finally, we see that $V_0 = (-7/3)(1) + (2/3)(5) = 1$ for all $\omega$.

Still another method for computing a replicating strategy involves working with the discounted prices and the discounted value process. The self-financing equation $V_0^* + G_t^* = V_t^*$ is the same thing as

$$V_{t-1}^*(\omega) + \sum_{n=1}^{N} H_n(t)(\omega)\, \Delta S_n^*(t)(\omega) = V_t^*(\omega)$$

So if you know $V_t^*$, then you can use this system of equations together with the predictability requirement to solve for $V_{t-1}^*$ along with the positions $H_1(t)\ldots, H_N(t)$. Hence you begin with $t = T$ and $V_T^* = X/B_T$ and you work backwards in time, computing the discounted value process $V^*$ of the replicating portfolio as well as the replicating trading strategy's positions in all the risky securities. This gives $V = V^*B$. Finally, using the definition of either $V$ or $V^*$, you compute $H_0$, the positions in the bank account. This is illustrated in the following example.

**Example 4.2 (continued)** Assuming $r = 0$, the equations are

$$V_1^* + H_1(2)(1) = 7/3$$

and

$$V_1^* + H_1(2)(-2) = 4/3$$

in states $\omega_1$ and $\omega_2$, respectively. Hence $V_1^* = 2$ and $H_1(2) = 1/3$ in these same states. Similarly, the equations

$$V_1^* + H_1(2)(2) = 0$$

and

$$V_1^* + H_1(2)(-1) = 0$$

give $V_1^* = 0$ and $H_1(2) = 0$ in states $\omega_3$ and $\omega_4$.

For the second iteration we have

$$V_0^*(\omega) + H_1(1)(\omega)(3) = 2, \quad \omega = \omega_1, \omega_2$$

$$V_0^*(\omega) + H_1(1)(\omega)(-1) = 0, \quad \omega = \omega_3, \omega_4$$

Hence $V_0^* = 1/2$ and $H_1(1) = 1/2$ for all $\omega$.

It remains to compute $H_0$. Using $H_0(1) = V_0^* - H_1(1)S_0$ we get $H_0(1) = -2$. Similarly, we get $H_0(2) = -2/3$ in states $\omega_1$ and $\omega_2$, and we get $H_0(2) = 0$ in states $\omega_3$ and $\omega_4$. Note that $E_Q[X] = (1/6)(7/3) + (1/12)(4/3) = 1/2$, the same as the value of $V_0^*$ that was already computed.

We now return to the earlier discussion of arbitrage pricing. If the actual traded price of a contingent claim differs from the value of the replicating portfolio, then one can find an arbitrage opportunity. To see this, let $P_t$ denote the actual time $t$ price of the contingent claim.

First suppose $P_t > V_t$. Then sell the contingent claim for $P_t$, collecting this amount. Simultaneously, begin the replicating trading strategy with an initial amount of capital equal to $V_t$. Invest the difference, $P_t - V_t$, at the bank



account rate. At time $T$ your liability on the contingent claim will be $P_T = X$, but this will coincide exactly with your replicating portfolio. One will precisely offset the other. Meanwhile, the investment in the bank account has become $(P_t - V_t)B_T/B_t > 0$, a sure profit.

On the other hand, if $P_t < V_t$, then you do the opposite of these transactions. You buy the contingent claim, follow the negative of the replicating trading strategy (thereby collecting $V_t$ dollars), and invest the difference $V_t - P_t$ in the bank account. At time $T$ your liability $V_T$ is exactly offset by the value $P_T = X$ of the contingent claim. Meanwhile, you now have $(V_t - P_t)B_T/B_t > 0$ in the bank.

Turning to another topic, if the bank account process $B$ is deterministic (see exercise 4.3) then a call option on a security is marketable if and only if the put option on the same security with the same exercise price is marketable. If they are both marketable, then one has the *put-call parity* relationship:

$$p = c + e\, E_Q[1/B_T] - S_0 \tag{4.2}$$

where $c$ and $p$ are the time $t = 0$ prices of the call and put options, respectively, both having the common exercise price $e$ and the underlying security $S$. This is one in a number of examples where the price of the option of interest (in this case the put) can be expressed as a linear combination of the prices of one or more conventional calls, the price of the underlying security, and the 'forward' price $E_Q[1/B_T]$. Here is another example.

**Example 4.3** Suppose the buyer acquires at time 0 an option which provides the right to choose at fixed time $t$, where $0 < t < T$, between a call option and a put option, both having the same exercise price $e$ and expiring at time $T$. This is called a *chooser* option. If $C_t$ and $P_t$ denote their respective time $t$ prices, then the option buyer will choose the call if and only if $C_t \geqslant P_t$, in which case the time $T$ payoff will be

$$(S_T - e)^+ 1_{\{C_t \geqslant P_t\}} + (e - S_T)^+ 1_{\{C_t < P_t\}}$$

where $1_A$, the indicator function of the event $A$, equals one if event $A$ occurs and equals zero if it does not. We are interested in computing the time 0 value of this option with the risk neutral valuation principle (4.1).

Now by adding and subtracting the term $(S_T - e)^+ 1_{\{C_t < P_t\}}$, it follows that the time $T$ payoff of the chooser option is equal to

$$(S_T - e)^+ + (e - S_T) 1_{\{C_t < P_t\}}$$

Hence by risk neutral valuation, if this chooser option is marketable, then its time 0 price is equal to the price of the ordinary call option $(S_T - e)^+$ plus the quantity

$$E_Q[(e - S_T) 1_{\{C_t < P_t\}} / B_T] \tag{4.3}$$

Assuming the interest rate $r$ is constant and the call option $(S_T - e)^+$ is marketable, expression (4.3) has a simple form. The put-call parity relationship (4.2) applies at time $t$, so the event $C_t < P_t$ is the same as the event $S_t < e(1+r)^{t-T}$. Hence

$$\begin{aligned}
E_Q[(e-S_T)1_{\{C_t<P_t\}}/B_T] &= E_Q[E_Q[(e-S_T)1_{\{C_t<P_t\}}/(1+r)^T|\mathscr{F}_t]] \\
&= E_Q[1_{\{C_t<P_t\}}E_Q[(e-S_T)/(1+r)^T|\mathscr{F}_t]] \\
&= E_Q[1_{\{C_t<P_t\}}(1+r)^{-t}\{e(1+r)^{t-T} - E_Q[S_T(1+r)^{t-T}|\mathscr{F}_t]\}] \\
&= E_Q[1_{\{C_t<P_t\}}(1+r)^{-t}[e(1+r)^{t-T} - S_t]] \\
&= E_Q[1_{\{S_t<e(1+r)^{t-T}\}}[e(1+r)^{t-T} - S_t]/(1+r)^t\,]
\end{aligned}$$

where the next to last equality uses the fact that the discounted price process $S^* = S/B$ is a martingale under $Q$. We recognize the final expression to be the time 0 price of an ordinary put option that has exercise price $e(1+r)^{t-T}$ and expires at time $t$. Hence under the indicated assumptions the time 0 price of the chooser option equals the price of a certain ordinary call plus the price of a certain ordinary put, with the latter expressable in terms of another call by the put-call parity relationship.

**Example 4.1 (continued)** With interest rate $r = 0$ and exercise price $e = 5$, consider the chooser option where the buyer chooses between the put and call at time $t = 1$. By our earlier calculations for the ordinary put and call, the chooser option buyer will choose the call if $S_1 = 8$ and will choose the put if $S_1 = 4$. Hence the chooser option will pay off the amounts 4, 1, 0, and 2 in states $\omega_1$ to $\omega_4$, respectively, at time 2. The time 0 price of this chooser option is

$$V_0 = \frac{1}{6}(4) + \frac{1}{12}(1) + \frac{1}{4}(0) + \frac{1}{2}(2) = 1\frac{3}{4}$$

Meanwhile, the put with exercise price 5 that expires at time 1 will pay off the amount 0 if $S_1 = 8$ and the amount 1 if $S_1 = 4$, so its time 0 price is $\frac{1}{4}(0) + \frac{3}{4}(1) = \frac{3}{4}$. The time 0 price of the call that expires at time 2 was calculated earlier and found to be 1, so this illustrates that the price of the chooser option equals the price of this put plus the price of this call.

Of course, options can be defined in terms of two or more underlying securities. For example, given a function $g : \mathbb{R}^N \to \mathbb{R}^+$ one can take $X = g(S_1(T), \ldots, S_N(T))$. Then if you know the joint probability distribution of the random variables $S_1(T), \ldots S_N(T)$ under the martingale measure, it is easy to compute the time 0 value of this contingent claim. In particular, with

$$g(s_1, \ldots, s_N) = (a_1 s_1 + \ldots + a_N s_N - e)^+$$



for positive scalars $a_1, \ldots a_N$ you could have a call option on a stock index. Or with

$$g(s_1, \ldots, s_N) = \max\{s_1, \ldots, s_N, e\}$$

you would have a contingent claim delivering the best of $N$ securities and the cash amount $e$.

**Example 4.4** Suppose $K = 9$, $N = 2$, $T = 2$, $r = 0$, and the price processes and information structure are as displayed in figure 4.1. There is a unique risk neutral probability measure $Q$ for this model; it is also displayed in figure 4.1. Now consider a call option with exercise price $e = 13$ on the time $T = 2$ value of the stock index $S_1(t) + S_2(t)$. In other words, this contingent claim $X = (S_1(2) + S_2(2) - 13)^+$ and is displayed in figure 4.1. As will be explained in section 4.4, $X$ is marketable. Hence its time 0 price is easily computed to be $E_Q X = 19/18$.

---

**Exercise 4.1** Consider the usual two-period model as in example 4.1 with $r = 0$. Compute the time $t = 0$ values and the replicating trading strategies for the following European options:

(a) A call option with exercise price 7.
(b) A put option with exercise price 7.
(c) An Asian put option with exercise price 7.
(d) A chooser option with exercise price 7 and decision time 1.

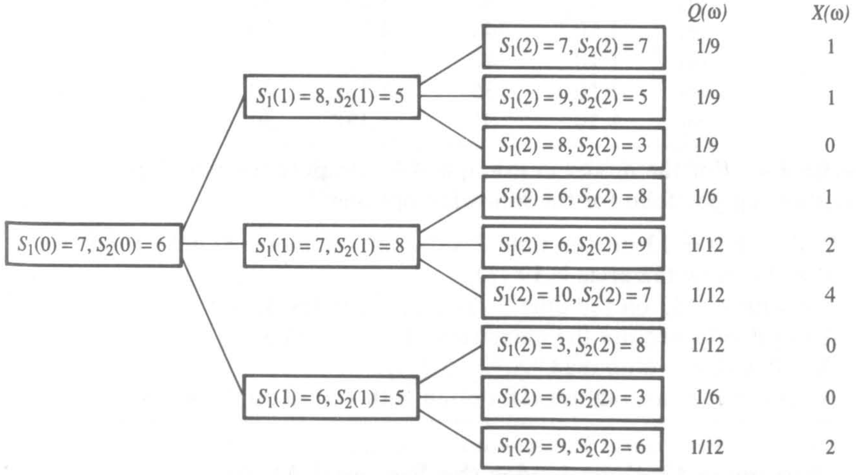


**Figure 4.1** Data for example 4.4

**Exercise 4.2** A *look-back* option is one where the payoff is based on the maximum (or perhaps the minimum) price of the underlying security during a recent time interval. Compute the time $t = 0$ value and the replicating strategy for the following look-back call option: $X = \max\{0, S_0 - 7, S_1 - 7, S_2 - 7\}$. As usual, work with the two period model of example 4.1 with $r = 0$.

**Exercise 4.3** Consider a European put and a European call option on the same security $S$, where they have the same expiration date $T$ and the same exercise price $e$.

(a) Show that (4.2) holds if both the put and call are marketable (Hint: use $(S_T - e)^+ - (e - S_T)^+ = S_T - e$ to show that the claim with time $T$ payment equal to the constant $e$ is also marketable).
(b) Show that if the bank account process $B$ is deterministic, then the call option is marketable if and only if the corresponding put option is marketable (Hint: specify the trading strategy which replicates the constant payment $e$).
(c) Show that if the bank account process $B$ is predictable but not deterministic, then it is possible for a call to be marketable even though the corresponding put as well as the claim with time $T$ payment equal to the constant $e$ are not marketable. Do this by considering the model with $T = 2$, $K = 6$, $N = 1$, $\mathscr{P}_1 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6\}\}$, $S_0 = 40$, options expiring at $T = 2$ with $e = 43\frac{48}{109}$, and prices as follows:

| $\omega$ | $B_1(\omega)$ | $S_1(\omega)$ | $B_2(\omega)$ | $S_2(\omega)$ |
|---|---|---|---|---|
| $\omega_1$ | 1.10 | 45 | 1.232 | 55 |
| $\omega_2$ | 1.10 | 45 | 1.232 | 40 |
| $\omega_3$ | 1.10 | 40 | 1.21 | 50 |
| $\omega_4$ | 1.10 | 40 | 1.21 | 35 |
| $\omega_5$ | 1.10 | 35 | 1.188 | 40 |
| $\omega_6$ | 1.10 | 35 | 1.188 | 30 |

**Exercise 4.4** For the model in example 4.4, compute the time 0 price and the replicating portfolio for the following options:

(a) Call with $e = 13$ on the time 2 value of the index $S_1 + S_2$ (hint: we already know the price is 19/18).
(b) Put with $e = 13$ on the time 2 value of the index $S_1 + S_2$.
(c) An option to acquire the maximum of $S_1(2)$ and $S_2(2)$.
(d) A call with $e = 6$ on $\max\{S_1(2), S_2(2)\}$.
(e) An Asian (i.e., averaging) call option with $e = 13$ on the index $S_1 + S_2$.

## 4.2 European Options Under the Binomial Model

The binomial model was introduced in section 3.4. It consists of a single risky security satisfying



$$S_t = S_0 u^{N_t} d^{t-N_t}, \quad t = 1, 2, \ldots, T$$

where $0 < d < 1 < u$ and $N = \{N_t; t = 1, \ldots, T\}$ is a binomial process with parameter $p$, $0 < p < 1$. Assuming the interest rate $r$ is constant with $u > 1 + r$, there exists a martingale measure; it is given by

$$Q(\omega) = q^n (1-q)^{T-n}, \quad q = \frac{1+r-d}{u-d}$$

where $\omega \in \Omega$ is any state corresponding to $n$ 'up moves' and $T - n$ 'down moves' by the risky security.

The probability distribution of $S_t$ under the martingale measure is given by (3.32). Hence if a contingent claim is of the form

$$X = g(S_T)$$

for a specified real-valued function $g$, then the time $t = 0$ value of $X$ is given by

$$V_0 = (1+r)^{-T} E_Q\, g(S_T) \tag{4.4}$$
$$= (1+r)^{-T} \sum_{n=0}^{T} \binom{T}{n} q^n (1-q)^{T-n} g(S_0 u^n d^{T-n})$$

The binomial model is a good illustration of a useful principle: if the underlying securities are Markov chains, then you probably can work out explicitly the probability distribution for the time $T$ values of the securities under the martingale measure, in which case you can compute explicit formulas for the values of contingent claims which are of the form $X = g(S_T)$.

**Example 4.5** For a call option with exercise price $e$ we have

$$g(s) = (s-e)^+ = \begin{cases} s - e, & s \geqslant e \\ 0, & s \leqslant e \end{cases}$$

Hence the time $t = 0$ price is

$$V_0 = (1+r)^{-T} \sum_{n=0}^{T} \binom{T}{n} q^n (1-q)^{T-n} \max\{0,\ S_0 u^n d^{T-n} - e\}$$

Note that

$$S_0 u^n d^{T-n} - e > 0 \Leftrightarrow (u/d)^n > e/(S_0 d^T)$$
$$\Leftrightarrow n \log(u/d) > \log(e/(S_0 d^T))$$
$$\Leftrightarrow n > \frac{\log(e/(S_0 d^T))}{\log(u/d)}$$

We define $\hat{n}$ to be the smallest non-negative integer $n$ such that this strict inequality is satisfied.

Now if $\hat{n} > T$, then $S_0 u^n d^{T-n} \leqslant e$ for all $n \leqslant T$, in which case $V_0 = 0$. On the other hand, if $0 \leqslant \hat{n} \leqslant T$, then there is a chance the call option will finish in the money, in which case $V_0 > 0$. In particular,

$$V_0 = (1+r)^{-T}\sum_{n=0}^{\hat{n}-1}\binom{T}{n}q^n(1-q)^{T-n}(0)$$
$$+ (1+r)^{-T}\sum_{n=\hat{n}}^{T}\binom{T}{n}q^n(1-q)^{T-n}[S_0u^nd^{T-n}-e]$$
$$= \frac{S_0}{(1+r)^T}\sum_{n=\hat{n}}^{T}\binom{T}{n}q^n(1-q)^{T-n}u^nd^{T-n}$$
$$- \frac{e}{(1+r)^T}\sum_{n=\hat{n}}^{T}\binom{T}{n}q^n(1-q)^{T-n}$$
$$= S_0\sum_{n=\hat{n}}^{T}\binom{T}{n}\left[\frac{qu}{1+r}\right]^n\left[\frac{(1-q)d}{1+r}\right]^{T-n}$$
$$- \frac{e}{(1+r)^T}\sum_{n=\hat{n}}^{T}\binom{T}{n}q^n(1-q)^{T-n}$$
$$= S_0\sum_{n=\hat{n}}^{T}\binom{T}{n}\hat{q}^n(1-\hat{q})^{T-n} - \frac{e}{(1+r)^T}\sum_{n=\hat{n}}^{T}\binom{T}{n}q^n(1-q)^{T-n}$$

where $\hat{q} \equiv qu/(1+r)$. It is left to the reader to use elementary algebra to verify that $1-\hat{q} = (1-q)d/(1+r)$ and $0 < \hat{q} < 1$. Hence both terms in the formula for $V_0$ involve the sum of $T-\hat{n}+1$ binomial probabilities.

**Example 4.6** Consider a look-back call option $X = (Y_T - e)^+$, where $Y_T = \max\{S_t : t = 0, 1, \ldots, T\}$, as introduced in section 3.4. Assume $d = u^{-1}$, so that $Y_T$ will take one of the $T+1$ values $S_0, S_0u, \ldots$, or $S_0u^T$. Under the martingale measure, the probability $P\{Y_T > S_0u^i\}$ is given for $i = 1, \ldots, T$ by formula (3.33), only with parameter $q$ in place of $p$. Hence in principle we have the probability distribution of $Y_T$, and so it is straightforward to compute a formula for the time $t=0$ price $(1+r)^{-T}E_Q(Y_T-e)^+$. Similar approaches will work for the look-back put option $X = (e - Y_T)^+$ as well as look-back puts and calls based on the minimum security price level reached before time $T$.

**Example 4.7** *Knockout* options are ones which expire worthless if the price level ever hits a specified level, say $k$. For example, suppose $S_0 < k, e < k, X = (S_T - e)^+$ if the maximum price $Y_T < k$, and $X = 0$ if $Y_T \geqslant k$. The ideas developed in section 3.4 and applied in example 4.6 can be used to value this option.

Suppose $d = u^{-1}$ and $i$ is such that $S_0u^i = k$. As explained in section 3.4, we know, at least in principle, the probability distribution under the martingale measure of $\tau_i$, the first passage time to security price level $k$. Note that $\tau_i$ takes one of the values 1, 2,…, $T$, or, if the security price never reaches the level $k$, infinity. Moreover, we know the conditional probability $Q\{S_T = S_0u^j \mid \tau_i = t\}$ is the same as the conditional probability $Q\{S_{T-t} = S_0u^{j-i} | S_0 = k\}$, where integer $j < i$. In other words, this conditional probability is the same as the probability that the price process which is at level $k$ at time $t$ will find itself exactly $i-j$ price levels lower after the remaining $T-t$ periods. Under the martingale measure, this conditional probability is

$$Q\{S_T = S_0u^j | \tau_i = t\} = \binom{T-t}{n} q^n (1-q)^{T-t-n}, \quad j < i$$

provided $n \equiv (T-t+j-i)/2$ is an integer (here $n$ is the number of 'up moves' during the last $T-t$ periods so that the price moves from level $k$ to level $S_0u^j$). Hence we can compute the joint probability distribution

$$Q\{S_T = S_0u^j, \tau_i = t\} = Q\{S_T = S_0u^j | \tau_i = t\}\ Q\{\tau_i = t\}$$

for $j < i$ and $t = 1, \ldots, T$. Next we easily compute for $j < i$

$$Q\{S_T = S_0u^j, \tau_i = \infty\} = Q\{S_T = S_0u^j\} - \sum_{t=1}^{T} Q\{S_T = S_0u^j, \tau_i = t\}$$

All these probabilities, of course, are under the martingale measure $Q$. Finally, we compute the price $(1+r)^{-T}E_QX$ of the knockout option by using $(1+r)^{-T}\sum_{j<i} Q\{S_T = S_0u^j, \tau_i = \infty\}(S_0u^j - e)^+$.

The option in example 4.7 is also called an *up-and-out* call. You could also have a *down-and-out* call ($k < S_0, k < e$), an *up-and-out* put ($S_0 < k, e < k$), and a *down-and-out* put ($k < S_0, k < e$). With all these options you can compute the time 0 price with a similar approach, that is, by deriving the joint probability distribution under the martingale measure $Q\{S_T = S_0u^j, \tau_i = \infty\}$.

Paired with each of these four knockout options are options that become activated if and only if the price level $k$ is ever reached. For example, an *up-and-in* call ($S_0 < k$, $e < k$) expires worthless if the maximum price remains strictly less than $k$, whereas the expiration value is $(S_T - e)^+$ if the maximum security price during [0, $T$] is greater than or equal to $k$. You also have *down-and-in* calls, *up-and-in* puts, and *down-and-in* puts. All eight of these options are called *barrier* options. Again, you can compute time 0 prices of these four 'in' options by deriving the joint probability distribution $Q\{S_T = S_0u^j, \tau_i < \infty\}$. Or, if you already know the price of the paired knockout option, then you can use the fact that the time 0 price of an ordinary call (or put) is equal to the sum of the prices of the barrier options in a pairing. For example, the price of the up-and-in call option with parameters $k$ and $e$ is equal to the price of the European call option with exercise price $e$ less the price of the up-and-out call option with parameters $k$ and $e$. This is because

the expiration value of the call option is equal to the expiration value of the up-and-in call plus the expiration value of the up-and-out call.

**Exercise 4.5** Consider the European call option under the binomial model with $T = 2$, $e = 1000, u = 1.1$, $d = 0.9$, and $r = 1/100$. Compute the option price as a function of the initial stock price $S_0$ and draw a precise graph of this function.

**Exercise 4.6** Suppose $S_0 = 80$, $T = 3$, $u = 1.5, d = 0.5$, and $r = 0.1$ are the parameters for the binomial model, and consider a call option with exercise price $e = 80$.

(a) Draw a lattice (i.e., recombining network) for the model and label the nodes with the security's price $S_t$.
(b) Compute $q$ and $\hat{q}$.
(c) Label each node with the corresponding option value.
(d) Determine the replicating portfolio.

**Exercise 4.7** Suppose $S_0 = 36$, $T = 2$, $u = 1.5$, $d = 2/3$, and $r = 0$ are the parameters for the binomial model. Compute the prices of the following barrier options by deriving the joint probability distribution $Q\{S_T, \tau\}$, where $\tau$ is the first passage time to the barrier $k$.

(a) Up-and-in call with $k = 54$ and $e = 24$.
(b) Up-and-out call with $k = 54$ and $e = 24$.
(c) Up-and-in put with $k = 54$ and $e = 40$.
(d) Up-and-out put with $k = 54$ and $e = 40$.
(e) Down-and-in call with $k = 24$ and $e = 30$.
(f) Down-and-out call with $k = 24$ and $e = 30$.
(g) Down-and-in put with $k = 24$ and $e = 54$.
(h) Down-and-out put with $k = 24$ and $e = 54$.

## 4.3 American Options

With European options, that is, with contingent claims, the payoff $X$ can only occur at a specified date $T$, the expiration date. American options are similar, only now the payoff can occur at any time $\tau$ on or before the specified expiration date $T$.

As with European options, you should think of an American option as a contract between two parties, a buyer and a seller. Specified as data is a non-negative, adapted stochastic process $Y = \{Y_t; t = 0, 1 \ldots, T\}$. If they make an agreement at time $t$, then at that time the buyer pays the seller an amount $Z_t$ equal to the time $t$ value of this option. The buyer then has the right to exercise this option at any time $\tau$, where $t \leqslant \tau \leqslant T$. If the option is exercised at time $\tau$, then the seller pays $Y_\tau$ to the buyer. An American option can only be exercised once. If it is never exercised, then no payoff occurs. The key problem, of course, is to determine the time $t$ value $Z_t$ of this option, that is, the *value process* $Z = \{Z_t; t = 0, 1, \ldots, T\}$ for this American option $Y$.



**Example 4.8** Setting $Y_t = (S_n(t) - e)^+$ gives rise to an *American call option* with exercise price $e$. Hence the buyer has the right to purchase one unit of security $n$ for the amount $e$ at any time on or before date $T$. Similarly, setting $Y_t = (e - S_n(t))^+$ leads to an *American put* which gives the option buyer the right to sell one unit of the security on or before time $T$ for the price $e$.

If you buy an American option, then you can always postpone the exercise decision until time $T$, so the value $Z_t$ of the American option $Y$ is at least as large as the value $V_t$ of the European option that has payoff $X = Y_T$. In addition, the possibility of being able to collect a desirable payoff at an earlier time tends to make American options more valuable than their European counterparts. Surprisingly, however, there are important situations where the two values coincide.

**(4.5)** Consider an American option $Y = \{Y_t; t = 0, 1, \ldots, T\}$ and the corresponding European option with time $T$ payoff $X = Y_T$. Let $V_t$ denote the time $t$ value of this European option. If $V_t \geqslant Y_t$ for all $t$ and $\omega \in \Omega$, then, for all $t$, $V_t$ is equal to $Z_t$, the time $t$ value of the American option, and it is optimal to wait until time $T$ to exercise.

The reasoning is quite simple: if you bought the American option and $V_t \geqslant Y_t$, then exercising at time $t$ would be foolish, because you can always guarantee yourself a time $t$ payoff of $V_t$. For example, you could turn around and sell the corresponding European option for $V_t$, or you could go short the portfolio which replicates the corresponding European option. Since it is optimal to wait until time $T$ to decide whether to exercise, the values of the two options must be the same.

**Example 4.1 (continued)** Suppose $Y_t = (S_t - 5)^+$, where $S_t$ is the price process in example 4.1. Recall $r = 0$ and $T = 2$. The value process $V_t$ for the corresponding European option $X = \max\{0, S_2 - 5\}$ was derived in section 4.1. Figure 4.2 displays on the information tree for this model the price process, the value process $V$, and the payoff process $Y$. Since $V_t \geqslant Y_t$ for all $t$, $V$ is equal to the value process $Z$ for the American option.

Alternatively, suppose the American option's payoff $Y$ is as displayed in figure 4.3. The time 2 payoffs are the same as in the figure 4.2 example, but now the time 0 and 1 payoffs $Y_t$ are greater (e.g., think of the American call option with $e = 4$, and suppose the stock pays a \$1 dividend between times 1 and 2). Hence the value process $V_t$ for the European option $X = Y_2$ will be the same as in the figure 4.2

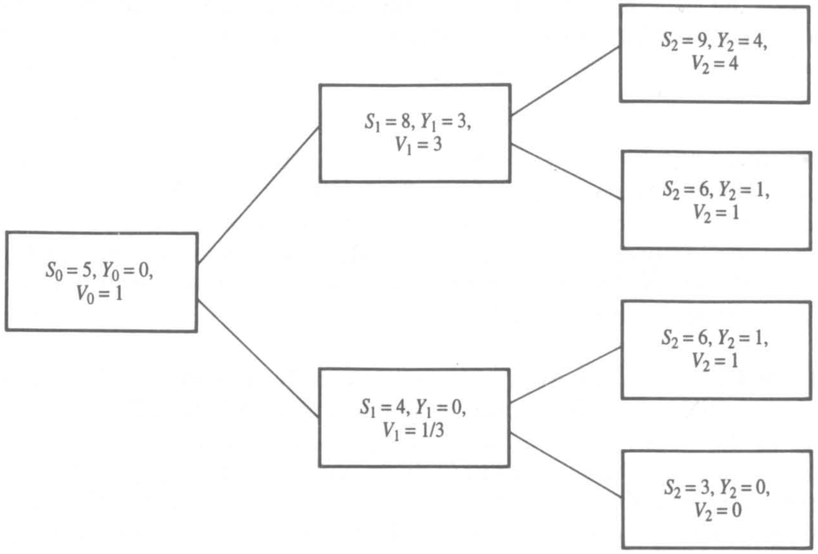


**Figure 4.2** Values of American and European options are the same for example 4.1

example, but now the payoff process will not satisfy $V \geqslant Y$. In particular, $Y_1 > V_1$ when $S_1 = 8$, and now there is no reason to suppose

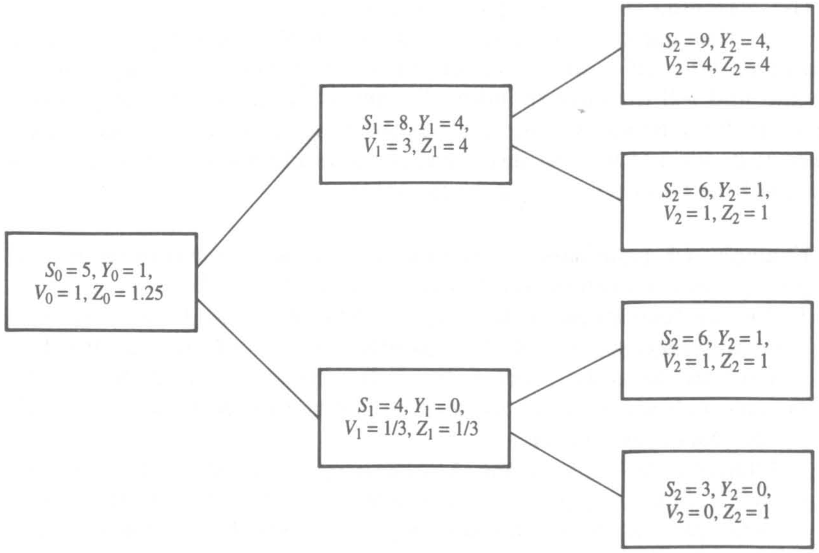


**Figure 4.3** Values of American and European options are different for example 4.1



the American option's value is given by $V$. The value process $Z$ for this American option is displayed in figure 4.3 and will be derived later.

In order to evaluate American options, it is necessary to introduce a new kind of stochastic process. The adapted stochastic process $Z = \{Z_t; t = 0, 1, \ldots, T\}$ is said to be a *supermartingale* (see section 3.3) if

$$E[Z_t \mid \mathscr{F}_s] \leqslant Z_s \qquad \text{all } 0 \leqslant s \leqslant t \leqslant T$$

Thus a supermartingale resembles a martingale, except that the conditional expected future value can be less than, instead of identical to, the current value. All martingales are supermartingales, but not vice versa. We know that the discounted value of a European option is a martingale under the risk neutral probability measure. It turns out that the discounted value of an American option is a supermartingale under the same measure.

It is necessary to introduce a second topic from the theory of stochastic processes. A *stopping time* is a random variable $\tau$ taking values in the set $\{0, 1, \ldots, T, \infty\}$ such that each event of the form $\{\tau = t\}, t \leqslant T$, is an element of the algebra $\mathscr{F}_t$. Thus you can evaluate whether the event $\{\tau = t\}$ occurs simply by examining $\mathscr{F}_t$, the information available at time $t$. For example, for a security with $S_0 = 10$, $\tau_1 \equiv \min\{t : S_t \geqslant 20\}$ is a stopping time, because you learn the event $\{\tau_1 = t\}$ by time $t$. However, the random variable $\tau_2 \equiv \max\{t : S_t \geqslant 20\}$ is not a stopping time, because you may not learn whether $\{\tau_2 = t\}$ until time $T$. Stopping times are allowed to take a value such as $\infty$ in order to provide for the possibility that the event of interest never occurs. For example, $\tau_1 = \infty$ if the event $\{S_t \geqslant 20\}$ does not occur by time $T$.

There are many stopping times associated with our security model's filtration. It is convenient to classify these stopping times by letting $\zeta(s, t)$ denote the set of (random variables which are) stopping times that take finite values in the closed interval $[s,t]$.

The process $Z$ in what follows will turn out to be the value process for the American option $Y$.

**(4.6)** Suppose there exists a risk neutral probability measure $Q$ and define the adapted stochastic process $Z = \{Z_t; t = 0, \ldots, T\}$ by

$$Z_t = \max_{\tau \in \zeta(t,T)} E_Q[Y_\tau B_t / B_\tau | \mathscr{F}_t]. \tag{4.7}$$

Then the process $Z/B$ is the smallest $Q$- supermartingale satisfying

$$Z_t \geqslant Y_t, \quad \textit{all } t, \omega. \tag{4.8}$$

Moreover, the stopping time

$$\tau(t) \equiv \min\{s \geqslant t \ : \ Z_s = Y_s\} \tag{4.9}$$

maximizes the right hand side of (4.7) for $t = 0, 1, \ldots, T$.