

# RISK

*and* FINANCIAL  
MANAGEMENT

Mathematical and  
Computational  
Methods

Charles S. Tapiero

 WILEY

# Risk and Financial Management

# Risk and Financial Management

Mathematical and Computational Methods

**CHARLES TAPIERO**

*ESSEC Business School, Paris, France*



John Wiley & Sons, Ltd

Copyright © 2004 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,  
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): [cs-books@wiley.co.uk](mailto:cs-books@wiley.co.uk)  
Visit our Home Page on [www.wileyeurope.com](http://www.wileyeurope.com) or [www.wiley.com](http://www.wiley.com)

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher.

Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to [permreq@wiley.co.uk](mailto:permreq@wiley.co.uk), or faxed to (+44) 1243 770571.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

### ***Other Wiley Editorial Offices***

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

### ***Library of Congress Cataloging-in-Publication Data***

Tapiero, Charles S.

Risk and financial management : mathematical and computational methods / Charles Tapiero.  
p. cm.

Includes bibliographical references.

ISBN 0-470-84908-8

1. Finance—Mathematical models. 2. Risk management. I. Title.

HG106 .T365 2004

658.15'5'015192—dc22

2003025311

### ***British Library Cataloguing in Publication Data***

A catalogue record for this book is available from the British Library

ISBN 0-470-84908-8

Typeset in 10/12 pt Times by TechBooks, New Delhi, India

Printed and bound in Great Britain by Biddles Ltd, Guildford, Surrey

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

This book is dedicated to:

Daniel  
Dafna  
Oren  
Oscar and  
Bettina

# Contents

Preface

xiii

## Part I: Finance and Risk Management

<b>Chapter 1</b>	<b>Potpourri</b>	<b>03</b>
1.1	Introduction	03
1.2	Theoretical finance and decision making	05
1.3	Insurance and actuarial science	07
1.4	Uncertainty and risk in finance	10
1.4.1	Foreign exchange risk	10
1.4.2	Currency risk	12
1.4.3	Credit risk	12
1.4.4	Other risks	13
1.5	Financial physics	15
	Selected introductory reading	16
<b>Chapter 2</b>	<b>Making Economic Decisions under Uncertainty</b>	<b>19</b>
2.1	Decision makers and rationality	19
2.1.1	The principles of rationality and bounded rationality	20
2.2	Bayes decision making	22
2.2.1	Risk management	23
2.3	Decision criteria	26
2.3.1	The expected value (or Bayes) criterion	26
2.3.2	Principle of (Laplace) insufficient reason	27
2.3.3	The minimax (maximin) criterion	28
2.3.4	The maximax (minimin) criterion	28
2.3.5	The minimax regret or Savage's regret criterion	28
2.4	Decision tables and scenario analysis	31
2.4.1	The opportunity loss table	32
2.5	EMV, EOL, EPPI, EVPI	33
2.5.1	The deterministic analysis	34
2.5.2	The probabilistic analysis	34
	Selected references and readings	38

<b>Chapter 3</b>	<b>Expected Utility</b>	<b>39</b>
3.1	The concept of utility	39
3.1.1	Lotteries and utility functions	40
3.2	Utility and risk behaviour	42
3.2.1	Risk aversion	43
3.2.2	Expected utility bounds	45
3.2.3	Some utility functions	46
3.2.4	Risk sharing	47
3.3	Insurance, risk management and expected utility	48
3.3.1	Insurance and premium payments	48
3.4	Critiques of expected utility theory	51
3.4.1	Bernoulli, Buffon, Cramer and Feller	51
3.4.2	Allais Paradox	52
3.5	Expected utility and finance	53
3.5.1	Traditional valuation	54
3.5.2	Individual investment and consumption	57
3.5.3	Investment and the CAPM	59
3.5.4	Portfolio and utility maximization in practice	61
3.5.5	Capital markets and the CAPM again	63
3.5.6	Stochastic discount factor, assets pricing and the Euler equation	65
3.6	Information asymmetry	67
3.6.1	‘The lemon phenomenon’ or adverse selection	68
3.6.2	‘The moral hazard problem’	69
3.6.3	Examples of moral hazard	70
3.6.4	Signalling and screening	72
3.6.5	The principal–agent problem	73
	References and further reading	75
<b>Chapter 4</b>	<b>Probability and Finance</b>	<b>79</b>
4.1	Introduction	79
4.2	Uncertainty, games of chance and martingales	81
4.3	Uncertainty, random walks and stochastic processes	84
4.3.1	The random walk	84
4.3.2	Properties of stochastic processes	91
4.4	Stochastic calculus	92
4.4.1	Ito’s Lemma	93
4.5	Applications of Ito’s Lemma	94
4.5.1	Applications	94
4.5.2	Time discretization of continuous-time finance models	96
4.5.3	The Girsanov Theorem and martingales*	104
	References and further reading	108
<b>Chapter 5</b>	<b>Derivatives Finance</b>	<b>111</b>
5.1	Equilibrium valuation and rational expectations	111

5.2	Financial instruments	113
5.2.1	Forward and futures contracts	114
5.2.2	Options	116
5.3	Hedging and institutions	119
5.3.1	Hedging and hedge funds	120
5.3.2	Other hedge funds and investment strategies	123
5.3.3	Investor protection rules	125
	References and additional reading	127

## Part II: Mathematical and Computational Finance

<b>Chapter 6</b>	<b>Options and Derivatives Finance Mathematics</b>	<b>131</b>
6.1	Introduction to call options valuation	131
6.1.1	Option valuation and rational expectations	135
6.1.2	Risk-neutral pricing	137
6.1.3	Multiple periods with binomial trees	140
6.2	Forward and futures contracts	141
6.3	Risk-neutral probabilities again	145
6.3.1	Rational expectations and optimal forecasts	146
6.4	The Black–Scholes options formula	147
6.4.1	Options, their sensitivity and hedging parameters	151
6.4.2	Option bounds and put–call parity	152
6.4.3	American put options	154
	References and additional reading	157
<b>Chapter 7</b>	<b>Options and Practice</b>	<b>161</b>
7.1	Introduction	161
7.2	Packaged options	163
7.3	Compound options and stock options	165
7.3.1	Warrants	168
7.3.2	Other options	169
7.4	Options and practice	171
7.4.1	Plain vanilla strategies	172
7.4.2	Covered call strategies: selling a call and a share	176
7.4.3	Put and protective put strategies: buying a put and a stock	177
7.4.4	Spread strategies	178
7.4.5	Straddle and strangle strategies	179
7.4.6	Strip and strap strategies	180
7.4.7	Butterfly and condor spread strategies	181
7.4.8	Dynamic strategies and the Greeks	181
7.5	Stopping time strategies*	184
7.5.1	Stopping time sell and buy strategies	184
7.6	Specific application areas	195

7.7	Option misses	197
	References and additional reading	204
	Appendix: First passage time*	207
<b>Chapter 8</b>	<b>Fixed Income, Bonds and Interest Rates</b>	<b>211</b>
8.1	Bonds and yield curve mathematics	211
8.1.1	The zero-coupon, default-free bond	213
8.1.2	Coupon-bearing bonds	215
8.1.3	Net present values (NPV)	217
8.1.4	Duration and convexity	218
8.2	Bonds and forward rates	222
8.3	Default bonds and risky debt	224
8.4	Rated bonds and default	230
8.4.1	A Markov chain and rating	233
8.4.2	Bond sensitivity to rates – duration	235
8.4.3	Pricing rated bonds and the term structure risk-free rates*	239
8.4.4	Valuation of default-prone rated bonds*	244
8.5	Interest-rate processes, yields and bond valuation*	251
8.5.1	The Vasicek interest-rate model	254
8.5.2	Stochastic volatility interest-rate models	258
8.5.3	Term structure and interest rates	259
8.6	Options on bonds*	260
8.6.1	Convertible bonds	261
8.6.2	Caps, floors, collars and range notes	262
8.6.3	Swaps	262
	References and additional reading	264
	Mathematical appendix	267
	A.1: Term structure and interest rates	267
	A.2: Options on bonds	268
<b>Chapter 9</b>	<b>Incomplete Markets and Stochastic Volatility</b>	<b>271</b>
9.1	Volatility defined	271
9.2	Memory and volatility	273
9.3	Volatility, equilibrium and incomplete markets	275
9.3.1	Incomplete markets	276
9.4	Process variance and volatility	278
9.5	Implicit volatility and the volatility smile	281
9.6	Stochastic volatility models	282
9.6.1	Stochastic volatility binomial models*	282
9.6.2	Continuous-time volatility models	00
9.7	Equilibrium, SDF and the Euler equations*	293
9.8	Selected Topics*	295
9.8.1	The Hull and White model and stochastic volatility	296
9.8.2	Options and jump processes	297

9.9	The range process and volatility	299
	References and additional reading	301
	Appendix: Development for the Hull and White model (1987)*	305
<b>Chapter 10</b>	<b>Value at Risk and Risk Management</b>	<b>309</b>
10.1	Introduction	309
10.2	VaR definitions and applications	311
10.3	VaR statistics	315
10.3.1	The historical VaR approach	315
10.3.2	The analytic variance–covariance approach	315
10.3.3	VaR and extreme statistics	316
10.3.4	Copulae and portfolio <i>VaR</i> measurement	318
10.3.5	Multivariate risk functions and the principle of maximum entropy	320
10.3.6	Monte Carlo simulation and VaR	324
10.4	VaR efficiency	324
10.4.1	VaR and portfolio risk efficiency with normal returns	324
10.4.2	VaR and regret	326
	References and additional reading	327
	<b>Author Index</b>	<b>329</b>
	<b>Subject Index</b>	<b>333</b>

# Preface

Another finance book to teach what market gladiators/traders either know, have no time for or can't be bothered with. Yet another book to be seemingly drowned in the endless collections of books and papers that have swamped the economic literate and illiterate markets ever since options and futures markets grasped our popular consciousness. Economists, mathematically inclined and otherwise, have been largely compensated with Nobel prizes and seven-figures earnings, competing with market gladiators – trading globalization, real and not so real financial assets. Theory and practice have intermingled accumulating a wealth of ideas and procedures, tested and remaining yet to be tested. Martingale, chaos, rational versus adaptive expectations, complete and incomplete markets and whatnot have transformed the language of finance, maintaining their true meaning to the mathematically initiated and eluding the many others who use them nonetheless.

This book seeks to provide therefore, in a readable and perhaps useful manner, the basic elements or economic language of financial risk management, mathematical and computational finance, laying them bare to both students and traders. All great theories are based on simple philosophical concepts, that in some circumstances may not withstand the test of reality. Yet, we adopt them and behave accordingly for they provide a framework, a reference model, inspiring the required confidence that we can rely on even if there is not always something to stand on. An outstanding example might be complete markets and options valuation – which might not be always complete and with an adventuresome valuation of options. Market traders make seemingly risk-free arbitrage profits that are in fact model-dependent. They take positions whose risk and rewards we can only make educated guesses at, and make venturesome and adventuresome decisions in these markets based on facts, fancy and fanciful interpretations of historical patterns and theoretical–technical analyses that seek to decipher things to come.

The motivation to write this book arose from long discussions with a hedge fund manager, my son, on a large number of issues regarding markets behaviour, global patterns and their effects both at the national and individual levels, issues regarding psychological behaviour that are rendering markets less perfect than what we might actually believe. This book is the fruit of our theoretical and practical contrasts and language – the sharp end of theory battling the long and wily practice of the market gladiator, each with our own vocabulary and misunderstandings. Further, too many students in computational finance learn techniques, technical analysis and financial decision making without assessing the dependence of such

analyses on the definition of uncertainty and the meaning of probability. Further, defining ‘uncertainty’ in specific ways, dictates the type of technical analysis and generally the theoretical finance practised. This book was written, both to clarify some of the issues confronting theory and practice and to explain some of the ‘fundamentals, mathematical’ issues that underpin fundamental theory in finance.

Fundamental notions are explained intuitively, calling upon many trading experiences and examples and simple equations-analysis to highlight some of the basic trends of financial decision making and computational finance. In some cases, when mathematics are used extensively, sections are starred or introduced in an appendix, although an intuitive interpretation is maintained within the main body of the text.

To make a trade and thereby reach a decision under uncertainty requires an understanding of the opportunities at hand and especially an appreciation of the underlying sources and causes of change in stocks, interest rates or assets values. The decision to speculate against or for the dollar, to invest in an Australian bond promising a return of five % over 20 years, are risky decisions which, inordinately amplified, may be equivalent to a gladiator’s fight for survival. Each day, tens of thousands of traders, investors and fund managers embark on a gargantuan feast, buying and selling, with the world behind anxiously betting and waiting to see how prices will rise and fall. Each gladiator seeks a weakness, a breach, through which to penetrate and make as much money as possible, before the hordes of followers come and disturb the market’s equilibrium, which an instant earlier seemed unmovable. Size, risk and money combine to make one richer than Croesus one minute and poorer than Job an instant later. Gladiators, too, their swords held high one minute, and history a minute later, have played to the arena. Only, it is today a much bigger arena, the prices much greater and the losses catastrophic for some, unfortunately often at the expense of their spectators.

Unlike in previous times, spectators are thrown into the arena, their money fated with these gladiators who often risk, not their own, but everyone else’s money – the size and scale assuming a dimension that no economy has yet reached.

For some, the traditional theory of decision-making and risk taking has fared badly in practice, providing a substitute for reality rather than dealing with it. Further, the difficulty of problems has augmented with the involvement of many sources of information, of time and unfolding events, of information asymmetries and markets that do not always behave competitively, etc. These situations tend to distort the approaches and the techniques that have been applied successfully but to conventional problems. For this reason, there is today a great deal of interest in understanding how traders and financial decision makers reach decisions and not only what decisions they ought to reach. In other words, to make better decisions, it is essential to deal with problems in a manner that reflects reality and not only theory that in its essence, always deals with structured problems based on specific assumptions – often violated. These assumptions are sometimes realistic; but sometimes they are not. Using specific problems I shall try to explain approaches applied in complex financial decision processes – mixing practice and theory. The approach we follow is at times mildly quantitative, even though much of the new approach to finance is mathematical and computational and requires an

extensive mathematical proficiency. For this reason, I shall assume familiarity with basic notions in calculus as well as in probability and statistics, making the book accessible to typical economics and business and maths students as well as to practitioners, traders and financial managers who are familiar with the basic financial terminology.

The substance of the book in various forms has been delivered in several institutions, including the MASTER of Finance at ESSEC in France, in Risk Management courses at ESSEC and at Bar Ilan University, as well as in Mathematical Finance courses at Bar Ilan University Department of Mathematics and Computer Science. In addition, the Montreal Institute of Financial Mathematics and the Department of Finance at Concordia University have provided a testing ground as have a large number of lectures delivered in a workshop for MSc students in Finance and in a PhD course for Finance students in the Montreal consortium for PhD studies in Mathematical Finance in the Montreal area. Throughout these courses, it became evident that there is a great deal of excitement in using the language of mathematical finance but there is often a misunderstanding of the concepts and the techniques they require for their proper application. This is particularly the case for MBA students who also thrive on the application of these tools. The book seeks to answer some of these questions and problems by providing as much as possible an interface between theory and practice and between mathematics and finance. Finally, the book was written with the support of a number of institutions with which I have been involved these last few years, including essentially ESSEC of France, the Montreal Institute of Financial Mathematics, the Department of Finance of Concordia University, the Department of Mathematics of Bar Ilan University and the Israel Port Authority (Economic Research Division). In addition, a number of faculty and students have greatly helped through their comments and suggestions. These have included, Elias Shiu at the University of Iowa, Lorne Switzer, Meir Amikam, Alain Bensoussan, Avi Lioui and Sebastien Galy, as well as my students Bernardo Dominguez, Pierre Bour, Cedric Lespiau, Hong Zhang, Philippe Pages and Yoav Adler. Their help is gratefully acknowledged.

PART I

# Finance and Risk Management

## CHAPTER 1

# Potpourri

### 1.1 INTRODUCTION

Will a stock price increase or decrease? Would the Fed increase interest rates, leave them unchanged or decrease them? Can the budget to be presented in Transylvania's parliament affect the country's current inflation rate? These and so many other questions are reflections of our lack of knowledge and its effects on financial markets performance. In this environment, uncertainty regarding future events and their consequences must be assessed, predictions made and decisions taken. Our ability to improve forecasts and reach consistently good decisions can therefore be very profitable. To a large extent, this is one of the essential preoccupations of finance, financial data analysis and theory-building. Pricing financial assets, predicting the stock market, speculating to make money and hedging financial risks to avoid losses summarizes some of these activities. Predictions, for example, are reached in several ways such as:

- ‘Theorizing’, providing a structured approach to modelling, as is the case in financial theory and generally called fundamental theory. In this case, economic and financial theories are combined to generate a body of knowledge regarding trades and financial behaviour that make it possible to price financial assets.
- Financial data analysis using statistical methodologies has grown into a field called financial statistical data analysis for the purposes of modelling, testing theories and technical analysis.
- Modelling using metaphors (such as those borrowed from physics and other areas of related interest) or simply constructing model equations that are fitted one way or another to available data.
- Data analysis, for the purpose of looking into data to determine patterns or relationships that were hitherto unseen. Computer techniques, such as neural networks, data mining and the like, are used for such purposes and thereby make more money. In these, as well as in the other cases, the ‘proof of the pudding is in the eating’. In other words, it is by making money, or at least making

it possible for others to make money, that theories, models and techniques are validated.

- Prophecies we cannot explain but sometimes are true.

Throughout these ‘forecasting approaches and issues’ financial managers deal practically with uncertainty, defining it, structuring it and modelling its causes, explainable and unexplainable, for the purpose of assessing their effects on financial performance. This is far from trivial. First, many theories, both financial and statistical, depend largely on how we represent and model uncertainty. Dealing with uncertainty is also of the utmost importance, reflecting individual preferences and behaviours and attitudes towards risk. Decision Making Under Uncertainty (DMUU) is in fact an extensive body of approaches and knowledge that attempts to provide systematically and rationally an approach to reaching decisions in such an environment. Issues such as ‘rationality’, ‘bounded rationality’ etc., as we will present subsequently, have an effect on both the approach we use and the techniques we apply to resolve the fundamental and practical problems that finance is assumed to address. In a simplistic manner, uncertainty is characterized by probabilities. Adverse consequences denote the risk for which decisions must be taken to properly balance the potential payoffs and the risks implied by decisions – trades, investments, the exercise of options etc. Of course, the more ambiguous, the less structured and the more uncertain the situations, the harder it is to take such decisions. Further, the information needed to make decisions is often not readily available and consequences cannot be predicted. Risks are then hard to determine. For example, for a corporate finance manager, the decision may be to issue or not to issue a new bond. An insurance firm may or may not confer a certain insurance contract. A Central Bank economist may recommend reducing the borrowing interest rate, leaving it unchanged or increasing it, depending on multiple economic indicators he may have at his disposal. These, and many other issues, involve uncertainty. Whatever the action taken, its consequences may be uncertain. Further, not all traders who are equally equipped with the same tools, education and background will reach the same decision (of course, when they differ, the scope of decisions reached may be that much broader). Some are well informed, some are not, some believe they are well informed, but mostly, all traders may have various degrees of intuition, introspection and understanding, which is specific yet not quantifiable. A historical perspective of events may be useful to some and useless to others in predicting the future. Quantitative training may have the same effect, enriching some and confusing others. While in theory we seek to eliminate some of the uncertainty by better theorizing, in practice uncertainty wipes out those traders who reach the wrong conclusions and the wrong decisions. In this sense, no one method dominates another: all are important. A political and historical appreciation of events, an ability to compute, an understanding of economic laws and fundamental finance theory, use of statistics and computers to augment one’s ability in predicting and making decisions under uncertainty are only part of the tool-kit needed to venture into trading speculation and into financial risk management.

## 1.2 THEORETICAL FINANCE AND DECISION-MAKING

Financial decision making seeks to make money by using a broad set of economic and theoretical concepts and techniques based on rational procedures, in a consistent manner and based on something more than intuition and personal subjective judgement (which are nonetheless important in any practical situation). Generally, it also seeks to devise approaches that may account for departures from such rationality. Behavioural and psychological reasons, the violation of traditional assumptions regarding competition and market forces and exchange combine to alter the basic assumptions of theoretical economics and finance.

Finance and financial instruments currently available through brokers, mutual funds, financial institutions, commodity and stock markets etc. are motivated by three essential problems:

- Pricing the multiplicity of claims, accounting for risks and dealing with the negative effects of uncertainty or risk (that can be completely unpredictable, or partly or wholly predictable)
- Explaining, and accounting for investors' behaviour. To counteract the effects of regulation and taxes by firms and individual investors (who use a wide variety of financial instruments to bypass regulations and increase the amount of money investors can make).
- Providing a rational framework for individuals' and firms' decision making and to suit investors' needs in terms of the risks they are willing to assume and pay for. For this purpose, extensive use is made of DMUU and the construction of computational tools that can provide 'answers' to well formulated, but difficult, problems.

These instruments deal with the uncertainty and the risks they imply in many different ways. Some instruments merely transfer risk from one period to another and in this sense they reckon with the *time phasing of events* to reckon with. One of the more important aspects of such instruments is to supply '*immediacy*', i.e. the ability not to wait for a payment for example (whereby, some seller will assume the risk and the cost of time in waiting for that payment). Other instruments provide a '*spatial diversification*', in other words, the distribution of risks across a number of independent (or almost independent) risks. For example, buying several types of investment that are less than perfectly correlated, maintaining *liquidity* etc. By liquidity, we mean the cost to instantly convert an asset into cash at its fair price. This liquidity is affected both by the existence of a market (in other words, buyers and sellers) and by the cost of transactions associated with the conversion of the asset into cash. As a result, risks pervading finance and financial risk management are varied; some of them are outlined in greater detail below.

Risk in finance results from the consequences of undesirable outcomes and their implications for individual investors or firms. A definition of risk involves their probability, individual and collective and consequences effects. These are relevant to a broad number of fields as well, each providing an approach to the

measurement and the valuation of risk which *is motivated by their needs and by the set of questions they must respond to and deal with*. For these reasons, the problems of finance often transcend finance and are applicable to the broad areas of economics and decision-making. Financial economics seeks to provide approaches and answers to deal with these problems. The growth of theoretical finance in recent decades is a true testament to the important contribution that financial theory has made to our daily life. Concepts such as financial markets, arbitrage, risk-neutral probabilities, Black–Scholes option valuation, volatility, smile and many other terms and names are associated with a maturing profession that has transcended the basic traditional approaches of making decisions under uncertainty. By the same token, hedging which is an important part of the practice finance is the process of eliminating risks in a particular portfolio through a trade or a series of trades, or contractual agreements. Hedging relates also to the valuation-pricing of derivatives products. Here, a portfolio is constructed (the hedging portfolio) that eliminates all the risks introduced by the derivative security being analyzed in order to replicate a return pattern identical to that of the derivative security. At this point, from the investor's point of view, the two alternatives – the hedging portfolio and the derivative security – are indistinguishable and therefore have the same value. In practice too, speculating to make money can hardly be conceived without hedging to avoid losses.

The traditional theory of decision making under uncertainty, integrating statistics and the risk behaviour of decision makers has evolved in several phases starting in the early nineteenth century. At its beginning, it was concerned with collecting data to provide a foundation for *experimentation and sampling theory*. These were the times when surveys and counting populations of all sorts began. Subsequently, statisticians such as Karl Pearson and R. A. Fisher studied and set up the foundations of *statistical data analysis*, consisting of the assessment of the reliability and the accuracy of data which, to this day, seeks to represent large quantities of information (as given explicitly in data) in an aggregated and summarized fashion, such as probability distributions and moments (mean, variance etc.) and states how accurate they are. Insurance managers and firms, for example, spend much effort in collecting such data to estimate mean claims by insured clients and the propensity of certain insured categories to claim, and to predict future weather conditions in order to determine an appropriate insurance premium to charge. Today, financial data analysis is equally concerned with these problems, bringing sophisticated modelling and estimation techniques (such as linear regression, ARCH and GARCH techniques which we shall discuss subsequently) to bear on the application of financial analysis.

The next step, expounded and developed primarily by R. A. Fisher in the 1920s, went one step further with *planning experiments* that can provide effective information. The issue at hand was then to plan the experiments generating the information that can be analysed statistically and on the basis of which a decision could, justifiably, be reached. This important phase was used first in testing the agricultural yield under controlled conditions (to select the best way to grow plants, for example). It yielded a number of important lessons, namely that the

procedure (statistical or not) used to collect data is intimately related to the kind of relationships we seek to evaluate. A third phase, expanded dramatically in the 1930s and the 1940s consisted in the construction of mathematical models that sought to bridge the gap between the process of data collection and the need of such data for specific purposes such as predicting and decision making. Linear regression techniques, used extensively in econometrics, are an important example. Classical models encountered in finance, such as models of stock market prices, currency fluctuations, interest rate forecasts and investment analysis models, cash management, reliability and other models, are outstanding examples.

In the 1950s and the 1960s the (Bayes) theory of *decision making under uncertainty* took hold. In important publications, Raiffa, Luce, Schlaiffer and many others provided a unified framework for integrating problems relating to data collection, experimentation, model building and decision making. The theory was intimately related to typical economic, finance and industrial, business and other problems. Issues such as the value of information, how to collect it, how much to pay for it, the weight of intuition and subjective judgement (as often used by behavioural economists, psychologists etc.) became relevant and integrated into the theory. Their practical importance cannot be understated for they provide a framework for reaching decisions under complex situations and uncertainty. Today, theories of decision making are an ever-expanding field with many articles, books, experiments and theories competing to provide another view and in some cases another vision of uncertainty, how to model it, how to represent certain facets of the economic and financial process and how to reach decisions under uncertainty. The DMUU approach, however, presumes that uncertainty is specified in terms of probabilities, albeit learned adaptively, as evidence accrues for one or the other event. It is only recently, in the last two decades, that theoretical and economic analyses have provided in some cases theories and techniques that provide an estimate of these probabilities. In other words, while in the traditional approach to DMUU uncertainty is exogenous, facets of modern and theoretical finance have helped 'endogenize' uncertainty, i.e. explain uncertain behaviours and events by the predictive market forces and preferences of traders. To a large extent, the contrasting finance fundamental theory and traditional techniques applied to reach decisions under uncertainty diverge in their attempts to represent and explain the 'making of uncertainty'. This is an important issue to appreciate and one to which we shall return subsequently when basic notions of fundamental theory including rational expectations and option pricing are addressed.

Today, DMUU is economics, finance, insurance and risk motivated. There are a number of areas of special interest we shall briefly discuss to better appreciate the transformations of finance, insurance and risk in general.

### 1.3 INSURANCE AND ACTUARIAL SCIENCE

Actuarial science is in effect one of the first applications of probability theory and statistics to risk analysis. Tetens and Barrois, already in 1786 and 1834

respectively, were attempting to characterize the 'risk' of life annuities and fire insurance and on that basis establish a foundation for present-day insurance. Earlier, the Gambling Act of 1774 in England (King George III) laid the foundation for life insurance. It is, however, to Lundberg in 1909, and to a group of Scandinavian actuaries (Borch, 1968; Cramer, 1955) that we owe much of the current mathematical theory of insurance. In particular, Lundberg provided the foundation for collective risk theory. Terms such as 'premium payments' required from the insured, 'wealth' or the 'firm liquidity' and 'claims' were then defined. In its simplest form, actuarial science establishes exchange terms between the insured, who pays the premium that allows him to claim a certain amount from the firm (in case of an accident), and the insurer, the provider of insurance who receives the premiums and invests and manages the moneys of many insured. The insurance terms are reflected in the 'insurance contract' which provides legally the 'conditional right to claim'. Much of the insurance literature has concentrated on the definition of the rules to be used in order to establish the terms of such a contract in a just and efficient manner. In this sense, 'premium principles' and a wide range of operational rules worked out by the actuarial and insurance profession have been devised. Currently, insurance is gradually being transformed to be much more in tune with market valuation of insurable contracts and financial instruments are being devised for this purpose. The problems of insurance are, of course, extremely complex, with philosophical and social undertones, seeking to reconcile individual with collective risk and individual and collective choices and interests through the use of the market mechanism and concepts of fairness and equity. In its proper time setting (recognizing that insurance contracts express the insured attitudes towards time and uncertainty, in which insurance is used to substitute certain for uncertain payments at different times), this problem is of course, conceptually and quantitatively much more complicated. For this reason, the quantitative approach to insurance, as is the case with most financial problems, is necessarily a simplification of the fundamental issues that insurance deals with.

Risk is managed in several ways including: '*pricing insurance, controls, risk sharing and bonus-malus*'. Bonus-malus provides an incentive not to claim when a risk materializes or at least seeks to influence insured behaviour to take greater care and thereby prevent risks from materializing. In some cases, it is used to discourage nuisance claims. There are numerous approaches to applying each of these tools in insurance. Of course, in practice, these tools are applied jointly, providing a capacity to customize insurance contracts and at the same time assuming a profit for the insurance firm.

In insurance and finance (among others) we will have to deal as well with special problems, often encountered in practical situations but difficult to analyse using statistical and analytical techniques. These essentially include *dependencies, rare events* and *man-made risks*. In insurance, correlated risks are costlier to assume while insuring rare and extremely costly events is difficult to assess. Earthquake and tornado insurance are such cases. Although, they occur, they do so with small probabilities. Their occurrence is extremely costly for the insurer,

however. For this reason, insurers seek the participation of governments for such insurance, study the environment and the patterns in weather changes and turn to extensive risk sharing schemes (such as reinsurance with other insurance firms and on a global scale). Dependencies can also be induced internally (endogenously generated risks). For example, when trading agents follow each other's action they may lead to the rise and fall of an action on the stock market. In this sense, 'behavioural correlations' can induce cyclical economic trends and therefore greater market variability and market risk. Man-made induced risks, such as terrorists' acts of small and unthinkable dimensions, also provide a formidable challenge to insurance companies. John Kay (in an article in the *Financial Times*, 2001) for example states:

The insurance industry is well equipped to deal with natural disasters in the developed world: the hurricanes that regularly hit the south-east United States; the earthquakes that are bound to rock Japan and California from time to time. Everyone understands the nature of these risks and their potential consequences. But we are ignorant of exactly when and where they will materialize. For risks such as these, you can write an insurance policy and assess a premium.

But the three largest disasters for insurers in the past 20 years have been man-made, not natural. The human cost of asbestos was greater even than that of the destruction of the World Trade Center. The deluge of asbestos-related claims was the largest factor in bringing the Lloyd's insurance market to its knees.

By the same token, the debacle following the deregulation of Savings and Loans in the USA in the 1960s led to massive opportunistic behaviours resulting in huge losses for individuals and insurance firms. These disasters have almost uniformly involved government interventions and in some cases bail-outs (as was the case with airlines in the aftermath of the September 11th attack on the World Trade Center). Thus, risk in insurance and finance involves a broad range of situations, sources of uncertainty and a broad variety of tools that may be applied when disasters strike. There are special situations in insurance that may be difficult to assess from a strictly financial point of view, however, as in the case of man-made risks. For example, environmental risks have special characteristics that are affecting our approach to risk analysis:

- Rare events: Relating to very large disasters with very small probabilities that may be difficult to assess, predict and price.
- Spillover effects: Having behavioural effects on risk sharing and fairness since persons causing risks may not be the sole victims. Further, effects may be felt over long periods of time.
- International dimensions: having power and political overtones.

For these reasons, some of the questions raised in conjunction with environmental risk that are of acute interest today are numerous, including among others:

- Who pays for it?
- What prevention if at all?
- Who is responsible if at all?

By the same token, the future of genetic testing promises to reveal information about individuals that, hitherto has been unknown, and thereby to change the whole traditional approach to insurance. In particular, randomness, an essential facet of the insurance business, will be removed and insurance contracts could/would be tailored to individuals' profiles. The problems that may arise subsequent to genetic testing are tremendous. They involve problems arising over the power and information asymmetries between the parties to contracts. Explicitly, this may involve, on the one hand, moral hazard (we shall elaborate subsequently) and, on the other, adverse selection (which will see later as well) affecting the potential future/non-future of the insurance business and the cost of insurance to be borne by individuals.

## 1.4 UNCERTAINTY AND RISK IN FINANCE

Uncertainty and risk are everywhere in finance. As stated above, they result from consequences that may have adverse economic effects. Here are a few financial risks.

### 1.4.1 Foreign exchange risk

*Foreign exchange risk* measures the risk associated with unexpected variations in exchange rates. It consists of two elements: an internal element which depends on the flow of funds associated with foreign exchange, investments and so on, and an external element which is independent of a firm's operations (for example, a variation in the exchange rates of a country).

Foreign exchange risk management has focused essentially on short-term decisions involving accounting exposure components of a firm's working capital. For instance, consider the case of captive insurance companies that diversify their portfolio of underwriting activities by reinsuring a 'layer' of foreign risk. In this case, the magnitude of the transaction exposure is clearly uncertain, compounding the exchange and exposure risks. Bidding on foreign projects or acquisitions of foreign companies will similarly entail exposures whose magnitudes can be characterized at best subjectively. Explicitly, in big-ticket export transactions or large-scale construction projects, the exporter or contractor will first submit a bid  $B(T)$  of say 100 million which is denominated in \$US (a foreign currency from the point of view of the decision maker) and which, if accepted, would give rise to a transaction exposure (asset or liability) maturing at a point in time  $T$ , say 2 years ahead. The bid will in turn be accepted or rejected at time  $t$ , say 6 months ahead ( $0 < t < T$ ), resulting in the transaction exposure which is uncertain until the resolution (time) standing at the full amount  $B(T)$  if the bid is accepted, or

being cancelled if the bid is rejected. Effective management of such uncertain exposures will require the existence of a futures market for foreign exchange allowing contracts to be entered into or cancelled at any time  $t$  over the bidding uncertainty resolution horizon  $0 < t < T$ . The case of foreign acquisition is a special case of the above more general problem with uncertainty resolution being arbitrarily set at  $t = T$ . Problems in long-term foreign exchange risk management – that is, long-term debt financing and debt refunding – in a multi-currency world, although very important, is not always understood and hedged. As global corporations expand operations abroad, foreign currency-denominated debt instruments become an integral part of the opportunities of financing options. One may argue that in a multi-currency world of efficient markets, the selection of the optimal borrowing source should be a matter of indifference, since nominal interest rates reflect inflation rate expectations, which, in turn, determine the pattern of the future spot exchange rate adjustment path. However, heterogeneous corporate tax rates among different national jurisdictions, asymmetrical capital tax treatment, exchange gains and losses, non-random central bank intervention in exchange markets and an ever-spreading web of exchange controls render the hypothesis of market efficiency of dubious operational value in the selection process of the least-cost financing option. How then, should foreign debt financing and refinancing decisions be made, since nominal interest rates can be misleading for decision-making purposes? Thus, a managerial framework is required, allowing the evaluation of the uncertain cost of foreign capital debt financing as a function of the ‘volatility’ (risk) of the currency denomination, the maturity of the debt instrument, the exposed exchange rate appreciation/depreciation and the level of risk aversion of the firm.

To do so, it will be useful to distinguish two sources of risk: internal and external. Internal risk depends on a firm’s operations and thus that depends on the exchange rate while external risk is independent of a firm’s operations (such as a devaluation or the usual variations in exchange rates). These risks are then expressed in terms of:

- Transaction risk, associated with the flow of funds in the firm
- Translation risk, associated with in-process, present and future transactions.
- Competition risk, associated with the firm’s competitive posture following a change in exchange rates.

The actors in a foreign exchange (risk) market are numerous and must be considered as well. These include the firms that import and export, and the intermediaries (such as banks), or traders. Traders behave just as market makers do. At any instant, they propose to buy and sell for a price. Brokers are intermediaries that centralize buy and sell orders and act on behalf of their clients, taking the best offers they can get. Over all, foreign exchange markets are competitive and can reach equilibrium. If this were not the case, then some traders could engage in arbitrage, as we shall discuss later on. This means that some traders will be able to make money without risk and without investing any money.

### 1.4.2 Currency risk

*Currency risk* is associated with variations in currency markets and exchange rates. A currency is not risky because its depreciation is likely. If it were to depreciate for sure and there were to be no uncertainty as to its magnitude and timing—there would not be any risk at all. As a result, a weak currency can be less risky than a strong currency. Thus, the risk associated with a currency is related to its randomness. The problems thus faced by financial analysts consist of defining a reasonable measure of exposure to currency risk and managing it. There may be several criteria in defining such an exposure. First, it ought to be denominated in terms of the relevant amount of currency being considered. Second, it should be a characteristic of any asset or liability, physical or financial, that a given investor might own or owe, defined from the investor's viewpoint. And finally, it ought to be practical. Currency risks are usually associated with macroeconomic variables (such as the trade gap, political stability, fiscal and monetary policy, interest rate differentials, inflation, leadership, etc.) and are therefore topics of considerable political and economic analysis as well as speculation. Further, because of the size of currency markets, speculative positions may be taken by traders leading to substantial profits associated with very small movements in currency values. On a more mundane level, corporate finance managers operating in one country may hedge the value of their contracts and profits in another foreign denominated currency by assuming financial contracts that help to relieve some of the risks associated with currency (relative or absolute) movements and shifts.

### 1.4.3 Credit risk

*Credit risk* covers risks due to upgrading or downgrading a borrower's creditworthiness. There are many definitions of credit risk, however, which depend on the potential sources of the risk, who the client may be and who uses it. Banks in particular are devoting a considerable amount of time and thoughts to defining and managing credit risk. There are basically two sources of uncertainty in credit risk: default by a party to a financial contract and a change in the present value (PV) of future cash flows (which results from changes in financial market conditions, changes in the economic environment, interest rates etc.). For example, this can take the form of money lent that is not returned. Credit risk considerations underlie capital adequacy requirements (CAR) regulations that are required by financial institutions. Similarly, credit terms defining financial borrowing and lending transactions are sensitive to credit risk. To protect themselves, firms and individuals turn to rating agencies such as Standard & Poors, Moody's or others (such as Fitch Investor Service, Nippon Investor Service, Duff & Phelps, Thomson Bank Watch etc.) to obtain an assessment of the risks of bonds, stocks and financial papers they may acquire. Furthermore, even after a careful reading of these ratings, investors, banks and financial institutions proceed to reduce these risks by risk management tools. The number of such tools is of course very large. For

example, limiting the level of obligation, seeking collateral, netting, recouping, insurance, syndication, securitization, diversification, swaps and so on are some of the tools a financial service firm or bank might use.

An exposure to credit risk can occur from several sources. These include an exposure to derivatives products (such as options, as we shall soon define) in exposures to the replacement cost (or potential increases in future replacement costs) due to default arising from market adverse conditions and changes. Problems of credit risk have impacted financial markets and global deflationary forces. 'Wild money' borrowed by hedge funds faster than it can be reimbursed to banks has created a credit crunch. Regulatory distortions are also a persistent theme over time. Over-regulation may hamper economic activity. The creation of wealth, while 'under-regulation' (in particular in emerging markets with cartels and few economic firms managing the economy) can lead to speculative markets and financial distortions. The economic profession has been marred with such problems. For example:

One of today's follies, says a leading banker, is that the Basle capital adequacy regime provides greater incentives for banks to lend to unregulated hedge funds than to such companies as IBM. The lack of transparency among hedge funds may then disguise the bank's ultimate exposure to riskier markets. Another problem with the Basle regime is that it forces banks to reinforce the economic cycle – on the way down as well as up. During a recovery, the expansion of bank profits and capital inevitably spurs higher lending, while capital shrinkage in the downturn causes credit to contract when it is most needed to business. (*Financial Times*, 20 October 1998, p. 17)

Some banks cannot meet international standard CARs. For example, Daiwa Bank, one of Japan's largest commercial banks, is withdrawing from all overseas business partly to avoid having to meet international capital adequacy standards. For Daiwa, as well as other Japanese banks, capital bases have been eroded by growing pressure on them to write off their bad loans and by the falling value of shares they hold in other companies, however, undermining their ability to meet these capital adequacy standards.

To address these difficulties the Chicago Mercantile Exchange, one of the two US futures exchanges, launched a new bankruptcy index contract (for credit default) working on the principle that there is a strong correlation between credit charge-off rates and the level of bankruptcy filings. Such a contract is targeted at players in the consumer credit markets – from credit card companies to holders of car loans and big department store groups. The data for such an index will be based on bankruptcy court data.

#### **1.4.4 Other risks**

There are other risks of course, some of which are defined below while others will be defined, explained and managed as we move along to define and use the tools of risk and computational finance management.

*Market risk* is associated with movements in market indices. It can be due to a stock price change, to unpredictable interest rate variations or to market liquidity, for example.

*Shape risk* is applicable to fixed income markets and is caused by non-parallel shifts of interest rates on straight, default-free securities (i.e. shifts in the term structure of interest rates). In general, rates risks are associated with the set of relevant flows of a firm that depend on the fluctuations of interest rates. The debt of a firm, the credit it has, indexed obligations and so on, are a few examples.

*Volatility risk* is associated with variations in second-order moments (such as process variance). It reduces our ability to predict the future and can induce preventive actions by investors to reduce this risk, while at the same time leading others to speculate wildly. Volatility risk is therefore an important factor in the decisions of speculators and investors. Volatility risk is an increasingly important risk to assess and value, owing to the growth of volatility in stocks, currency and other markets.

*Sector risk* stems from events affecting the performance of a group of securities as a whole. Whether sectors are defined by geographical area, technological specialization or market activity type, they are topics of specialized research. Analysts seek to gain a better understanding of the sector's sources of uncertainty and their relationship to other sectors.

*Liquidity risk* is associated with possibilities that the bid-ask spreads on security transactions may change. As a result, it may be impossible to buy or sell an asset in normal market conditions in one period or over a number of periods of time. For example, a demand for an asset at one time (a house, a stock) may at one time be oversubscribed such that no supply may meet the demand. While a liquidity risk may eventually be overtaken, the lags in price adjustments, the process at hand to meet demands, may create a state of temporary (or not so temporary) shortage.

*Inflation risk*: inflation arises when prices increase. It occurs for a large number of reasons. For example, agents, traders, consumers, sellers etc. may disagree on the value of products and services they seek to buy (or sell) thereby leading to increasing prices. Further, the separation of real assets and financial markets can induce adjustment problems that can also contribute to and motivate inflation. In this sense, a clear distinction ought to be made between financial inflation (reflected in a nominal price growth) and real inflation, based on the real terms value of price growth. If there were no inflation, discounting could be constant (i.e. expressed by fixed interest rates rather than time-varying and potentially random) since it could presume that future prices would be sustained at their current level. In this case, discounting would only reflect the time value of money and not the predictable (and uncertain) variations of prices. In inflationary states, discounting can become nonstationary (and stochastic), leading to important and substantial problems in modelling, understanding how prices change and evolve over time.

Importantly inflation affects economic, financial and insurance related issues and problems. In the insurance industry, for example, premiums and benefits

calculations induced by real as well as nominal price variations, i.e. inflation, are difficult to determine. These variations in prices alter over time the valuation of premiums in insurance contracts introducing a risk due to a lack of precise knowledge about economic activity and price level changes. At the same time, changes in the nominal value of claims distributions (by insurance contract holders), increased costs of living and lags between claims and payment render insurance even more risky. For example, should a negotiated insurance contract include inflation-sensitive clauses? If not, what would the implications be in terms of consumer protection, the time spans of negotiated contracts and, of course, the policy premium? In this simple case, a policyholder will gradually face declining payments but also a declining protection. In case of high inflation, it is expected that the policyholder will seek a renegotiation of his contract (and thereby increased costs for the insurer and the insured). The insurance firm, however, will obtain an unstable stream of payments (in real terms) and a very high cost of operation due to the required contract renegotiation. Unless policyholders are extremely myopic, they would seek some added form of protection to compensate on the one hand for price levels changes and for the uncertainty in these prices on the other. In other words, policyholders will demand, and firms will supply, inflation-sensitive policies. Thus, inflation clearly raises issues and problems that are important for both the insurer and the insured. For this reason, protection from inflation risk, which is the loss at a given time, given an uncertain variation of prices, may be needed. Since this is not a 'loss' per se, but an uncertainty regarding the price, inflation-adjusted loss valuation has to be measured correctly. Furthermore, given an inflation risk definition, the apportioning of this risk between the policyholder and the firm is also required, demanding an understanding of risk attitudes and behaviours of insured and insurer alike. Then, questions such as: who will pay for the inflation risk? how? (i.e. what will be the insurance policy which accounts expressly for inflation) and how much? These issues require that insurance be viewed in its inter-temporal setting rather than its static actuarial approach.

To clarify these issues, consider whether an insurance firm should *a priori* absorb the inflation risk pass it on to policyholders by an increased load factor (premium) or follow a posterior procedure where policyholders increase payments as a function of the published inflation rate, cost of living indices or even the value of a given currency. These are questions that require careful evaluation.

## 1.5 FINANCIAL PHYSICS

Recently, domains such Artificial Intelligence, Data Mining and Computational Tools, as well as the application of constructs and themes reminiscent of financial problems, have become fashionable. In particular, a physics-like approach has been devised to deal with selected financial problems (in particular with option valuation, volatility smile and so on). The intent of physical models is to explain (and thereby forecast) phenomena that are not explained by the fundamental theory. For example, trading activity bursts, bubbles and long and short cycles, as

well as long-run memory, that are poorly explained or predicted by fundamental theory and traditional models are typical applications. The physics approach is essentially a modelling approach, using metaphors and processes/equations used in physics and finding their parallel in economics and finance. For example, an individual consumer might be thought to be an atom moving in a medium/environment which might correspond in economics to a market. The medium results from an infinite number of atoms acting/interacting, while the market results from an infinite number of consumers consuming and trading among themselves. Of course, these metaphors are quite problematic, modelling simplifications, needed to render intractable situations tractable and to allow aggregation of the many atoms (consumers) into a whole medium (market). There are of course many techniques to reach such aggregation. For example, the use of Brownian motion (to represent the uncertainty resulting from many individual effects, individually intractable), originating in Bachelier's early studies in 1905, conveniently uses the Central Limit Theorem in statistics to aggregate events presumed independent. However, this 'seeming normality', resulting from the aggregation of many independent events, is violated in many cases, as has been shown in many financial data analyses. For example, data correlation (which cannot be modelled or explained easily), distributed (stochastic) volatility and the effects of long-run memory not accounted for by traditional modelling techniques, etc. are such cases. In this sense, if there is any room for financial physics it can come only after the failure of economic and financial theory to explain financial data. The contribution of physics to finance can be meaningful only by better understanding of finance—however complex physical notions may be. The true test is, as always, the 'proof of the pudding'; in other words, whether models are supported by the evidence of financial data or making money where no one else thought money could be made.

### SELECTED INTRODUCTORY READING

- Bachelier, L. (1900) *Théorie de la spéculation*, Thèse de Mathématique, Paris.
- Barrois, T. (1834) Essai sur l'application du calcul des probabilités aux assurances contre l'incendie, *Mem. Soc. Sci. De Lille*, 85–282.
- Beard, R.E., T. Pentikainen and E. Pesonen (1979) *Risk Theory* (2nd edn), Methuen, London.
- Black, F., and M. Scholes (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.
- Borch, K.H. (1968) *The Economics of Uncertainty*, Princeton University Press, Princeton, N. J.
- Bouchaud, J.P., and M. Potters (1997) *Théorie des Risques Financiers*, Aléa-Saclay/Eyrolles, Paris.
- Cootner, P.H. (1964) *The Random Character of Stock Prices*. MIT Press, Cambridge, MA.
- Cramer, H. (1955) *Collective Risk Theory* (Jubilee Volume), Skandia Insurance Company.
- Hull, J. (1993) *Options, Futures and Other Derivatives Securities* (2nd edn), Prentice Hall, Englewood Cliffs, NJ.
- Ingersoll, J.E., Jr (1987) *Theory of Financial Decision Making*, Rowman & Littlefield, New Jersey.
- Jarrow, R.A. (1988) *Finance Theory*, Prentice Hall, Englewood Cliffs, NJ.
- Kalman, R.E. (1994) Randomness reexamined, *Modeling, Identification and Control*, **15**(3), 141–151.

- Lundberg, F. (1932) Some supplementary researches on the collective risk theory, *Skandinavisk Aktuarietidskrift*, **15**, 137–158.
- Merton, R.C. (1990) *Continuous Time Finance*, Cambridge, M.A, Blackwell.
- Modigliani, F., and M. Miller (1958) The cost of capital and the theory of investment, *American Economic Review*, **48**(3), 261–297.
- Tetens, J.N. (1786) *Einleitung zur Berchnung der Leibrenten und Antwartschaften*, Leipzig.

## CHAPTER 2

# Making Economic Decisions under Uncertainty

### 2.1 DECISION MAKERS AND RATIONALITY

Should we invest in a given stock whose returns are hardly predictable? Should we buy an insurance contract in order to protect ourselves from theft? How much should we be willing to pay for such protection? Should we be rational and reach a decision on the basis of what we know, or combine our prior and subjective assessment with the unfolding evidence? Further, do we have the ability to use a new stream of statistical news and trade intelligently? Or ‘bound’ our procedures? This occurs in many instances, for example, when problems are very complex, outpacing our capacity to analyse them, or when information is so overbearing or so limited that one must take an educated or at best an intuitive guess. In most cases, steps are to be taken to limit and ‘bound’ our decision processes for otherwise no decision can be reached in its proper time. These ‘bounds’ are varied and underlie theories of ‘bounded rationality’ based on the premise that we can only do the best we can and no better! However, when problems are well defined, when they are formulated properly – meaning that the alternatives are well-stated, the potential events well-established, and their conditional consequences (such as payoffs, costs, etc.) are determined, we can presume that a rational procedure to decision making can be followed. If, in addition, the uncertainties inherent in the problem are explicitly stated, a rational decision can be reached.

What are the types of objectives we may consider? Although there are several possibilities (as we shall see below) it is important to understand that no criterion is the objectively correct one to use. The choice is a matter of economic, individual and collective judgement – all of which may be imbued with psychological and behavioural traits. Utility theory, for example (to be seen in Chapter 3), provides an approach to the selection of a ‘criterion of choice’ which is both consistent and rational, making it possible to reconcile (albeit not always) a decision and its

economic and risk justifications. It is often difficult to use, however, as we shall see later on for it requires parameters and an understanding of human decision making processes that might not be available.

To proceed rationally it is necessary for an individual decision-maker (an investor for example) to reach a judgement about: the alternatives available, the sources of uncertainties, the conditional outcomes and preferences needed to order alternatives. Then, combine them without contradicting oneself (i.e. by being rational) in selecting the best course of action to follow. Further, to be rational it is necessary to be self-consistent in stating what we believe or are prepared to accept and then accept the consequences of our actions. Of course, it is possible to be 'too rational'. For example, a decision maker who refuses to accept any dubious measurements or assumptions will simply never make a decision! He then incurs the same consequences as being irrational. To be a practical investor, one must accept that there is a 'bounded rationality' and that an investment will in the end bear some risk one did not plan on assuming. This understanding is an essential motivation for financial risk management. That is, we can only be satisfied that we did the best possible analysis we could, given the time, the information and the techniques available at the time the decision to invest (or not) was made. Appropriate rational decision-making approaches, whether these are based on theoretical and/or practical considerations, would thus recognize both our capacities and their limit.

### **2.1.1 The principles of rationality and bounded rationality**

Underlying rationality is a number of assumptions that assume (Ariel Rubinstein, 1998):

- knowledge of the problem,
- clear preferences,
- an ability to optimize,
- indifference to equivalent logical descriptions of alternative and choice sets.

Psychologists and economists have doubted these. The fact that decisions are not always rational does not mean that there are no underlying procedures to the decision-making process. A systematic approach to departures from rationality has been a topic of intense economic and psychological interest of particular importance in finance, emphasizing 'what is' rather than 'what ought to be'.

For example, decision-makers often have a tendency to 'throw good money after bad', also known as sunk costs. Although it is irrational, it is often practised. Here are a few instances: Having paid for the movie, I will stay with it, even though it is a dreadful and time-consuming movie. An investment in a stock, even if it has failed repeatedly, may for some irrational reason generate a loyalty factor. The reason we are so biased in favour of bringing existing projects to fruition irrespective of their cost is that such behaviour is imbedded in our brains. We resist the conceptual change that the project is a failure and refuse to change our decision process to admit such failure. The problem is psychological: once we

have made an irreversible investment, we imbue it with extra value, the price of our emotional 'ownership'. There are many variations of this phenomenon. One is the 'endowment effect' in which a person who is offered \$10 000 for a painting he paid only \$1000 for refuses the generous offer. The premium he refuses is accounted for by his pride in an exceptionally good judgement—truly, perhaps the owner's wild fantasy that make such a painting wildly expensive. Similarly, once committed to a bad project one becomes bound to its outcome. This is equivalent to an investor to being OTM (on the money) in a large futures position and not exercising it. Equivalently, it is an alignment, not bounded by limited responsibility, as would be the case for stock options traders; and therefore it leads to maintaining an irrational risky position.

Currently, psychology and behavioural studies focus on understanding and predicting traders' decisions, raising questions regarding markets' efficiency (meaning: being both rational and making the best use of available information) and thereby raising doubts regarding the predictive power of economic theory. For example, aggregate individual behaviour leading to herding, black sheep syndrome, crowd psychology and the tragedy of the commons, is used to infuse a certain reality in theoretical analyses of financial markets and investors' decisions. It is with such intentions that funds such as ABN AMRO Asset Management (a fund house out of Hong Kong) are proposing mutual investment funds based on 'behavioural finance principles' (IHT Money Report, 24–25 February 2001, p. 14). These funds are based on the assumption that investors make decisions based on multiple factors, including a broad range of identifiable emotional and psychological biases. This leads to market mechanisms that do not conform to or are not compatible with fundamental theory (as we shall see later on) and therefore, provide opportunities for profits when they can be properly apprehended. The emotional/psychological factors pointed out by the IHT article are numerous. *'Investors' mistakes are not due to a lack of information but because of mental shortcuts inherent in human decision-making that blinds investors. For example, investors overestimate their ability to forecast change and they inefficiently process new information. They also tend to hold on to bad positions rather than admit mistakes.'* In addition, image bias can keep investors in a stock even when this loyalty flies in the face of balance sheet fundamentals. Over-reaction to news can lead investors to dump stocks when there is no rational reason for doing so. Under-reaction is the effect of people's general inability to admit mistakes. This is a trait that is also encountered by analysts and fund managers as much as individual investors. These factors are extremely important for they underlie financial practice and financial decision-making, drawing both on theoretical constructs and an appreciation of individual and collective (market) psychology. Thus, to construct a rational approach to making decisions, we can only claim to do the best we can and recognize that, however thorough our search, it is necessarily bounded.

Rationality is also a 'bounded' qualitative concept that is based on essentially three dimensions: analysis of information, perception of risk and decision-making. It may be defined and used in different ways. 'Classical rationality', underlying important economic and financial concepts such as 'rational

expectations' and 'risk-neutral pricing' (we shall attend to this later on in great detail), suppose that the investor/decision maker uses all available information, perceives risk without bias and makes the best possible investment decision he can (given his ability to compute) with the information he possesses at the time the decision is made. By contrast, a 'Bayesian rationality', which underlies this chapter, has a philosophically different approach. Whereas 'rational expectations' supposes that an investor extrapolates from the available information the true distribution of payoffs, Bayesian rationality supposes that we have a prior subjective distribution of payoffs that is updated through a learning mechanism with unfolding new information. Further, 'rational expectations' supposes that this prior or subjective distribution is the true one, imbedding future realizations while the Bayes approach supposes that the investor's belief or prior distribution is indeed subjective but evolving through learning about the true distribution. These 'differences of opinion' have substantive impact on how we develop our approach to financial decision making and risk management. For 'rational expectations', the present is 'the present of the future' while Bayesian rationality incorporates learning from one's bias (prejudice or misconception) into risk measurement and hence decision making, the bias being gradually removed as learning sets in. In this chapter we shall focus our attention on Bayes decision making under uncertainty.

## 2.2 BAYES DECISION MAKING

The basic elements of Bayes rational decision making involve behaviours including:

- (1) A decision to be taken from a set of known alternatives.
- (2) Uncertainty defined in terms of events with associated known (subjective) probabilities.
- (3) Conditional consequences resulting from the selection of a decision and the occurrence of a specific event (once uncertainty, ex-post, is resolved).
- (4) A preference over consequences, i.e. there is a well-specified preference function or procedure for selecting a specific alternative among a set of given alternatives.

An indifferent decision maker does not really have a problem. A problem arises when certain outcomes are preferred over others (such as making more money over less) and when preferences are sensitive to the risks associated with such outcomes. What are these preferences? There are several possibilities, each based on the information available – what is known and not known and how we balance the two and our attitude toward risk (or put simply, how we relate to the probabilities of uncertain outcomes, their magnitude and their adverse consequences). For these reasons, *risk management* in practice is very important, impacting events' desirability and their probabilities. There are many ways to do so, as we shall see below.

### 2.2.1 Risk management

Risk results from the direct and indirect adverse consequences of outcomes and events that were not accounted for, for which we are ill-prepared, and which effects individuals, firms, financial markets and society at large. It can result from many reasons, both internally induced and occurring externally. In the former case, consequences are the result of failures or misjudgements, while, in the latter, these are the results of uncontrollable events or events we cannot prevent. As a result, a definition of risk involves (i) consequences, (ii) their probabilities and their distribution, (iii) individual preferences and (iv) collective, market and sharing effects. These are relevant to a broad number of fields as well, each providing an approach to measurement, valuation and minimization of risk which *is motivated by psychological needs and the need to deal with problems that result from uncertainty and the adverse consequences they may induce.*

Risk management is broadly applied in finance. Financial economics, for example, deals intensively with hedging problems in to order eliminate risks in a particular portfolio through a trade or a series of trades, or through contractual agreements reached to share and induce a reduction of risk by the parties involved. Risk management consists then in using financial instruments to negate the effects of risk. It might mean a judicious use of options, contracts, swaps, insurance contracts, investment portfolio design etc. so that risks are brought to bearable economic costs. These tools cost money and, therefore, risk management requires a careful balancing of the numerous factors that affect risk, the costs of applying these tools and a specification of (or constraints on) tolerable risks an economic optimization will be required to fulfil. For example, options require that a premium be paid to limit the size of losses just as the insured are required to pay a premium to buy an insurance contract to protect them in case an adverse event occurs (accidents, thefts, diseases, unemployment, fire, etc.). By the same token, ‘value at risk’ (see Chapter 10) is based on a quantile risk constraint, which provides an estimate of risk exposure. Each profession devises the tools it can apply to manage the more important risks to which it is subjected.

The definition of risk, risk measurement and risk management are closely related, one feeding the other to determine the proper/optimal levels of risk. In this process a number of tools are used based on:

- ex-ante risk management,
- ex-post risk management and
- robustness.

Ex-ante risk minimization involves the application of *preventive controls*; preventive actions of various forms; information seeking, statistical analysis and forecasting; design for reliability; insurance and financial risk management etc. Ex-post risk minimization involves by contrast control audits, the design of optional, flexible-reactive schemes that can deal with problems once they have occurred and limit their consequences. Robust design, unlike ex-ante and ex-post risk minimization, seeks to reduce risk by rendering a process insensitive to its

adverse consequences. Thus, *risk management consists of altering the states a system may reach in a desirable manner (financial, portfolio, cash flow etc.), and their probabilities or reducing their consequences to planned or economically tolerable levels.* There are many ways to do so, however, each profession devises the tools it can apply or create a market for. For example, insurance firms use reinsurance to share the risks insured while financial managers use derivative products to contain unsustainable risks.

Risk management tools are applied in insurance and finance in many ways. *Control* seeks to ascertain that ‘what is intended occurs’. It is exercised in a number of ways rectifying decisions taken after a nonconforming event or problem has been detected. For example, auditing a trader, controlling a portfolio performance over time etc. are such instances. The disappearance of \$750 million at AIB (Allied Irish Bank) in 2002 for example, accelerated implementation of control procedures within the bank and its overseas traders.

*Insurance* is a medium or a market for risk, substituting payments now for potential damages (reimbursed) later. The size of such payments and the potential damages that may occur with various probabilities, can lead to widely distributed market preferences and thereby to a possible exchange between decision-makers of various preferences. Insurance firms have recognized the opportunities of such differences and have, therefore, provided mechanisms for pooling, redistributing and capitalizing on the ‘willingness to pay to avoid losses’. It is because of such attitudes, combined with goals of personal gain, social welfare and economic efficiency, that markets for fire and theft insurance, as well as sickness, unemployment, accident insurance, etc., have come to be as important as they are today. It is because of persons’ or institutions’ desires to avoid too great a loss (even with small probabilities), which would have to be borne alone, that markets for reinsurance (i.e., sub-selling portions of insurance contracts) and mutual protection insurance (based on the pooling of risks) have also come into being. Today, risk management in insurance has evolved and is much more in tune with the valuation of insurance risks by financial markets. Understanding the treatment of risk by financial markets; the ‘law of the single price’ (which we shall consider below); risk diversification (when is possible) and risk transfer techniques using a broad set of financial instruments currently used and traded in financial markets; the valuation of risk premiums and the estimation of yield curves (see also Chapter 8); mastering financial statistical and simulation techniques; and finally devising applicable risk metrics and measurement approaches for insurance firms – all have become essential for insurance risk management.

While insurance is a *passive form of risk management*, based on exchange mechanisms only (or, equivalently, ‘passing the buck’ to some willing agent), loss prevention and technological innovations are *active means of managing risks*. Loss prevention is a means of altering the probabilities and the states of undesirable, damaging states. For example, maintaining one’s own car properly is a form of loss prevention seeking to alter the chances of having an accident. Similarly, driving carefully, locking one’s own home effectively, installing fire alarms, etc. are all forms of loss prevention. Of course, insurance and loss prevention are, in

fact, two means to the similar end of risk protection. Car insurance rates tend, for example, to be linked to a person's past driving record. Certain clients (or areas) might be classified as 'high risk clients', required to pay higher insurance fees. Inequities in insurance rates will occur, however, because of an imperfect knowledge of the probabilities of damages and because of the imperfect distribution of information between the insured and insurers. Thus, situations may occur where persons might be 'over-insured' and have no motivation to engage in loss prevention. Such outcomes, known as 'moral hazard' (to be seen in greater detail in Chapter 3), counter the basic purposes of insurance. It is a phenomenon that can recur in a society in widely different forms, however. Over-insuring unemployment may stimulate persons not to work, while under-insuring may create uncalled-for social inequities. Low car insurance rates (for some) can lead to reckless driving, leading to unnecessary damages inflicted on others, on public properties, etc. Risk management, therefore, seeks to ensure that risk protection does not become necessarily a reason for not working. More generally, risk management in finance considers both risks to the investor and their implications for returns, 'pricing one at the expense of the other'. In this sense, finance, has gone one step further in using the market to price the cost an investor is willing to sustain to prevent the losses he may incur. Financial instruments such as options provide a typical example. For this reason, given the importance of financial markets, many insurance contracts have to be reassessed and valued using basic financial instruments.

*Technological innovation* means that the structural process through which a given set of inputs is transformed into an output is altered. For example, building a new six-lane highway can be viewed as a way for the public to change the 'production-efficiency function' of transport servicing. Environmental protection regulation and legal procedures have, in fact, had a technological impact by requiring firms to *change* the way in which they convert inputs into outputs, by considering as well the treatment of refuse. Further, pollution permits have induced companies to reduce their pollution emissions in a given by-product and sell excess pollution to less efficient firms.

*Forecasting, learning, information and its distribution* is also an essential ingredient of risk management. Banks learn every day how to price and manage risk better, yet they are still acutely aware of their limits when dealing with complex portfolios of structured products. Further, most non-linear risk measurement and assessment are still 'terra incognita' asymmetries. Information between insured and insurers, between buyers and sellers, etc., are creating a wide range of opportunities and problems that provide great challenges to risk managers and, for some, 'computational headaches' because they may be difficult to value. These problems are assuming added importance in the age of internet access for all and in the age of 'total information accessibility'. Do insurance and credit card companies have access to your confidential files? Is information distribution now swiftly moving in their favour? These are issues creating 'market inefficiencies' as we shall see in far greater detail in Chapter 9.

*Robustness* expresses the insensitivity of a process to the randomness of parameters (or mis-specification of the model) on which it is based. The search for

robust solutions and models has led to many approaches and techniques of optimization. Techniques such as VaR (Value at Risk), scenario optimization, regret and ex-post optimization, min-max objectives and the like (see Chapter 10) seek to construct robust systems. These are important tools for risk management; we shall study them here at length. They may augment the useful life of a portfolio strategy as well as provide a better guarantee that ‘what is intended will likely occur’, even though, as reality unfolds over time, working assumptions made when the model was initially constructed turn out to be quite different.

Traditional decision problems presume that there are *homogeneous* decision makers, deciding as well what information is relevant. In reality, decision makers may be heterogeneous, exhibiting broadly varying preferences, varied access to information and a varied ability to analyse (forecast) and compute it. In this environment, decision-making becomes an extremely difficult process to understand and decisions become difficult to make. For example, when there are few major traders, the apprehension of each other’s trades induces an endogenous uncertainty, resulting from a mutual assessment of intentions, knowledge, knowhow etc. A game may set in based on an appreciation of strategic motivations and intentions. This may result in the temptation to collude and resort to opportunistic behaviour.

## 2.3 DECISION CRITERIA

The selection of a decision criterion is an essential part of DMUU, expressing decision-makers’ impatience and attitudes towards uncertain outcomes and valuing them. Below we shall discuss a few commonly used approaches.

### 2.3.1 The expected value (or Bayes) criterion

Preferences for decision alternatives are expressed by sorting their expected outcomes in an increasing order. For monetary values, the Expected Monetary Value (or EMV) is calculated and a choice is made by selecting the greatest EMV. For example, given an investment of 3 million dollars yielding an uncertain return one period hence (with a discount rate of 7%), and given in the returns in the table below, what is the largest present expected value of the investment? For the first, alternative we calculate the EMV of the investment one period hence and obtain:  $EMV = 4.15$ . The current value of the investment is thus equal to the present value of the expected return (EMV less the cost of the investment) or:

$$V(I) = (1 + r)^{-1}EMV - I = 4.15 * (1 + 0.07)^{-1} - 3 = 0.878$$

Probability	0.10	0.20	0.30	0.15	0.15	0.10
Return	-4	-1	5	7	8	10

When there is more than one alternative (measured by the initial outlay and forecasting of future cash flows), a decision is then reached by comparing the

economic properties of each investment alternative. For example, consider another investment proposal consisting of an initial outlay of 1 million dollars only (rather than 3) with a prospective cash flow given by the following:

Probability	0.10	0.20	0.30	0.15	0.15	0.10
Return	-8	-3	5	3	4	8

If we maintain the same EMV criterion, we note that:

$$V(I) = (1 + r)^{-1} EMV - I = 1.95 * (1 + 0.07)^{-1} - 1 = 0.822$$

which clearly ranks the first investment alternative over the second (in terms of the EMV criterion). In both cases the EMV is positive and therefore both projects seem to be economically worthwhile. There may be other considerations, for example, an initial outlay of 3 (rather than 1) million dollars for sure compared to an uncertain cash flow in the future (with prospective potential losses, albeit probabilistic, in the future). The attitude towards these losses are often important considerations to consider as well. Such considerations require the application of other criteria for decision making, as we shall briefly outline below. Note that it is noteworthy that such an individual approach does not deal with the market valuation of such cash flow streams and expresses only an individual's judgement (and not market valuation of the cash flow, that is the consensus of judgements of participants on a market price). Financial analysis, as we shall see subsequently, provides a market-sensitive discounting to these uncertain streams of cash.

### 2.3.2 Principle of (Laplace) insufficient reason

The Laplace principle states that, when the probabilities of the states of nature in a given problem are not known, we assume they are equally likely. In other words, a state of utmost ignorance will be replaced by assigning to each potential state the same probability! In this case, when we return to our first investment project, we are faced with the following prospect:

Probability	0.166	0.166	0.166	0.166	0.166	0.166
Return	-4	-1	5	7	8	10

and its present EMV is,

$$V(I) = (1 + r)^{-1} EMV - I = 5 * (1 + 0.07)^{-1} - 3 = 1.672$$

which implies that 'not knowing' can be worth money! This is clearly not the case, since reaching a decision on this basis can lead to losses since the probability we have assumed are not necessarily the true ones. Gathering information in these cases may be useful, since it may be used to reduce the potential (miscalculated) expected losses.

### 2.3.3 The minimax (maximin) criterion

The criterion consists in selecting the decision that will have the least maximal loss regardless of what future (state) may occur. It is used when we seek protection from the worst possible events and expresses generally an attitude of abject pessimism. Consider again the two investment projects with cash flows I and II specified below and for simplicity, assume that they require initially the same investment outlay. The flows to compare are:

Probability	0.10	0.20	0.30	0.15	0.15	0.10
Return I	-4	-1	-5	-7	8	10
Return II	-8	-1	5	7	8	10

The worst prospect in the first project is  $-7$  million dollars while it is  $-8$  million dollars in the second project. The minimum of the maximum loss is therefore  $-7$  million dollars, which provides a criterion (albeit very pessimistic) justifying the selection of the first investment project.

The minimax criterion takes the smallest of the available maximums. In this case, the projects have an equal maximum value and the investor is indifferent between the two. It is a second-best objective. Who cares about getting the gold medal as long as we get the silver! Honour is safe and the player satisfied. This criterion can be extended using this sporting analogy. A bronze is third best, good enough; while fourth best may be just participating, providing a reward in itself. Maximin is a loss-averse mindset. As long as we do get the best of all worst possible outcomes the investor is satisfied.

### 2.3.4 The maximax (minimin) criterion

This is an optimist's criterion, banking on the best possible future, yielding the hoped for largest possible profits. It is based on the belief or the urge to profit as much as possible, regardless of the probability of desirable or other events. Again, returning to our previous example, we note that both projects have a maximal gain of 10 million dollars and therefore the maximum-maximum gain (maximax criterion) will indicate indifference in selecting one or the other project, as was the case for the minimax criterion. As Voltaire's *Candide* would put it: 'We live in the best of all possible worlds' as he travelled in a world ravaged by man, as a prelude story to the French Revolution.

The minimin criterion is a pessimist's point of view. Regardless of what happens, only the worst case can happen. On the upside, such a point of view, leads only to upbeat news. My house has not burned today! Amazing!

### 2.3.5 The minimax regret or Savage's regret criterion

The previous criteria involving maximums and minimums were evaluated ex-ante. In practice, payoffs and probabilities are not easily measured. Thus, these criteria

express a philosophical outlook rather than an objective to base a decision on. Ex-post, unlike ex-ante, decision-making is reached once information is revealed and uncertainty is resolved. Each decision has then a regret defined by the difference between the gain made and the gain that could have been realized had we selected the best decision (associated with the event that actually occurs). An expected ‘regret’ decision-maker would then seek to minimize the expectation of such a regret, while a minimax regret decision-maker would seek to select the decision providing the least maximal regret.

The cost of a decision’s regret represents the difference between the ex-ante payoffs that would be received with a given outcome compared to the maximum possible ex-post payoff received. Savage, Bell and Loomes and Sugden (see references) have pointed out the relevance of this criterion to decision-making under uncertainty by suggesting that decision makers may select an act by minimizing the regrets associated with potential decisions. Behaviourally, such a criterion would be characteristic of people attached to their past. Their past mistakes haunt their present day, hence, they do the best they can to avoid them in the future. Specifically, assume that we select an action (decision) and some event occurs. The decision/event combination generates a *payoff table*, expressing the conditional consequences of that decision when, ex-post, the event occurs. For example, the following table gives the payoff on a portfolio dependent on two different decisions on the portfolio allocation.

	Event A	Event B	Event C	Event D	Event E	Event F
Probability	0.10	0.20	0.30	0.15	0.15	0.10
Return I	−3	−1	−5	−7	8	10
Return II	−8	1	6	7	8	12

The decision/event combination may then generate a ‘regret’ for the decision – for it is possible that we could have done better! Was decision 1 the better one? This is an opportunity loss, since a profit could have been made – had we known what events were to occur. If event B is the one that happens then clearly, based on an ex-post basis, decision 2 is the better one. If decisions were reversible then it might be possible to compensate (at least partially) for the fact that we took, a posteriori, a ‘wrong’ decision. Such a characteristic is called ‘flexibility’ and is worth money that decision makers are willing to pay for. What would I be willing to pay to have taken effectively decision 1 instead of 2 when event B happens? Options for example, provide such an opportunity, as we shall see in Chapter 6. An option would give us the right but not the obligation to make a decision in the future, once uncertainty is resolved. In most cases, these are decisions to sell or buy. But applications to real world problems have led to options to switch from one technology to another for example.

For example, say that we expect the demand for a product to grow significantly, and as a result we decide to expand the capacity of our plants. Assume that in fact, this expectation for demand growth does not materialize and we are left with

a large excess capacity, unable to reduce it except at a substantial loss. What can we do then, except regret our decision! Similarly, assume that we expect peace to come on earth and decide to spend less on weapons development. Optimism, however much it may be wanted, may not, unfortunately, be justified and instead we find ourselves facing a war for which we may be ill-prepared. What can we do? Not much, except regret our decision. The *regret* (also called the *Savage regret*) criterion, then, seeks to minimize the regret we may have in adopting a decision. This explains why some actions are taken to reduce the possibility of such extreme regrets (as with the buying of insurance, steps taken to reduce the risks of bankruptcy, buying options to limit downside risk, in times of peace prepare for war – Sun-Tze and so on). Examples to this effect will be considered below using the opportunity loss table in the next section (Table 2.3).

### **Example: Regret and the valuation of firms**

Analysts' valuation of stocks are growing in importance. Analyst recommendations have a great impact on investors, but their effects are felt particularly when analysts are 'disappointed' by a stock performance and revise their recommendations downwards. In these cases, the effects can be disastrous for the stock price in consideration. In practice, analysts use a number of techniques that are based on firms' reports. Foremost is the net return multiple factor. It is based on the ratio of the stock value of the firm to its net return. The multiple factor is then selected by comparing firms that have the same characteristics. It is then believed that *the larger the risk, the smaller the multiple factor*. In practice, analysts price stocks quite differently. A second technique is based on the firm's future discounted (at the firm's internal rate of return) cash flow. In practice, the future cash flow is based on forecasts that may not be precise. Finally, the third technique is based on assets value (which is the most conservative one). In other words, there is not a uniform agreement regarding which objective to use in valuing a firm's stock. Financial fundamental theory has made an important contribution by providing a set of proper circumstances to resolve this issue. This will be considered in Chapter 6 in particular.

### **Example: The firm and risk management**

Consider a firm operating in a given industry. Evidently, competition with other firms, as well as explicit (or implicit) government intervention through regulation, tax rebates for special environmental protection investments, grants or subsidized capital budgets in distress areas, etc., are instances where firms are required to be sensitive to uncertainty and risk. Managers, of course, will seek to reduce and manage the risk implied by such uncertainty and seek ways to augment the market control (by vertical integration, acquisition of competition, etc.), or they may diversify risks by seeking activities in unrelated markets.

In the example Table 2.1 we have constructed a list of uncertainties and risks faced by firms and how these may be met. The list provided is by no means exhaustive and provides only an indication of the kind of problems that we can address. For example, competition can be an important source of risk which may be met by many means such as strategic M & A, collusion practices, diversification and so on.

**Table 2.1** Sources of uncertainty and risks.

Uncertainty and risks	Protective actions taken
Long-range changes in market growth	Research and development on new products, diversification to other markets
Inflation	Indexation of assets, and accounts receivable
Price uncertainty of input materials	Building up inventory, contract with suppliers (essentially futures), buying options and hedging techniques
Competition	Mergers and acquisition, cartels, price-fixing, advertising and marketing effort, diversification

## 2.4 DECISION TABLES AND SCENARIO ANALYSIS

Decision tables and trees are simple mechanisms for structuring some decision problems involving uncertainty and solving them. It requires that an objective, the problem's states and probabilities be given. To construct a payoff table we proceed as follows:

- Identify the alternative courses of action, mutually exclusive, and collectively exhaustive, which are variables (at least two) we can control directly.
- Consider all possible and relevant states of a problem. Each state represents one and only one potential event; each state may itself be defined in terms of multiple other states, however; states represent events which are mutually exclusive; they are collectively exhaustive; one and only one state will actually result.
- Assign to each state a probability of occurrence. This probability should be based on the information we have regarding the problem and, since states are mutually exclusive and collectively exhaustive, these probabilities (summed over all states) should be equal to one.

All conditional (payoff or cost) consequences are then assembled in a table format – see Table 2.2 where  $[c_{ij}]$  are the conditional costs of alternative  $i$  if event  $j$  occurs.

**Table 2.2** The payoff table.

States	1	2	3	...	...	...	$n$
Probabilities	$p_1$	$p_2$	$p_3$	...	...	...	$p_n$
$A_1$	$c_{11}$	$c_{12}$	$c_{13}$	...	...	...	$c_{1n}$
$A_2$	$c_{21}$	$c_{22}$	$c_{23}$	...	...	...	$c_{2n}$
$A_3$	$c_{31}$	$c_{32}$	$c_{33}$	...	...	...	$c_{3n}$
...	...	...	...	...	...	...	...
$A_m$	$c_{m1}$	$c_{m2}$	$c_{m3}$				$c_{mn}$

For example, for a credit manager, what are the relevant states to consider when a customer comes in and demands a loan? Simply grant the loan (state 1) or not (state 2). If the loan is not reimbursed on time (and reimbursement delays are introduced) there may be other ways to express these states. For example, a first state would stand for no delay, a second, would stand for a one-period delay, a third state for a two-period delay, and so on. The entries in the table tell us what will be the conditional payoffs (or costs) associated with each action. The sample Table 2.2 specifies  $n$  states  $1, 2, 3, 4, \dots, n$  and  $m$  alternatives. When alternative  $A_i, i = 1, 2, \dots, m$  is taken and say state  $j$  occurs with probability  $p_j, j = 1, 2, \dots, n$ , then the cost (or payoff) is  $c_{ij}$  (or  $\pi_{ij}$ ). Thus, in such a decision problem, there are:

- (1)  $n$  potential, mutually exclusive and exhaustive states,
- (2)  $m$  alternative actions, one of which only can be selected,
- (3)  $nm$  conditional consequences we should be able to define.

If we use an expected cost (or expected payoff) criterion, then the decision selected would be the one yielding the least expected cost (or, equivalently, the largest expected payoff). The expected monetary cost of alternative  $i$  is then:

$$EMC_i = \sum_{j=1}^n p_{ij} c_{ij}$$

while the least cost alternative  $k$  selected is:

$$k \in \text{Min}_{i \in \{1, \dots, n\}} \{EMC_i\}$$

### Problem

Cash management consists of managing the short-term flow of funds in order to meet a potential need or demand for cash. Cash is kept primarily because of its need in the future. Assume, for example, that an investor has the following needs for money:

Quantities	100	300	500	700	900
Probabilities	0.05	0.25	0.50	0.15	0.05

- (1) What are the potential courses of action?
- (2) What are the problem states? And their probabilities?
- (3) What are the conditional costs if the bank rate is 20 % yearly?

#### 2.4.1 The opportunity loss table

Say that action  $i$  has been selected and event  $j$  occurs and thus payoff  $\pi_{ij}$  is gained. If we were equipped with this knowledge prior to making a decision, it is possible that another decision would bring greater profits. Assume such knowledge and let the maximum payoff, based on the best decision be

$$\text{Max}_j [\pi_{ij}]$$

**Table 2.3** Opportunity loss table.

States	1	2	3	...	...	...	n
Probabilities	$p_1$	$p_2$	$p_3$	...	...	...	$p_n$
$A_1$	$l_{11}$	$l_{12}$	$l_{13}$	...	...	...	$l_{1n}$
$A_2$	$l_{21}$	$l_{22}$	$l_{23}$	...	...	...	$l_{2n}$
$A_3$	$l_{31}$	$l_{32}$	$l_{33}$	...	...	...	$l_{3n}$
...	...	...	...	...	...	...	...
$A_m$	$l_{m1}$	$l_{m2}$	$l_{m3}$				$l_{mn}$

The difference between this maximum payoff and the payoff obtained by taking any other decision is called the opportunity loss, denoted by:

$$l_{ij} = \text{Max}_j [\pi_{ij}] - \pi_{ij}$$

The opportunity loss table is therefore a matrix as given in Table 2.3.

Thus, the opportunity loss is the difference between the costs or profits actually realized and the costs or profits which would have been realized if the decision had been the best one possible. A project might seem like a good investment, but it means that we have lost the opportunity to do something else that might be more profitable. This loss may be likened to the additional income a trader would have realized had he been an inside trader, benefiting from information regarding stock prices before they reach the market! As a result, we can verify that the difference between the expected profits of any two acts is equal in magnitude but opposite in sign to the difference between their expected losses. By the same token, the difference between the expected costs of any two acts is equal in magnitude and identical in sign to the difference between their expected opportunity losses. With these definitions on hand we can also state that: *the cost of uncertainty is the expected opportunity loss of the best possible decision under a given probability distribution.*

## 2.5 EMV, EOL, EPPI, EVPI

EMV, EOL, EPPI and EVPI are terms associated with a decision; they will be elucidated through an application. Assume that data supplied by a Port Authority points to a number of development alternatives for the port. Uncertainty regarding the economic state of the country, geopolitical developments and so on, lead to a number of scenarios to be considered and against which each of these alternatives must be assessed. Each alternative can generate, ex-post, a sense of satisfaction at having followed the proper course of action as well as a sense that a suboptimal alternative was taken. Four scenarios are assumed each to lead to the following results, summarized in the table below where entries are payoffs (losses):

Scenario	1	2	3	4	5	6
Probability	0.1	0.15	0.25	0.05	0.3	0.15
Alternative 1	-200	-100	150	400	-300	700
Alternative 2	300	-150	300	600	100	500
Alternative 3	-500	300	400	-100	400	100
Alternative 4	400	600	-100	-250	-300	100

### 2.5.1 The deterministic analysis

An alternative is selected irrespective of the probabilities of forthcoming events. Given a number of alternatives and specified events, a decision can be taken. A number of criteria are used, such as maximax, maximin, minimax regret and the 'equally likely' (Laplace) criteria as stated earlier. Under these criteria, we see that alternatives 1 and 2 are always better than alternatives 3 and 4. Explicitly, the following results are obtained:

Criterion	Decision	Payoff
Maximax	Alternative 1	700
Maximin	Alternative 2	-150
Minimax regret	Alternative 1	700
Equally likely	Alternative 2	275

### 2.5.2 The probabilistic analysis

Probabilistic analysis characterizes the likelihood of forthcoming events by associating a probability with each event. It uses a number of potential criteria but we shall be concerned essentially with the EMV – expected monetary value index of performance. The results for our example are given by the following:

Probabilistic analysis: The Port Authority  
 Expected value – Summary report  
 Decision      Expected payoff  
 Alternative 1    37.50  
 Alternative 2    217.50  
 Alternative 3    225.00 \*  
 Alternative 4    17.50

Calculations were made as follows:

Alternative 1:  $0.1(-200) + 0.15(-100) + 0.25(150) + 0.05(400) + 0.3(-300) + 0.15(700) = 37.50$

Alternative 2:  $0.1(300) + 0.15(-150) + 0.25(300) + 0.05(600) + 0.3(100) + 0.15(500) = 217.50$

Alternative 3:  $0.1(-500) + 0.15(300) + 0.25(400) + 0.05(-100) + 0.3(400) + 0.15(100) = 225.00$

Alternative 4:  $0.1(400) + 0.15(600) + 0.25(-100) + 0.05(-2500) + 0.3(-300) + 0.15(100) = 17.50$

The EMV (expected monetary value) consists of valuing each alternative by its EMV. The 'best' choice (in an EMV context) is 225. In other words, ex-ante, the best decision we can take is alternative 3. By contrast, if a decision could be taken ex-post, once uncertainty is revealed and removed, the cost of each decision is given by its opportunity loss, whose expectation is the EOL (expected opportunity loss). This value is calculated explicitly through the opportunity loss table below:

Table of opportunity losses, calculations

Scenario	1	2	3	4	5	6
Probability	0.1	0.15	0.25	0.05	0.3	0.15
Alternative 1	400 - (-200)	600 - (-100)	400 - 150	600 - 400	400 - (-300)	700 - 700
Alternative 2	400 - 300	600 - (-150)	400 - 300	600 - 600	400 - 100	700 - 500
Alternative 3	400 - (-500)	600 - 300	400 - 400	600 - (-100)	400 - 400	700 - 100
Alternative 4	400 - 400 = 0	600 - 600	400 - (-100)	600 - (-250)	400 - (-300)	700 - 100

Table of opportunity losses

Scenario	1	2	3	4	5	6
Probability	0.1	0.15	0.25	0.05	0.3	0.15
Alternative 1	60	105	62.5	10	210	0
Alternative 2	10	110.5	25	0	90	30
Alternative 3	90	45	0	35	0	90
Alternative 4	0	0	125	42.5	210	90

Entries are calculated as follows. Say that scenario 1 realizes itself. The best alternative would then be alternative 4 yielding a payoff of 400. We replace in the table the entry 400 by 0 and then calculate in the first column corresponding to Scenario 1 the relative losses had we selected a suboptimal alternative. Now compute for each alternative the expected opportunity loss, which is the sum of columns for each row. Verify that the sums  $EMV + EOL$  are equal for each alternative, called the EPPI, or the Expected Profit under Perfect Information. Further, note that the recommended alternative under an EOL criterion is also alternative 3 as in the expected payoff (EMV) case. This is always the case and should not come as any surprise, since selecting the largest EMV is equivalent to the smallest EOL. Since,

$$EMV + EOL = EPPI$$

Note that the EOL for the third alternative equals 260 and, therefore, note that the EPPI is 485, which is the same for all alternatives. The EPPI means that if, ex-post, we always have the best alternative, then in expectation our payoff would be 485. Since, ex-ante, it is only 225 (=EMV), the potential for improving the ex-ante payoff EMV by better forecasts of the scenarios, by a better management of uncertainty (through contracts of various sorts that manage risk) cannot be larger than the EOL or 260. Such an approach would be slightly more complex if

we were to introduce sample surveys, information guesses etc. used to improve our assessment of the states, the probabilities and the economic value of such an assessment.

Optimal Decision: Alternative 3; Expected payoff : 225.00  
 Probabilistic analysis  
 Expected value of perfect information

State	Prob.	Decision	Payoff	Prob.*Payoff
Scenario 1	0.1000	Alternative 4	400.00	40.00
Scenario 2	0.1500	Alternative 4	600.00	90.00
Scenario 3	0.2500	Alternative 3	400.00	100.00
Scenario 4	0.0500	Alternative 2	600.00	30.00
Scenario 5	0.3000	Alternative 3	400.00	120.00
Scenario 6	0.1500	Alternative 1	700.00	105.00
Expected payoff with perfect information (EPPI)				485.00
Expected payoff without perfect information (EMV)				225.00
Expected value of perfect information (EVPI)				260.00

If we integrate other sources of information, it is possible to improve the probability estimates and, therefore, improve the optimal decision. The value of information, of a sample on the basis of which such information is available, is called EVSI (or the expected value of sample information). It is a gain obtained by improving our assessment of the events/states probabilities. Finally, if gains and losses are weighted in a different manner, then we are led to approaches based on disappointment (giving greater weight to losses, relative to gains) and elation (when the prospects of 'doing better than expected' is more valued because of the self-gratification it produces). Avoidance of losses, motivated by disappointment, can also lead to selecting alternatives that have smaller gain expectation but reduce the probability of having made the 'wrong choice', in the sense of ending the development project with losses. We shall return to this approach in Chapter 3.

### Problem

The Corporate Financial Officer Vice of HardKoor Co. has the problem of raising some additional capital. To do so, it is possible to sell 10 000 convertible bonds. A preliminary survey of the capital market indicates that they could be sold at the present time for \$100 per bond. However, the company is currently engaged in a union contract dispute and there is a possibility of a strike. If the strike were to take place, the selling price of the bonds would be decreased by 20 %. There is also a possibility of winning a large, exclusive contract which, if obtained, would mean the bonds could be sold for 30 % more. The VP Finance would like to raise the maximum amount of capital, and so must decide whether to offer the bonds now or wait for the situation to become clearer.

- (a) What are the alternatives?
- (b) What are the sources and the types of uncertainty?
- (c) What action should be taken if an EMV criterion is used? – if a minimax criterion is taken? – if a maximin criterion is taken? – if a regret criterion is taken?

- (d) If the probability of a strike is felt to be 0.4, while the probability of the contract being awarded is 0.8, what action is best if the EMV criterion is applied (note that it is necessary to calculate the proceeds for the various outcomes), and if the expected opportunity loss (EOL) criterion is used?
- (e) Give one example of how the principle of bounded rationality was apparently used in formulating the problem?

### SELECTED REFERENCES AND READINGS

- Bell, D.E. (1982) Regret in decision making under uncertainty, *Operations Research*, **30**, 961–981.
- Bell, D.E. (1983) Risk premiums for decision regrets, *Management Science*, **29**, 1156–1166.
- Loomes, G., and R. Sugden (1982) Regret theory: An alternative to rational choice under uncertainty, *Economic Journal*, **92**, 805–824.
- Loomes, G., and R. Sugden (1987) Some implications of a more general form of regret theory, *Journal of Economic Theory*, **41**, 270–287.
- Luce, R.D., and H. Raiffa (1958) *Games and Decisions*, John Wiley & Sons, Inc., New York.
- Raiffa, H., and R. Schlaiffer (1961) *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business, Harvard University, Boston, MA.
- Rubinstein, A. (1998) *Modeling Bounded Rationality*, MIT Press, Boston, MA.
- Savage, L.J. (1954) *The Foundations of Statistics*, John Wiley & Sons, Inc., New York.
- Winkler, R.L. (1972) *Introduction to Bayesian Inference and Decision*, Holt, Rinehart & Winston, New York.

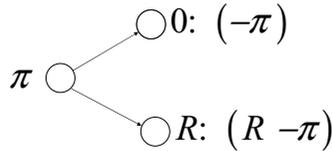
## CHAPTER 3

# Expected Utility

### 3.1 THE CONCEPT OF UTILITY

When the expected monetary value (EMV) is used as the sole criterion to reach a decision under uncertainty, it can lead to results we might not have intended. Outstanding examples to this effect are noted by observing people gambling in a casino or acquiring insurance. For example, in Monte Carlo, Atlantic City or Las Vegas, we might see people gambling (investing!) their wealth on ventures (such as putting \$100 on number 8 in roulette), knowing that these ventures have a negative expected return. To explain such an ‘irrational behaviour’, we may argue that not all people value money evenly. Alternatively we may rationalize that the prospect of winning  $36 * 100 = \$3600$  in a second at the whim of the roulette is worth taking the risk. After all, someone will win, so it might as well be me! Both an attitude towards money and the willingness to take risks, originating in a person’s initial wealth, emotional state and the pleasure to be evoked in some way by such risk, are reasons that may justify a departure from the Bayes EMV criterion. If all people were ‘straight’ expected payoff decision-makers, then there would be no national lotteries and no football or basketball betting. Even the mafia might be much smaller! People do not always use straight expected payoffs to reach decisions, however. The subjective valuation of money and people’s attitudes towards risk and gambling provide the basic elements that characterize gambling and the utility of money associated with such gambling. Utility theory seeks to represent how such subjective valuation of wealth and attitude towards risk can be quantified so that it may provide a rational foundation for decision-making under uncertainty.

Just as in Las Vegas we might derive ‘pleasure from gambling’, we may be also concerned by the loss of our wealth, even if it can happen with an extremely small probability. To protect ourselves from large losses, we often turn to insurance. Do we insure our house against fire? Do we insure our belongings against theft? Should we insure our exports against currency fluctuations or against default payment by foreign buyers? Do we invest in foreign lands without seeking insurance against national takeovers? And so on. In these situations and in order to avoid large losses, we willingly pay money to an insurance firm – the premium needed to buy such insurance. In other words, we transfer our risk to the insurer who in



**Figure 3.1** A lottery.

turn makes money by collecting the premium. Of course, how much premium to pay for how much risk insured underscores our ability to sustain a great loss and our attitude towards risk. Thus, just as our gambler was willing to pay a small amount of money to earn a very large one (albeit with a very small probability), we may be willing to pay a small amount (the premium) to prevent and protect ourselves from having to face a large loss, even if it occurs with a very small probability. In both cases, the Bayes expected payoff (EMV) criterion breaks down, for otherwise there would be no casinos and no insurance firms. Yet, they are here and provide an important service to society. Due to the importance of utility theory to economics and finance, providing a normative framework for decision-making under uncertainty and risk management, we shall outline its basic principles. Subsequently, we shall see how the concepts of expected utility have been used importantly in financial analysis and financial decision-making.

### 3.1.1 Lotteries and utility functions

Lotteries consist of the following: we are asked to pay a price  $\pi$  (say it is \$5) for the right to participate in a lottery and earn, potentially, another amount,  $R$ , called the reward (which is say \$1 000 000), with some probability,  $p$ . If we do not win the lottery, the loss is  $\pi$ . If we win, the payoff is  $R$ . This lottery is represented graphically in Figure 3.1 where all cash expenditures are noted. Lotteries of this sort appear in many instances. A speculator buys a stock expecting to make a profit (in probability) or losing his investment. Speculators are varied, however, owning various lotteries and possessing varied preferences for these lotteries. It is the exchange between speculators and investors that create a ‘financial market’ which, once understood, can provide an understanding and a valuation of lotteries pricing.

If we use an EMV criterion for valuing the lottery, as seen in the previous chapter, then the value of the lottery would be:

$$\text{Expected value of lottery} = p(R - \pi) - (1 - p)\pi = pR - \pi < 0$$

By participating in the lottery, we will be losing money in an expected sense. In other words, if we had ‘an infinite amount of money’ and were to play the lottery forever, then in the long run we would lose  $\$(\pi - pR)$ ! Such odds for lotteries are not uncommon, and yet, however irrational they may seem at first, many people play such lotteries. For example, people who value the prospect of ‘winning big’ even with a small probability much more than the prospect of

‘losing small’ even with a large probability, buy lottery tickets. This uneven valuation of money means that we may not be able to compare two sums of money easily. People are different in many ways, not least in their preferences for outcomes that are uncertain. An understanding of human motivations and decision making is thus needed to reconcile observed behaviour in a predictable and theoretical framework. This is in essence what expected utility theory is attempting to do. Explicitly, it seeks to define a scale that values money by some function, called the utility function  $U(\cdot)$ , whose simple expectation provides the scale for comparing alternative financial and uncertain prospects. The larger the expected utility, the ‘better it is’.

More precisely, the function  $U(\cdot)$  is a transformation of the value of money that makes lotteries of various sums comparable. Namely, the two sums  $(R - \pi)$  and  $(-\pi)$ , can be transformed into  $U(R - \pi)$  and  $U(-\pi)$ , and then the lottery would be,

- Make  $U(R - \pi)$  with a probability  $p$ .
- Make (lose)  $U(-\pi)$  with a probability  $1 - p$ .

while its expected value, which tells how valuable it is compared to other lotteries, is:

$$\text{Expected utility} = EU = pU(R - \pi) + (1 - p)U(-\pi)$$

This means that:

- If  $EU = 0$ , we are indifferent whether we participate in the lottery or not.
- If  $EU > 0$ , we are better off participating in the lottery.
- If  $EU < 0$ , we are worse off participating in the lottery.

Thus, participation in a lottery is measured by its expected utility. Further, the price  $\$p$  we will be willing to pay – the premium, for the prospect of winning  $\$R$  with probability  $p$  – is the price that renders the expected utility null, or  $EU = 0$ , found by the solution to

$$EU = 0 = pU(R - \pi) + (1 - p)U(-\pi)$$

which can be solved for  $\pi$  when the utility function is specified.

By the same token, expected utility can be used by an investor to compare various lotteries, various cash flows and payments, noting that the value of each has an expected utility, known for certain and used to scale the uncertain prospects. The ‘expected utility’ approach to decision-making under uncertainty is thus extremely useful, providing a rational approach ‘eliminating the uncertainty from decision-making’ and bringing it back to a problem under certainty, which we can solve explicitly and numerically. But there remains the nagging question: how can we obtain such utility functions? And how justified are we in using them? Von Neumann and Morgenstern, two outstanding mathematicians and economists, concluded in the late 1940s, that for expected utility to be justified as a scaling function for uncertain prospects the following holds:

- (1) The higher the utility the more desirable the outcome. This makes it possible to look for the best decision by seeking the decision that makes the expected utility largest.
- (2) If we have three possibilities (such as potential investment alternatives), then if possibility '1' is 'better' than '2' and '2' is better than '3', then necessarily '1' is better than '3'. This is also called the transitivity axiom.
- (3) If we are indifferent between two outcomes or potential acts, then necessarily the expected utilities will be the same.

These three assumptions, underlie the rational framework for decision making under uncertainty that expected utility theory provides.

### 3.2 UTILITY AND RISK BEHAVIOUR

An expected utility provides a quantitative expression of a decision makers' desires for higher rewards as well as his attitude towards the 'risks' of such rewards. Say that  $\{R, P(\cdot)\}$  is a set of rewards  $R$  assumed to occur with probability  $P(\cdot)$  and let  $u(\cdot)$  define a utility function. The basic utility theorem states that the expected utility provides an objective index to evaluate the desirability of rewards, or:

$$E(u(R)) = \int u(R)P(R) dR; R \in \Re$$

Given uncertain prospects, a rational decision-maker will then select that prospect whose expected utility is largest. For example, the EU of an alternative prospect  $i$  with probability outcomes  $(\pi_{ij}, p_{ij})$  is:

$$EU_i = \sum_{j=1}^n p_{ij}u(\pi_{ij})$$

and the optimal alternative  $k$  is found by:

$$k \in \text{Max}_{i \in [1, n]} \{EU_i\}$$

In this decision approach, the function  $u(\cdot)$ , stands for the investor's psychology. For example, we might construe that  $u'(\cdot) > 0$  implies *greed*,  $u''(\cdot) < 0$  implies *fear*, while *risk tolerance and prudence* are implied by the signs of the third derivative  $u'''(\cdot) > 0$  and  $u'''(\cdot) < 0$  respectively. Given a probability distribution for rewards,  $P(R)$ , the basic assumptions regarding continuous utility functions are that alternative rewards:

- (1) can be compared (comparability).
- (2) can be ranked such that preferred alternatives have greater utility.
- (3) have strong independence.
- (4) have transitive preferences (transitivity).
- (5) are indifferent if their utilities are equal.

### 3.2.1 Risk aversion

Expected utility provides an investor preference for uncertain payoffs, expressing thereby his attitude toward the risk associated with such payoffs. Three attitudes are defined: (1) risk aversion (2) risk loving and (3) risk neutrality. Risk aversion expresses a risk-avoidance preference and thus a preference for more conservative gambles. For example, a risk-averse investor may be willing to pay a premium to reduce risk. A risk lover would rather enjoy the gamble than an investment risk provides. Finally, risk neutrality implies that rewards are valued at their objective value by the expectation criterion (EMV). In other words, the investor would be oblivious to risk. For risk-averse investors, the desire for greater rewards with smaller probabilities will decrease (due to the increased risk associated with such rewards); such an attitude will correspond to a negative second derivative of the utility function or equivalently to an assumption of concavity, as we shall see below. And, vice versa, for a risk loving decision-maker the second derivative of the utility function will be positive. To characterize quantitatively a risk attitude, two approaches are used:

- Risk aversion directly relates to the risk premium, expressed by the difference between the expected value of a decision and its certainty (riskless) equivalent reward.
- Risk aversion is expressed by a decreasing preference for an increased risk, while maintaining a mean preserving spread.

These two definitions are equivalent for concave utility functions, as we shall see below.

#### Certainty equivalence and risk premium

Assume an uncertain reward  $\tilde{R}$  whose expected utility is  $E(u(\tilde{R}))$ . Its equivalent sure amount of money, given by the expected utility of that amount, is called the *certainty equivalent* which we shall denote here by  $\bar{R}$  and is given by

$$u(\bar{R}) = E(u(\tilde{R})) \quad \text{and} \quad \bar{R} = u^{-1}\{E[u(\tilde{R})]\}$$

Note that the certainty equivalent is not equal to the expected value  $\hat{R} = E(\tilde{R})$  for it embodies as well the cost of risk associated with the uncertain prospect valued by its expected utility. The difference  $\rho = \hat{R} - \bar{R}$ , expresses the *risk premium* a decision maker would be willing to pay for an outcome that provides for sure the expected return compared to the certainty equivalent. It can be null, positive or negative. In other words, the risk premium is:

$$\text{Risk premium } (\rho) = \text{Expected return } (\hat{R}) - \text{The certainty equivalent } (\bar{R})$$

An alternative representation of the risk premium can be reached by valuing the expected utility of the random payoff:  $\tilde{R} = \hat{R} + \tilde{\varepsilon}$  where  $E(\tilde{\varepsilon}) = 0$ ,  $\text{var}(\tilde{\varepsilon}) = \sigma^2$  and  $\sigma^2$  denotes the payoff spread. In this case, note that a Taylor series expansion

around the mean return yields:

$$Eu(\tilde{R}) = Eu(\hat{R} + \tilde{\varepsilon}) = u(\hat{R}) + \sigma^2 \frac{u''(\hat{R})}{2}$$

Similarly, a first-order Taylor series expansion of the certainty equivalent utility around the mean return (since there are no uncertain elements associated with it) yields:

$$u(\bar{R}) = u(\hat{R} - \rho) = u(\hat{R}) - \rho u'(\hat{R})$$

Equating these two equations, we obtain the risk premium calculated earlier but expressed in terms of the derivatives of the utility function and the return variance, or:

$$\rho = -\frac{1}{2}\sigma^2 \frac{u''(\hat{R})}{u'(\hat{R})}$$

This risk premium can be used as well to define the index of risk behaviour suggested by Arrow and Pratt. In particular, Pratt defines an *index of absolute risk aversion* expressing the quantity by which a fair bet must be altered by a risk-averse decision maker in order to be indifferent between accepting and rejecting the bet. It is given by:

$$\rho_a(\tilde{R}) = \frac{\rho}{\sigma^2/2} = -\frac{u''(\hat{R})}{u'(\hat{R})}$$

### Prudence and robustness

When a decision-maker's expected utility is not (or is mildly) sensitive to other sources of risk, we may state that the expected utility is 'robust' or expresses a prudent attitude by the decision-maker. A prudent investor, for example, who adopts a given utility function to reach an investment decision, expresses both his desire for returns and the prudence he hopes to assume in obtaining these returns, based on the functional form of the utility function he chooses. Thus, an investor with a precautionary (prudence) motive will tend to save more to hedge against the uncertainty that arises from additional sources of risk not accounted for by the expected utility of uncertain returns. This notion of prudence was first defined by Kimball (1990) and Eeckhoudt and Kimball (1991) and is associated with the optimal utility level (measured by the relative marginal utilities invariance), which is, or could be, perturbed by other sources of risk. Explicitly, say that  $(w, \tilde{R})$  is the wealth of a person and the random payoff which results from some investment. If we use the expected marginal utility, then at the optimum investment decision:

$$\begin{aligned} Eu'(w + \tilde{R}) &> u'(w) && \text{if } u' \text{ is convex} \\ Eu'(w + \tilde{R}) &< u'(w) && \text{if } u' \text{ is concave} \end{aligned}$$

The risk premium  $\psi$  that the investor pays for 'prudence' is thus the amount of money required to maintain the marginal utility for sure at its optimal wealth

level. Or:

$$u'(w - \psi) = Eu'(w + \tilde{R}) \quad \text{and} \quad \psi = w - u'^{-1}[Eu'(w + \tilde{R})]$$

Proceeding as before (by using a first term Taylor series approximation on the marginal utility), we find that:

$$\psi = \frac{1}{2} \text{var}(\tilde{R}) \left[ -\frac{u'''(w)}{u''(w)} \right]$$

The square bracket term is called the *degree of absolute prudence*. For a risk-averse decision maker, the utility second-order derivative is negative ( $u' \leq 0$ ) and therefore prudence will be positive (negative) if the third derivative  $u'''$  is positive (negative). Further, Kimball also shows that if the risk premium is positive and decreases with wealth  $w$ , then  $\psi > \pi$ . As a result,  $\psi - \pi$  is a premium an investor would pay to render the expected utility of an investment invariant under other sources of risks.

The terms expected utility, certainty equivalent, risk premium, Arrow–Pratt index of risk aversion and prudence are used profusely in insurance, economics and financial applications, as we shall see later on.

### 3.2.2 Expected utility bounds

In many instances, calculating the expected utility can be difficult and therefore bounds on the expected utility can be useful, providing a first approximation to the expected utility. For risk-averse investors with utility function  $u(\cdot)$  and  $u''(\cdot) \leq 0$ , the expected utility has a bound from above, known as *Jensen's inequality*. It is given by:

$$\begin{aligned} Eu(\tilde{R}) &\leq u(\hat{R}) \quad \text{when} \quad u''(\cdot) \leq 0 \\ Eu(\tilde{R}) &\geq u(\hat{R}) \quad \text{when} \quad u''(\cdot) \geq 0 \end{aligned}$$

and vice versa when it is the utility function of a risk-loving investor (i.e.  $u''(\cdot) \geq 0$ ). When rewards have known mean and known variance however, Willasen (1981, 1990) has shown that for risk-averse decision-makers, the expected utility can be bounded from below as well. In this case, we can bound the expected utility above and below by:

$$u(\hat{R}) \geq Eu(\tilde{R}) \geq \hat{R}^2 u(\alpha_2 / \hat{R}) / \alpha_2; \quad \alpha_2 = E(\tilde{R}^2)$$

The first bound is, of course, Jensen's inequality, while the second inequality provides a best lower bound. It is possible to improve on this estimate by using the best upper and lower Tchebycheff bounds on expected utility (Willasen, 1990). This inequality is particularly useful when we interpret and compare the effects of uncertainty on the choice of financial decisions, as we shall see in the example below. Further, it is also possible to replace these bounds by polynomials such that:

$$Eu(\tilde{R}) \leq EA(\tilde{R}); \quad Eu(\tilde{R}) \geq EB(\tilde{R})$$

where  $A(\cdot)$  and  $B(\cdot)$  are polynomials of the third degree. To do so, second- and third-order Taylor series approximations are taken for the utility functions (using thereby the decision-makers' prudence). For example, consider the following portfolio prospect with a mean return of  $\hat{R}$  and a variance  $\sigma^2$ . Say that mean returns are also a function of the variance, expressing the return-risk substitution, with:

$$\hat{R} = \hat{R}(\sigma), \partial \hat{R}(\sigma) / \partial \sigma >, \hat{R}(0) = R_f$$

where  $R_f$  denotes the riskless rate of return. It means that the larger the returns uncertainty, the larger the required expected payoff. Using the Jensen and Willason inequalities, we have for any portfolio, the following bounds on the expected utility:

$$\frac{u(\hat{R}(\sigma)(1 + \nu))}{1 + \nu} \leq Eu(\tilde{R}) \leq u(\hat{R}(\sigma)); \nu = \frac{\sigma^2}{\hat{R}^2}; \quad E(\tilde{R}^2) = \hat{R}^2 + \sigma^2$$

Thus, lower and upper bounds of the portfolio expected utility can be constructed by maximizing (minimizing) the lower (upper) bounds over feasible  $(\hat{R}, \sigma)$  portfolios. Further, if we set  $\hat{R} = R_f + \lambda\sigma$  where  $\lambda$  is used as a measure for the price of risk (measured by the return standard deviation and as we shall see subsequently), we have equivalently the following bounds:

$$\frac{u((R_f + \lambda\sigma)(1 + \nu))}{1 + \nu} \leq Eu(\tilde{R}) \leq u(R_f + \lambda\sigma)$$

The definition of an appropriate utility function is in general difficult. For this reason, other means are often used to express the desirability of certain outcomes. For example, some use targets, expressing the desire to maintain a given level of cash, deviations from which induce a dis-utility. Similarly, constraints (as they are defined by specific regulation) as well as probability constraints can also be used to express a behavioural attitude towards outcomes and risks. Such an approach has recently been found popular in financial circles that use 'value at risk' (VaR) as an efficiency criterion (see Chapter 10 in particular). Such assumptions regarding decision-makers' preferences are often used when we deal with practical problems.

### 3.2.3 Some utility functions

A utility function is selected because it represents the objective of an investor faced with uncertain payoffs and his attitude towards risk. It can also be selected for its analytical convenience. In general, such a selection is difficult and has therefore been one of the essential reasons in practice for seeking alternative approaches to decision making under uncertainty. Below we consider a number of analytical utility functions often used in theoretical and practical applications.

(1) *The exponential utility function:*  $u(w) = 1 - e^{-aw}$ ,  $a > 0$  is a concave function. For this function,  $u'(w) = a e^{-aw} > 0$ ,  $u''(w) = -a^2 e^{-aw} < 0$  while the index of absolute risk aversion  $R_A$  is constant and given by:  $R_A(w)$

$= -u''/u' = a > 0$ . Further  $u'''(w) = a^3 e^{-aw} > 0$  and therefore the degree of prudence is  $a$  while the prudence premium is,  $\psi = \frac{1}{2}a \text{ var}(\tilde{R})$ .

(2) *The logarithmic utility function:*  $u(w) = \log(\beta + \gamma w)$ , with  $\beta > 0, \gamma > 0$  is strictly increasing and strictly concave and has a strictly decreasing absolute risk aversion. Note that,  $u'(w) = \gamma/(\beta + \gamma w) > 0, u''(w) = -\gamma^2/(\beta + \gamma w)^2 < 0$  while,  $R_A(w) = \gamma/(\beta + \gamma w) = u'(w)$  which is decreasing in wealth.

(3) *The quadratic utility function:*  $u(w) = w - \rho w^2$  is a concave function for all  $\rho \geq 0$  since  $u' = 1 - 2\rho w, u' \geq 0 \rightarrow w \leq 1/2\rho$  and  $u'' = -2\rho \leq 0$ . As a result, the Arrow-Pratt index of absolute risk aversion is

$$R_A(w) = -\frac{u''[E(w)]}{u'[E(w)]} = \frac{2\rho}{1 - 2\rho w}$$

and the prudence is null (since the third derivative is null).

(4) *The cubic utility function:*  $u(w) = w^3 - 2kw^2 + (k^2 + g^2)w, k^2 > 3g^2$  is strictly increasing and strictly concave and has a decreasing absolute risk aversion if  $0 \leq w \leq \frac{2}{3}k - \frac{1}{2}\sqrt{k^2 - 3g^2}$ .

(5) *The power utility function:*  $u(w) = (w - \delta)^\beta, 0 < \beta < 1$  is strictly increasing and has a strictly absolute risk aversion on  $[\delta, \infty)$  since:  $u' = \beta(w - \delta)^{\beta-1}$ , and  $u'' = -(1 - \beta)\beta(w - \delta)^{\beta-2}$ . The risk aversion index is thus,  $R_A(w) = -(1 - \beta)(w - \delta)^{-1}$ .

(6) *The HARA (hyperbolic absolute risk aversion)* has a utility function given by:

$$u(w) = \frac{1 - \gamma}{\gamma} \left[ \frac{aw}{1 - \gamma} + b \right]^\gamma$$

while its first and second derivatives as well as its index of absolute risk aversion are given by:

$$u' = a \left[ \frac{aw}{1 - \gamma} + b \right]^{\gamma-1} > 0, u'' = -a^2 \left[ \frac{aw}{1 - \gamma} + b \right]^{\gamma-2} < 0;$$

$$R_a(w) = -\frac{u''}{u'} = \frac{a}{b + aw/(1 - \gamma)} > 0$$

This utility function includes a number of special cases. In particular, when  $\gamma$  tends to one, we obtain the logarithmic utility.

### 3.2.4 Risk sharing

Two firms sign an agreement for a joint venture. A group of small firms organize a cooperative for marketing their products. The major aerospace companies in the US west coast set up a major research facility for deep space travel. A group of 70 leading firms form a captive insurance firm in the Bahamas to insure their managers against kidnappings, and so on. These are all instances of risk sharing. Technically, when we combine together a number of (independent) participants and split among them a potential loss or gain, the resulting variance of the loss or gain for each of the participants will be smaller. Assuming that this variance

is an indicator of the ‘risk’, and if decision makers are assumed to be risk averse, then the more partners in the venture the smaller the individual risk sustained by each partner. Such arguments underly the foundations of insurance firms (that create the means for risk sharing), of major corporations based on numerous shareholders etc. Assuming that our preference is well defined by a utility function  $U(\cdot)$ , how would we know if it is worthwhile to share risk? Say that the net benefits (profits less costs) of a venture is  $\$X$  whose probability distribution is  $p(\tilde{X})$ . If we do nothing, nothing is gained and nothing is lost and therefore the ‘value’ of doing nothing is  $U(0)$ . The venture with its  $n$  participants, however, will have an expected utility  $EU(\tilde{X}/n)$ . Thus, if sharing is worthwhile the expected utility of the venture ought to be greater than the utility of doing nothing! Or,  $EU(\tilde{X}/n) > U(0)$ .

### Problems

- (1) Formulate the problem of selecting the optimal size of a risk-sharing pool.
- (2) How much does a member of the pool benefit from participating in sharing.

## 3.3 INSURANCE, RISK MANAGEMENT AND EXPECTED UTILITY

How much would it be worth paying for car insurance (assuming that there is such a choice)? This simple question highlights an essential insurance problem. If we are fully insured and the premium is  $\$\pi$ , then the expected utility is, for sure,  $U(w - \pi)$  where  $w$  is our initial wealth. If we self-insure for a risk whose probability distribution is  $p(\tilde{X})$ , then using the expected utility theory paradigm, we should be willing to pay a premium  $\pi$  as long as  $U(w - \pi) > EU(w - \tilde{X})$ . In fact the largest premium we would be willing to pay solves the equation above, or

$$\pi^* = w - U^{-1}(EU(w - \tilde{X}))$$

Thus, if the utility function is known, we can find out the premium  $\pi^*$  above which we would choose to self-insure.

### Problem

For an exponential, HARA and logarithmic utility function, what is the maximal premium an individual will be willing to pay for insurance?

#### 3.3.1 Insurance and premium payments

Insurance risk is not reduced but is transferred from an individual to an insurance firm that extracts a payment in return called the premium and profits from it by investing the premium and by risk reducing aggregation. In other words, it is the difference in risk attitudes of the insurer and the insured, as well as the price insured, and insurers are willing to pay for that to create an opportunity for the insurance business.

Say that  $\tilde{X}$  is a risk to insure (a random variable) whose density function is  $F(\tilde{X})$ . Insurance firms, typically, seek some rule to calculate the premium they ought to charge policyholders. In other words, they seek a ‘rule’  $\Upsilon$  such that a premium can be calculated by:

$$P = \Upsilon(F(\tilde{X}))$$

Although there are alternative ways to construct this rule, the more prominent ones are based on the application of the expected utility paradigm and traditionally based on a factor loading the mean risk insured. The expected utility approach seeks a ‘fair’ premium  $P$  which increases the firm expected utility, or:

$$U(W) \leq EU(W + P - \tilde{X})$$

where  $W$  is the insurance firm’s capital. The loading factor approach seeks, however, to determine a loading parameter  $\lambda$  providing the premium to apply to the insured and calculated by  $P/n = (1 + \delta)E(\tilde{x})$ , where  $\tilde{x}$  denotes the individual risk in a pool of  $n$  insured, i.e.  $\tilde{x} = \tilde{X}/n$  and  $P/n$  is an individual premium share. For the insured, whose utility function is  $u(\cdot)$  and whose initial wealth is  $w$ , the expected utility of insurance ought to be greater than the expected utility of self-insurance. As a result, a premium  $P$  is feasible if the expected utilities of both the insurer and the insured are larger with insurance, or:

$$u(w - P/n) \geq Eu(w - \tilde{x}_i), \tilde{X} = \sum_{i=1}^n \tilde{x}_i; U(W) \leq EU(W + P - \tilde{X})$$

Note that in this notation, the individual risk is written as  $\tilde{x}_i$  which is assumed to be identically and independently distributed for all members of the insurance pool. Of course, since an insurance firm issues many policies, assumed independent, it will profit from risk aggregation. However, if risks are correlated, the variance of  $\tilde{X}$  will be much greater, prohibiting in some cases the insurance firm’s ability or willingness to insure (as is the case in natural disaster, agricultural and weather related insurance).

Insurance ‘problems’ arise when it is necessary to resolve the existing disparities between the insured and the insurer, which involves preferences and insurance terms that are specific to both the individual and the firm. These lead to extremely rich topics for study, including the important effects of moral hazard, adverse selection resulting from information asymmetry which will be studied subsequently, risk correlation, rare events with substantive damages, insurance against human-inspired terrorists acts etc.

Risk sharing, risk transfer, reinsurance and other techniques of risk management are often used to spread risk and reduce its economic cost. For example, let  $\tilde{x}$  be the insured risk; the general form of reinsurance schemes associated with an insurer (I), an insured (i) and a reinsurer (r) and consisting in sharing risk can be

written as follows:

$$\begin{aligned} \text{Insured: } R_i(\tilde{x} | a, q) &= \begin{cases} \tilde{x} & \tilde{x} \leq a \\ (1-q)\tilde{x} & \tilde{x} \geq a \end{cases} \\ \text{Insurer: } R_I(\tilde{x} | a, b, c, q) &= \begin{cases} 0 & \tilde{x} \leq a \\ q(\tilde{x} - a) & a < \tilde{x} \leq b \\ c & b > \tilde{x} \end{cases} \\ \text{Reinsurer: } R_r(\tilde{x} | b, c, q) &= \begin{cases} 0 & \tilde{x} \leq b \\ q(b - \tilde{x}) - c & \tilde{x} \geq b \end{cases} \end{aligned}$$

Here, if a risk materializes and it is smaller than ‘ $a$ ’, then no payment is made by the insurance firm while the insured will be self-insured up to this amount. When the risk is between the lower level ‘ $a$ ’ and the upper one ‘ $b$ ’, then only a proportion  $q$  is paid where  $1-q$  is a co-participation rate assumed by the insured. Finally, when the risk is larger than ‘ $b$ ’, then only  $c$  is paid by the insurer while the remaining part  $\tilde{x} - c$  is paid by a reinsurer. In particular for a proportional risk scheme we have  $R(\tilde{x}) = q\tilde{x}$  while for an excess-loss reinsurance scheme we have:

$$R_I(\tilde{x} | a) = \begin{cases} 0 & \tilde{x} \leq a \\ \tilde{x} - a & \tilde{x} > a \end{cases}$$

where  $a$  is a deductible specified by the insurance contract. A reinsurance scheme is thus economically viable if the increase in utility is larger than the premium  $P_r$  to be paid to the reinsurer by the insurance firm. In other words, for utility functions  $u_I(\cdot)$ ,  $u_i(\cdot)$ ,  $u_r(\cdot)$  for the individual, the insurance and the reinsurance firms with premium payments:  $P_i$ ,  $P_I$ ,  $P_r$ , the following conditions must be held:

$$u_i(w - R_i(\tilde{x} | a, q) - P_i) \geq Eu_i(w - \tilde{x}) \quad (\text{the individual condition})$$

$$u_I(W) \leq Eu_I(W + P_i - P_I - R_I(\tilde{x} | a, b, c, q)) \quad (\text{the insurance firm condition})$$

$$u_r(W_r) \leq Eu_r(W_r + P_I - P_r - R_r(\tilde{x} | b, c, q)) \quad (\text{the reinsurer condition})$$

Other rules for premium calculation have also been suggested in the insurance literature. For example, some say that in insurance ‘you get what you give’. In this sense, the premium payments collected from an insured should equal what he has claimed plus some small amounts to cover administrative expenses. These issues are in general much more complex because the insurer benefits from risk aggregation over the many policies he insures, a concept that is equivalent to portfolio risk diversification. In other words, if the insurance firm is large enough it might be justified in using a small (risk-free) discount rate in valuing its cash flows, compared to an individual insured, sensitive with the uncertain losses associated with the risk insured. For this reason, the determination of the loading rate is often a questionable parameter in premium determination. Recent research has greatly improved the determination of insurance premiums by indexing insurance risk to market risk and using derivative markets (such as options) to value insurance contracts (and thereby the cost of insurance or premium).

### 3.4 CRITIQUES OF EXPECTED UTILITY THEORY

Theory and practice do not always concur when we use expected utility theory. There are many reasons for such a statement. Are decision-makers irrational? Are they careless? Are they uninformed or clueless? Do they lack the proper incentives to reach a rational decision. Of course, the axioms of rationality that underlie expected utility theory may be violated. Empirical and psychological research has sought to test the real premises of decision-makers under uncertainty. To assess potential violations, we consider a number of cases. Consider first the example below called the St Petersburg Paradox that has motivated the development of the utility approach.

#### 3.4.1 Bernoulli, Buffon, Cramer and Feller

Daniel Bernoulli in the early 1700s suggested a problem whose solution was not considered acceptable in practice, albeit it seemed to be appropriate from a theoretical viewpoint. This is called the St Petersburg Paradox. The paradox is framed in a tossing game stating how much one would be willing to pay for a game where a fair coin is thrown until it falls ‘heads’. If it occurs at the  $r$ th throw, the player receives  $2^r$  dollars from the bank. Thus, the gain doubles at each throw. In an expected sense, the probability of obtaining ‘heads’ at the  $k$ th throw is  $1/2^k$ , since the pay-out is also equal to  $2^k$ , the expected value of the game is:

$$\sum_{k=1}^{\infty} (1/2^k)2^k = 1 + 1 + \dots = \infty$$

Thus, the fair amount to pay to play this game is infinite, which clearly does not reflect the decision makers’ behaviour. Bernoulli thus suggested a logarithmic utility function whose expected utility:

$$Eu(x) = \sum_{i=1}^{\infty} p(i)u(2^{i-1}) = \sum_{i=1}^{\infty} \left(\frac{1}{2^i}\right) \log(2^{i-1}) = \sum_{i=1}^{\infty} \left(\frac{i-1}{2^i}\right) \log(2)$$

Since

$$\sum_{i=1}^{\infty} \left(\frac{1}{2^i}\right) = 1 \quad \text{and} \quad \sum_{i=1}^{\infty} \left(\frac{i-1}{2^i}\right) = 1$$

The expected utility of the game equals  $Eu(x) = \log(2)$ .

Mathematicians such as Buffon, Cramer, Feller and others have attempted to provide a solution that would seem to be appropriate. Buffon and Cramer suggest that the game be limited (in the sense that the bank has a limited amount of money and, therefore, it can only pay a limited amount). Say that the bank has only a million dollars. In this case, we will have the following amounts,

$$\sum_{k=1}^{19} (1/2^k)2^k + 10^6 \sum_{k=20}^{\infty} (1/2^k) = 1 + 1 + \dots + 1 + 1.19 \approx 21$$

Therefore, the fair amount to play this game is 21 dollars only. Any larger amount would be favourable to the bank.

Gabriel Cramer, on the other hand, suggested a square root utility function. Then for the St Petersburg game, we have:

$$\begin{aligned} Eu(x) &= \sum_{i=1}^{\infty} p(i)u(i) = \sum_{i=1}^{\infty} \left(\frac{1}{2^i}\right) \sqrt{2^i - 1} \\ &= \frac{1}{2} + \frac{\sqrt{2}}{2^2} + \frac{(\sqrt{2})^2}{2^2} + \dots + \frac{(\sqrt{2})^j}{2^{j+1}} + \dots = \frac{1}{2} \frac{1}{1 - \sqrt{2}/2} = \frac{1}{2 - \sqrt{2}} \end{aligned}$$

And therefore, for Cramer, the value of the game is  $1/(2 - \sqrt{2})$ . Feller suggests another approach however, seeking a mechanism for the gains and payments to be equivalent in the long run. In other words, a lottery will be fair if:

$$\begin{aligned} \frac{\text{Accumulated gains}}{\text{Accumulated fees}} &= \frac{N_n}{R_n} \rightarrow 1 \text{ as } n \rightarrow \infty \\ \text{or } P \left\{ \left| \frac{N_n}{R_n} - 1 \right| < \varepsilon \right\} &\rightarrow \infty \text{ as } n \rightarrow \infty \end{aligned}$$

Feller noted that the game is fair if  $R_n = n \log_2(n)$ . Thus if the accumulated entrance fee to the game is proportional to the number of games, it will not be a fair game.

### 3.4.2 Allais Paradox

The strongest attack on expected utility theory can be found in Allais' Paradox, which doubts the strong independence assumption needed for consistent choice in expected utility. Allais proved that the assumption of linearity in probabilities applied in calculating the expected utility is often doubtful in practice. Explicitly, the independence axiom also called the 'sure-thing' principle asserts that two alternatives that have a common outcome under a particular state of nature should imply that ordering should be independent of the value of their common outcome. This is not always the case and counter examples abound, in particular due to Allais.

For example, let us confront people with two lotteries. First, we have to pick one of the two gambles given by  $(p_1, p_2)$  below. The first gamble consists of \$100 000 for sure (probability 1) while the other is \$5 000 000 with probability 0.1, 1 million with probability 0.89 and nothing with probability 0.01 as stated below.

$$p_1 \Rightarrow \{ 1 \quad 100\,000 \quad \text{and} \quad p_2 \Rightarrow \begin{cases} 0.1 & 5\,000\,000 \\ 0.89 & 1\,000\,000 \\ 0.01 & 0 \end{cases}$$

A second set of gambles  $(p_3, p_4)$  consists of:

$$p_3 \Rightarrow \begin{cases} 0.1 & 5\,000\,000 \\ 0.9 & 0 \end{cases} \quad \text{and} \quad p_4 \Rightarrow \begin{cases} 0.11 & 1\,000\,000 \\ 0.89 & 0 \end{cases}$$

Confronted with choosing between  $p_1$  and  $p_2$ , people chose  $p_1$  while confronting  $p_1$ ,  $p_3$  and  $p_4$ , people preferred  $p_3$  which is in contradiction with the strong independence axiom of utility theory. In other words, if gamble 1 is selected over gamble 2, while in presenting people with gambles 1, 3 and 4 results in their selecting 3 over 1, there is necessarily a contradiction, since if we were to compare gambles 3 and 2, clearly, 3 is not as good as 2. This contradiction means, therefore, that application of expected utility theory does not always represent investors' and decision-maker's psychology. Since then, a large number of studies have been done, seeking to bridge a gap between investors' psychology and the concepts of utility theory. Some essential references include Kahnemann and Tverski (1979), Machina on anticipated utility (1982, 1987) and Quiggin (1985) as well as many other others. In these approaches, the expected utility framework is 'extended' by stating that an uncertain prospect can be measured by an 'expected utility'  $u(\cdot)$  interpreted either as the choice of a utility function (as was the case in traditional expected utility) or by a preferred probability distribution ( $P$ ) (or a function  $g(P)$  assumed over the probability distribution). Then, the probabilities used to calculate the expected utility would be 'subjective' estimates, or beliefs, about the probabilities of returns, imbedding 'something else' above and beyond the objective assessment of uncertain prospects. Thus, the objective index used to value the relative desirability of uncertain prospects is also a function of the model used for probabilities  $P(\cdot)$ . This is in contrast to the utility function, expressing a behaviour imbedded in the choice of the function  $u(\cdot)$  only, which stands solely for the investor's psychology, as we saw above. For example, what if we were to determine probabilities  $P^*(\cdot)$  such that the price of random prospects  $\tilde{R}$  could be uniquely defined by the following expected value?

$$\pi = \int_{\mathfrak{R}} \tilde{R} dP^*(\tilde{R})$$

In this case, once such a well-defined transformation of these probabilities is reached, all uncertain prospects may be valued uniquely, thereby simplifying greatly the problem of financial valuation of risk assets such as stocks, default bonds and the like. This approach, defined in terms of economic exchange market mechanisms (albeit subject to specific assumptions regarding markets and individual behaviours), underlies the modern theories of finance and 'risk-neutral pricing'. This is also an essential topic of our study in subsequent chapters.

### 3.5 EXPECTED UTILITY AND FINANCE

Finance provides many opportunities for applications of expected utility. For example, portfolio management consists essentially of selecting an allocation strategy among  $n$  competing alternatives (stocks, bonds, etc.), each yielding an uncertain payoff. Each stock purchase is an alternative which can lead to a (speculative) profit or loss with various (known or unknown) probabilities. When selecting several stocks and bonds to invest in, balancing the potential gains with the risks of losing part or all of the investment, the investor in effect constructs a

portfolio with a risk/reward profile which is preferred in an expected utility sense. Similarly, to evaluate projects, contracts, investments in real estate, futures and forward contracts, etc. financial approaches have been devised based directly on (or inspired by) expected utility. Below, we shall review some traditional techniques for valuing cash flows and thereby introduce essential notions of financial decision making using expected utility. Typical models include the CAPM (Capital Asset Pricing Model) as well the SDF (Stochastic Discount Factor) approach.

### 3.5.1 Traditional valuation

Finance values money and cash flow, the quantity of it, the timing of it and the risk associated with it. A number of techniques and approaches that are subjective – defined usually by corporate financial officers or imposed by managerial requirements – have traditionally been used. For example, let  $C_0, C_1, C_2, \dots, C_n$  be a prospective cash flow in periods  $i = 1, 2, \dots, n$ . Such a cash flow may be known for sure, may be random, payments may be delayed unexpectedly, defaulted etc. To value these flows, various techniques can be used, each assuming a body of presumptions regarding the cash flow and its characteristics. Below, we consider first a number of ‘traditional’ approaches including ‘the payback period’, ‘the accounting rate of return’ and the traditional ‘NPV’ (Net Present Value).

#### Payback period

The payback period is the number of years required for an investment to be recovered by a prospective cash flow. CFOs usually specify the number of years needed for recovery. For example, if 4 years is the specified time to recover an investment, then any project with a prospective cash flow of recovery less or equal to 4 years is considered acceptable. While, any investment project that does not meet this requirement is rejected. This is a simple and an arbitrary approach, although in many instances it is effective in providing a first cut approach to multiple investment opportunities. For example, say that we have an investment of \$100 000 with a return cash flow (yearly and cumulatively) given by the following table:

Year	1	2	3	4	5	6	7	8	9
Return	-5	5	10	20	40	30	20	10	10
Cumulative	-5	0	10	30	70	100	120	130	140

Thus, only after the sixth year is the investment is recovered. If management specifies a period of 4 years payback, then of course the investment will be rejected.

#### The accounting rate of return (ARR)

The ARR is a ratio of average profit after depreciation and average investment book value. There are, of course, numerous accounting procedures in calculating these terms, making this approach as arbitrary as the payback period.

A decision may be made to specify a required ARR. Any such ratio 'better' than the ARR selected would imply that the investment project is accepted.

### The internal rate of return (IRR)

The net present value (NPV) approach uses a discount rate  $R$  for the time value of money, usually called the rate of return. This discount rate need not be the risk-free rate (even if future cash flows are known for sure). Instead, they are specified by CFOs and used to provide a firm's valuation of the prospective cash flow. Typically, it consists of three components: the real interest rate, the inflation rate and a component adjusting for investment risk.

Discount (Nominal) Rate = (Real + Inflation + Risk Compensating) Rates

Each of these rates is difficult to assess and, therefore, much of finance theory and practice seeks to calculate these rates. The NPV of a cash flow over  $n$  periods is given by:

$$NPV = C_0 + \frac{C_1}{1+R} + \frac{C_2}{(1+R)^2} + \frac{C_3}{(1+R)^3} + \dots + \frac{C_n}{(1+R)^n}$$

where  $R$  is the discount rate applied to value the cash flow. The IRR, however, is found by finding the rate that renders the NPV null ( $NPV = 0$ ), or by solving for  $R^*$  is:

$$0 = C_0 + \frac{C_1}{1+R^*} + \frac{C_2}{(1+R^*)^2} + \frac{C_3}{(1+R^*)^3} + \dots + \frac{C_n}{(1+R^*)^n}$$

Each project may therefore have its own IRR which in turn can be used to rank alternative investment projects. In fact, one of the essential problems CFOs must deal with is selecting an IRR to enable them to select/accept investment alternatives. If the IRR is larger than a strategic discount factor, specified by the CFO, then investment is deemed economical and therefore can be made. There are, of course, many variants of the IRR, such as the FIE (fixed equivalent rate of return) which assumes a fixed IRR with funds, generated by the investment, reinvested at the IRR. Such an assumption is not always realistic, however, tending to overvalue investment projects. Additional approaches based on 'risk analysis' and the market valuation of risk have therefore been devised, seeking to evaluate the probabilities of uncertain costs and uncertain payoffs (and thereby uncertain cash flows) of the investment at hand. In fact, the most significant attempt of fundamental finance has been to devise a mechanism that takes the 'arbitrariness' out of investment valuation by letting the market be the mechanism to value risk (i.e. by balancing supply and demand for risky assets at an equilibrium price for risky assets). We shall turn to this important approach subsequently.

### Net present value (NPV) and random cash flows

When cash flows are random we can use the expected utility of the random quantities to calculate the NPV. We consider first a simple two-period example. Let  $\tilde{C}$  be an uncertain cash flow whose expected utility is  $Eu(\tilde{C})$ . Its certainty equivalent, is  $CE$  where  $Eu(\tilde{C}) = u(CE)$  or  $CE = u^{-1}(Eu(\tilde{C}))$ . Since  $CE$  is a sure

quantity, the discount rate applied to value the reception of such a quantity for one period hence is the risk free rate  $R_f$ . In other words, for a one-period model the  $PV$  is:

$$PV = \frac{CE}{1 + R_f} = \frac{u^{-1}(Eu(\tilde{C}))}{1 + R_f} = \frac{E(\tilde{C}) - P}{1 + R_f}$$

where  $P$  is the risk premium. Equivalently we can calculate the  $PV$  by using the expected cash flow but discounted at a rate  $k$  (incorporating the risk inherent in the cash flow). Namely,

$$PV = \frac{E(\tilde{C}) - P}{1 + R_f} = \frac{E(\tilde{C})}{1 + k}$$

As a result, we see that the risk-free rate and the risk premium combine to determine the risk adjusted rate as follows:

$$1 - \frac{P}{E(\tilde{C})} = \frac{1 + R_f}{1 + k} \rightarrow k - R_f = (1 + k) \frac{P}{E(\tilde{C})} \quad \text{or} \quad k = \frac{1 + R_f}{1 - P/E(\tilde{C})} - 1$$

In particular note that  $k - R_f$  defines the ‘excess discount rate’. It is the rate of return needed to compensate for the uncertainty in the cash flow  $\tilde{C}$ . To calculate the appropriate discount rate to apply, a concept of equilibrium reflecting investors’ homogeneity is introduced. This is also called ‘the capital market equilibrium’ which underlies the CAPM as we shall see below. Over multiple periods of time, calculations of the  $PV$  for an uncertain future cash stream, yields similarly:

$$PV = C_0 + \frac{E(\tilde{C}_1)}{(1 + k)^1} + \frac{E(\tilde{C}_2)}{(1 + k)^2} + \frac{E(\tilde{C}_3)}{(1 + k)^3} + \dots$$

However, interest rates (and similarly, risk-free rates, risk premiums, etc.) may vary over time – reflecting the effects of time (also called the term structure) either in a known or unknown manner. Thus, discounting must reflect the discount adjustments to be applied to both uncertainties in the cash flow and the discount rate to apply because of the timing of payments associated with these cash flows. If the discount (interest) rates vary over time, and we recognize that each instant of time discounting accounts for both time and risk of future cash flows, the present value is then given by:

$$PV = C_0 + \frac{E(\tilde{C}_1)}{(1 + k_1)^1} + \frac{E(\tilde{C}_2)}{(1 + k_2)^2} + \frac{E(\tilde{C}_3)}{(1 + k_3)^3} + \dots = \sum_{i=0} \frac{E(\tilde{C}_i)}{(1 + k_i)^i}$$

where  $k_1, k_2, k_3, \dots, k_n \dots$  express the term structure of the risk-adjusted rates. Finally, note that if we use the certainty cash equivalents  $C_i$ , associated with a utility  $u(C_i) = Eu(\tilde{C}_i)$ , then:

$$PV = \sum_{i=0} \frac{C_i}{(1 + R_{f,i})^i}$$

where  $R_{f,i}$  is the risk-free term structure of interest rate for a discount over  $i$  periods. In Chapter 7, we shall discuss these issues in greater detail.

### 3.5.2 Individual investment and consumption

A number of issues in finance are stated in terms of optimal consumption problems and portfolio holdings. Say that an individual maximizes the expected utility of consumption, separable in time and state and is constrained by his wealth accumulation equation (the returns on savings and current wage income). Technically, assume that an individual investor has currently a certain amount of money invested in a portfolio consisting of  $N_0$  shares of a stock whose current price is  $p_0$  and a riskless investment in a bond whose current price is  $B_0$ . In addition, the investor has a wage income of  $s_0$ . Thus, current wealth is:

$$W_0 = N_0 p_0 + B_0 + s_0$$

Let  $c_0$  be a planned current consumption while the remaining part  $W_0 - c_0$  is reinvested in a portfolio consisting of  $N_1$  shares of a stock whose price is  $p_0$  and a bond whose current price is  $B_1$ . At the next time period, time '1', the investor consumes all available income. Disposable savings  $W_0 - c_0$  are thus invested in a portfolio whose current wealth is  $N_1 p_0 + B_1$ , or initially:

$$W_0 - c_0 = N_1 p_0 + B_1$$

At the end of the period, the investor's wealth is random due to a change in the stock price  $\Delta \tilde{p}$  and is wholly consumed, or:

$$\tilde{W}_1 = N_1(p_0 + \Delta \tilde{p}) + B_1(1 + R_f) \quad \text{and} \quad c_1 = \tilde{W}_1$$

Given the investor's utility function, there are three decisions to reach: how much to consume now, how many shares of stock to buy and how much to invest in bonds. The problem can be stated as the maximization of:

$$U = u_0(c_0) + \frac{1}{1 + R} E u_1(\tilde{W}_1)$$

with,  $u_0(\cdot)$  and  $u_1(\cdot)$  the utilities of the current and next (final) period consumption. An individual's preference is expressed here twice. First we use an individual discount rate  $R$  for the expected utility of consumption at retirement  $E u_1(\tilde{W}_1)$ . And, second, we have used the expected utility as a mechanism to express the effects of uncertainty on the value of such uncertain payments. Define a cash (certainty) equivalent to such expected utility by  $C_1$  or  $C_1 = u_1^{-1}(E u_1(\tilde{W}_1))$ . Since this is a 'certain cash equivalent', we can also write in terms of cash worth:

$$U = C_0 + \frac{C_1}{1 + R_f}, \quad C_0 = u_0^{-1}(u_0(c_0)) = c_0$$

Note that once we have used a certain cash amount we can use the risk-free rate  $R_f$  to discount that amount. As a result,

$$\frac{1 + R_f}{1 + R} = \frac{u_1^{-1}(E u_1(\tilde{W}_1))}{E u_1(\tilde{W}_1)}$$

which provides a relationship between the discounted expected utility and the risk-free discount rate for cash. Using the expected utility discount rate, we have:

$$U = u_0(N_0p_0 + B_0 + s_0 - N_1p_0 - B_1) + \frac{1}{1+R}Eu_1(N_1(p_0 + \Delta\tilde{p}) + B_1(1 + R_f))$$

A maximization of the current utility provides the investment strategy  $(N_1, B_1)$ , found by the solution of:

$$p_0 \frac{\partial u_0(c_0)}{\partial N_1} = \frac{1}{1+R}E\left(\frac{\partial u_1(c_1)}{\partial N_1}(p_0 + \Delta\tilde{p})\right), \quad \frac{\partial u_0}{\partial B_1} = E\left(\frac{\partial u_1}{\partial B_1}\right)$$

This portfolio allocation problem will be dealt with subsequently in a general manner but has already interesting implications. For example, if the utility of consumption is given by a logarithmic function,  $u_0(c) = \ln(c)$  and  $u_1(c) = \ln(c)$ , we have:

$$\frac{p_0}{c_0} = \frac{1}{1+R}E\left(\frac{p_0 + \Delta\tilde{p}}{c_1}\right) \quad \text{or} \quad \eta_0 = \frac{1}{1+R}E(\eta_1), \quad \eta_i = \frac{p_i}{c_i}, \quad i = 0, 1$$

meaning that the price per unit consumption  $\eta_i$  is an equilibrium whose value is the rate  $R$ . Further, the condition for the bond yields

$$\frac{\partial u_0(c_0)}{\partial c_0} = E\left(\frac{\partial u_1(c_1)}{\partial c_1}\right) \quad \text{or} \quad \frac{1}{c_0} = E\left(\frac{1}{c_1}\right)$$

It implies that the current marginal utility of consumption equals the next period's expected marginal consumption. The investment policy is thus a solution of the following two equations for  $(N_1, B_1)$  (where  $\tilde{W}_1 = N_1(p_0 + \Delta\tilde{p}) + B_1(1 + R_f)$ ):

$$\frac{p_0}{(N_0p_0 + B_0 + s_0 - N_1p_0 - B_1)} = \frac{1}{1+R}E\left(\frac{p_0 + \Delta\tilde{p}}{\tilde{W}_1}\right)$$

$$\frac{1}{N_0p_0 + B_0 + s_0 - N_1p_0 - B_1} = E\left(\frac{1}{\tilde{W}_1}\right)$$

or  $p_0RE(1/\tilde{W}_1) = E(\Delta\tilde{p}/\tilde{W}_1)$ . To solve this equation numerically, we still need to specify the probability of the stock price.

### Problem

Assume that

$$\Delta\tilde{p} = \begin{cases} Hp_0 & \text{w.p. } \pi \\ Lp_0 & \text{w.p. } 1 - \pi \end{cases}$$

where w.p. means with probability, and find the optimal portfolio. In particular, set  $\pi = 0.6, H = 0.3, L = -0.2$  and  $R_f = 0.1$ , then show that the optimal investment

policy is an all-bond investment. However, if  $H = 0.5$ , we have:

$$B_1^* = 0.085(N_0 p_0 + B_0 + s_0), N_1^* = -0.6444 \left( N_0 + \frac{B_0 + s_0}{p_0} \right)$$

### 3.5.3 Investment and the CAPM

Say that an investor has an initial wealth level  $W_0$ . Let  $\tilde{k}$  be a random rate of return of a portfolio with known mean and known variance given respectively by  $\hat{k}$ ,  $\sigma_k^2$  respectively. Say that part of the individual wealth,  $S_1$ , is invested in the risky asset while the remaining part  $B_1 = W_0 - S_1$  is invested in a non-risky asset whose rate of return is the risk-free rate  $R_f$ . The wealth one period hence is thus:

$$\tilde{W}_1 = B_1(1 + R_f) + S_1(1 + \tilde{k})$$

The demand for the risky asset is thus given by optimizing the expected utility function:

$$\text{Max}_{S_1 \geq 0} Eu[B_1(1 + R_f) + S_1(1 + \tilde{k})] \quad \text{or} \quad \text{Max}_{S_1 \geq 0} Eu[W_0(1 + R_f) + S_1(\tilde{k} - R_f)]$$

The first two derivatives conditions are:

$$\frac{d(Eu(\tilde{W}_1))}{dS_1} = E(u'(\tilde{W}_1)(\tilde{k} - R_f)) = 0 \quad \frac{d^2(Eu(\tilde{W}_1))}{d(S_1)^2} = E(u''(\tilde{W}_1)(\tilde{k} - R_f)^2) < 0$$

This is always satisfied when the investor is risk-averse. Consider the first-order condition, which we rewrite for convenience as follows:

$$E[u'(\tilde{W}_1)\tilde{k}] = E[u'(\tilde{W}_1)R_f]$$

By definition of the covariance, we have:

$$E(u'(\tilde{W}_1)\tilde{k}) = \hat{k}E(u'(\tilde{W}_1)) + \text{cov}(u'(\tilde{W}_1), \tilde{k})$$

Thus,

$$\hat{k}E(u'(\tilde{W}_1)) + \text{cov}(u'(\tilde{W}_1), \tilde{k}) = R_f E(u'(\tilde{W}_1))$$

Since the derivative of the expected utility is not null, we divide this expression by it and obtain:

$$\hat{k} + \frac{\text{cov}(u'(\tilde{W}_1), \tilde{k})}{E(u'(\tilde{W}_1))} = R_f$$

which clearly outlines the relationship between the expected returns of the risky and the non-risky asset and provides a classical result called the Capital Asset Pricing Model (CAPM). If we write the CAPM regression equation (to be seen below) by:

$$\hat{k} = R_f + \beta(R_m - R_f)$$

then, we can recuperate the beta factor often calculated for stocks and risky investments:

$$\beta = -\frac{\text{cov}(u'(\tilde{W}_1), \tilde{k})}{(R_m - R_f)E(u'(\tilde{W}_1))}$$

where  $R_m$  is the expected rate of return of a market portfolio or the stock market index. However, the beta found in the previous section implies:

$$\beta = \frac{\text{cov}(\tilde{R}_m, \tilde{k})}{\sigma_m^2} = -\frac{\text{cov}(u'(\tilde{W}_1), \tilde{k})}{(R_m - R_f)E(u'(\tilde{W}_1))}$$

and therefore:

$$\hat{k} - R_f = (R_m - R_f) \frac{\text{cov}(\tilde{R}_m, \tilde{k})}{\sigma_m^2} = -\frac{\text{cov}(u'(\tilde{W}_1), \tilde{k})}{E(u'(\tilde{W}_1))}$$

This sets a relationship between individuals' utility of wealth, the market mechanism and a statistical estimate of market parameters.

An equivalent approach consists in constructing a portfolio consisting of a proportional investment  $y_i = S_i/W_0$  in a risky asset  $i$  with a rate of return  $\tilde{k}_i$  while the remaining part is invested in a market index whose rate of return is  $\tilde{R}_m$  (rather than investing in a riskless bond). The rate of return of the portfolio is then  $\tilde{k}_p = y_i\tilde{k}_i + (1 - y_i)\tilde{R}_m$  with mean and variance:

$$\begin{aligned} \hat{k}_p &= y_i\hat{k}_i + (1 - y_i)R_m, \quad R_m = E\tilde{R}_m \\ \sigma_p^2 &= y_i^2\sigma_i^2 + (1 - y_i)^2\sigma_m^2 + 2y_i(1 - y_i)\sigma_{im}; \quad \sigma_{im} = \text{cov}(\tilde{k}_i, \tilde{k}_m) \end{aligned}$$

where variances  $\sigma^2$  are appropriately indexed according to the return variable they represent. The returns–risk substitution is found by calculating by chain differentiation:

$$\frac{d\hat{k}_p}{d\sigma_p} = \frac{d\hat{k}_p/dy_i}{d\sigma_p/dy_i}$$

where:

$$\frac{d\hat{k}_p}{dy_i} = \hat{k}_i - R_m; \quad \frac{d\sigma_p}{dy_i} = \frac{y_i(\sigma_i^2 + \sigma_m^2 - 2\sigma_{im}) + \sigma_{im} - \sigma_m^2}{\sigma_p}$$

A portfolio invested only on the market index (i.e.  $y_i = 0$ ) will lead to:

$$\left(\frac{d\sigma_p}{dx_i}\right)_m = \left(\frac{d\sigma_p}{dx_i}\right)_{x_i=0} = \frac{\sigma_{im} - \sigma_m^2}{\sigma_p} \quad \text{and} \quad \left(\frac{d\hat{k}_p}{d\sigma_p}\right)_m = \frac{(\hat{k}_i - R_m)\sigma_m}{\sigma_{im} - \sigma_m^2}$$

However since for all investors, the preference for returns is a linear function of the returns' standard deviation, (assuming they all maximize a quadratic utility function!), we have:

$$\hat{k} = R_f + \lambda\sigma$$

where  $\sigma$  is the volatility of the portfolio. For all assets we have equivalently:  $\hat{k}_i = R_f + \lambda\sigma_i$  and for the portfolio as well, or:

$$\hat{k}_p = R_f + \lambda\sigma_p \rightarrow \frac{d\hat{k}_i}{d\sigma_i} = \lambda \quad \text{and} \quad \left( \frac{d\hat{k}_p}{d\sigma_p} \right)_m = \frac{(\hat{k}_i - R_m)\sigma_m}{\sigma_{im} - \sigma_m^2}$$

and therefore,

$$\left( \frac{d\hat{k}_p}{d\sigma_p} \right)_m = \frac{(\hat{k}_i - R_m)\sigma_m}{\sigma_{im} - \sigma_m^2} = \lambda \quad \text{which leads to: } \hat{k}_i = R_m - \lambda\sigma_m + \frac{\lambda\sigma_{im}}{\sigma_m}$$

But also  $R_m = R_f + \lambda\sigma_m$  which we insert in the previous equation, leading thereby to a linear expression for risk discounting which assumes the form of the (CAPM) or:

$$\hat{k}_i = R_f + \frac{\lambda\sigma_{im}}{\sigma_m} \quad \text{and explicitly, } \hat{k}_i = R_f + \frac{\lambda \operatorname{cov}(\tilde{k}_i, \tilde{R}_m)}{\sigma_m}$$

as seen earlier. Note that this expression can be written in a form easily amenable to a linear regression in returns and providing an estimate for the  $\beta_i$  factor:

$$\hat{k}_i = R_f + \beta_i(R_m - R_f); \quad \beta_i = \frac{\lambda \operatorname{cov}(\tilde{k}_i, \tilde{R}_m)}{(R_m - R_f)\sigma_m}$$

Since  $\lambda = (R_m - R_f)/\sigma_m$  is the market price of risk, we obtain also the following expression for the beta factor,

$$\beta_i = \frac{\operatorname{cov}(\tilde{k}_i, \tilde{R}_m)}{\sigma_m^2}$$

With this ‘fundamental’ identity on hand, we can calculate the risk premium of an investment as well as the betas for traded securities.

### 3.5.4 Portfolio and utility maximization in practice

Market valuation of a portfolio and individual valuation of a portfolio are not the same. The latter is based on an individual preference for the assets composition of the portfolio and responding to specific needs. For example, denote by \$W\$ a budget to be invested and let  $y_i, i = 1, 2, 3, \dots, n$  be the dollars allocated to each of the available alternatives with a resulting uncertain payoff

$$\tilde{R} = \sum_{i=1}^n \tilde{r}_i(y_i)$$

The portfolio investment problem is then formulated by solving the following expected utility maximization problem:

$$\begin{aligned} \text{Maximize } Eu(\tilde{R}) &= Eu \left( \sum_{i=1}^n \tilde{r}_i(y_i) \right) \text{ subject to : } \sum_{i=1}^n y_i \leq W, y_i \geq 0, i \\ &= 1, 2, \dots, n \end{aligned}$$

where  $u(\cdot)$  is the individual utility function, providing a return risk ordering over all possible allocations. This problem has been solved in many ways. It clearly sets up a transformation of an uncertain payoffs problem into a problem which is deterministic and to which we can apply well-known optimization and numerical techniques. If  $u(\cdot)$  is a quadratic utility function and the rates of return are linear in the assets allocation (i.e.  $\tilde{r}_i(y_i) = \tilde{r}_i y_i$ ), then we have:

$$\begin{aligned} \text{Maximize}_{y_1, y_2, y_3, \dots, y_n} Eu(\tilde{R}) &= Eu\left(\sum_{i=1}^n \tilde{r}_i(y_i)\right) = \sum_{i=1}^n \hat{r}_i y_i - \mu \sum_{j=1}^n \sum_{i=1}^n \rho_{ij} y_i y_j; \\ \rho_{ij} &= \text{cov}(\tilde{r}_i, \tilde{r}_j) \end{aligned}$$

where  $\hat{r}_i$  is the mean rate of return on asset  $i$ ,  $\rho_{ij}$  is the covariance between the returns on two assets ( $i, j$ ) and  $\mu$  is a parameter expressing the investor's risk aversion ( $\mu > 0$ ) or risk loving ( $\mu < 0$ ). This defines a well known quadratic optimization problem that can be solved using standard computational software when the index of risk aversion is available. There are many other formulations of this portfolio problem due to Harry Markowitz, as well as many other techniques for solving it, such as scenario optimization, multi-criterion optimization and others.

The Markowitz approach had a huge impact on financial theory and practice. Its importance is due to three essential reasons. First, it justifies the well-known belief that it is not optimal to put all one's eggs in one basket (or the 'principle of diversification'). Second, a portfolio value is expressed in terms of its mean return and its variance, which can be measured by using statistical techniques. Further, the lower the correlation, the lower the risk. In fact, two highly and negatively correlated assets can be used to create an almost risk-free portfolio. Third and finally, for each asset there are two risks, one diversifiable through a combination of assets and the other non-diversifiable to be borne by the investor and for which there may be a return compensating this risk.

Markowitz suggested a creative approach to solving the quadratic utility portfolio problem by assuming a specific index of risk aversion. The procedure consists in solving two problems. The first problem consists in maximizing the expected returns subject to a risk (variance) constraint (Problem 1 below) and the second problem consists in minimizing the risk (measured by the variance) subject to a required expected return constraint (Problem 2 below). In other words,

$$\begin{array}{ll} \text{Problem 1:} & \text{Problem 2:} \\ \text{Maximize}_{y_1, y_2, y_3, \dots, y_n} \sum_{i=1}^n \hat{r}_i y_i & \text{Minimize}_{y_1, y_2, y_3, \dots, y_n} \sum_{j=1}^n \sum_{i=1}^n \rho_{ij} y_i y_j \\ \text{Subject to:} & \text{Subject to:} \\ \sum_{j=1}^n \sum_{i=1}^n \rho_{ij} y_i y_j \leq \lambda & \sum_{i=1}^n \hat{r}_i y_i \geq \mu \end{array}$$

An optimization of these problems provides the efficient set of portfolios defined in the  $(\lambda, \mu)$  plane.

The importance of Markowitz's (1959) seminal work cannot be overstated. It laid the foundation for portfolio theory whereby rational investors determine the optimal composition of their portfolio on the basis of the expected returns, the standard deviations of returns and the correlation coefficients of rates of return. Sharpe (1964) and Lintner (1965) gave it an important extension leading to the CAPM to measure the excess premium paid to hold a risky financial asset, as we saw earlier.

### 3.5.5 Capital markets and the CAPM again

The CAPM for the valuation of assets is essentially due to Markowitz, James Tobin and William Sharpe. Markowitz first set out to show that 'diversification pay', in other words an investment in more than one security, provides an opportunity to 'make money with less risk', compared to the prior belief that the optimal investment strategy consists of putting all of one's money in the 'best basket'. For risk-averse investors this was certainly a strategy to avoid. Subsequently, Tobin (1956) showed that when there is a riskless security, the set of efficient portfolios can be characterized by a 'two-fund separation theorem' which showed that an efficient portfolio can be represented by an investor putting some of his money in the riskless security and the remaining moneys invested in a representative fund constructed from the available securities. This led to the CAPM, stated explicitly by Sharpe in 1964 and Lintner in 1965. Both assumed that investors are homogeneous and mean-variance utility maximizers (or, alternatively, investors are quadratic utility maximizers). These assumptions led to an equilibrium of financial markets where securities' risks are measured by a linear function (due to the quadratic utility function) given by the risk-free rate and a beta multiplied by the relative returns of the 'mutual fund' (usually taken to be the market average rate of return). In this sense, the CAPM approach depends essentially on quadratic utility maximizing agents and a known risk-free rate. Thus, the returns of an asset  $i$  can be estimated by a linear regression given explicitly by:

$$k_i = \alpha_i + \beta_i R_m + \varepsilon_i$$

where  $R_m$  is the market rate of return calculated by the rate of return of the stock market as a whole,  $\beta_i$  is an asset specific parameter while  $\varepsilon_i$  is the statistical error. The CAPM has, of course, been subject to criticism. Its assumptions may be too strong: for example, it implies that all investors hold the same portfolio – which is the market portfolio, by definition fully diversified. In addition, in order to invest, it is sufficient to know the beta associated with a stock, since it is the parameter that fully describes the asset/stock return. Considerable effort has been devoted to estimating this parameter through a statistical analysis of stocks' risk–return history. The statistical results obtained in this manner should then clarify whether such a theory is applicable or not. For example, for *one-factor models (market premium in the CAPM)* the following regression is run:

$$(k_j - R_f) = \alpha_j + \beta_j [R_m - R_f] + \varepsilon_j$$

where  $k_j$  is the stock (asset)  $j$  rate of return at time  $t$ , when the risk-free rate at that time was  $R_f$ ,  $(\alpha_j, \beta_j)$  are regression parameters, while  $R_m$  is the market rate of return. Finally,  $\varepsilon_j$  is the residual value, an error term, assumed to be normally distributed with mean zero and known variance. Of course, if  $(\alpha_j \neq 0)$  then this will violate a basic assumption of no excess returns of the CAPM. In addition,  $R_m$  must also, according to the CAPM theory, capture the market portfolio. If, again, this is not the case, then it will also violate a basic assumption of the CAPM. Using the regression equation above, the risk consists now of (assuming perfect diversification, or equivalently, no correlation):

$$\text{var}(k_j - R_f) = \beta_j^2 \sigma_m^2 + \sigma_\varepsilon^2, \sigma_m^2 = \text{var}(R_m); \sigma_\varepsilon^2 = \text{var}(\varepsilon_j)$$

expressing risk as a summation of beta-squared times the index-market plus the residual risk.

The problems with a one-factor model, although theory-independent (since it is measured by simple financial statistics), are its assumptions. Namely, it assumes that the regression is stable and that nonstationarities and residual risks are known as well. Further, to estimate the regression parameters, long time series are needed, which renders their estimate untimely (and, if used carefully, it is of limited value). The generalization of the one-factor model into a *multiple-factor model* is also known as *APT*, or *arbitrage pricing theory*. Dropping the time index, it leads to:

$$(k_j - R_f) = \alpha_j + \beta_{j1}[R_m - R_f] + \cdots + \beta_{jK}[R_K - R_f] + \varepsilon_j$$

where  $R_K - R_f$  is the expected risk premium associated with factor  $K$ . The number of factors that can be used is large, including, among many others, the yield, interest-rate sensitivity, market capitalization, liquidity, leverage, labour intensity, recent performance (momentum), historical volatility, inflation, etc. This model leads also to risks defined by the matrices calculated (the orthogonal factors of APT) using the multivariate regression above, or

$$\Sigma = \mathbf{B}' \mathbf{\Gamma} \mathbf{B} + \mathbf{\Phi}$$

where  $\Sigma$  is the variance–covariance matrix of assets returns,  $\mathbf{B}$  is the matrix of assets' exposures to the different risk factors,  $\mathbf{\Gamma}$  is the vector of factor risk premiums (i.e. in excess of the risk-free rate for period  $t$ ) and finally,  $\mathbf{\Phi}$  is the (diagonal) matrix of asset residual risks. These matrices, unfortunately involve many parameters and are therefore very difficult to estimate.

A number of approaches based on the APT are available, however. One approach, the *fundamental factor model* assumes that the matrix  $\mathbf{B}$  is given and proceeds to estimate the vector  $\mathbf{\Gamma}$ . A second approach, called the *macroeconomic approach*, takes  $\mathbf{\Gamma}$  as given and estimates the matrix  $\mathbf{B}$ . Finally, the *statistical approach* seeks to estimate  $(\mathbf{B}, \mathbf{\Gamma})$  simultaneously. These techniques each have their problems and are therefore used in varying circumstances, validated by the validity of the data and the statistical results obtained.

### 3.5.6 Stochastic discount factor, assets pricing and the Euler equation

The financial valuation of assets is essentially based on defining an approach accounting for the time and risk of future payoffs. To do so, financial practice and theories have sought to determine a discounting mechanism that would, appropriately, reflect the current value of uncertain payoffs to be realized at some future time. Thus, techniques such as classical discounting based on a pre-specified discount rate (usually a borrowing rate of return provided by banks or some other interest rate) were used. Subsequently, a concept of risk-adjusted discount rate applied to discounting the mean value of a stream of payoffs was used. A particularly important advance in determining an appropriate discount mechanism was ushered in, first by the CAPM approach, as we saw above, and subsequently by the use of risk-neutral pricing, to be considered in forthcoming chapters. Both approaches use ‘the market mechanism’ to determine the appropriate discounting process to value an asset’s future payoffs. We shall consider these issues at length when we seek a risk-neutral approach to value options and derivatives in general. These approaches are not always applicable, however, in particular when markets are incomplete and the value of a portfolio may not be determined in a unique manner. In these circumstances, attempts have been made to maintain the framework inspired by rational expectations and at the same time be consistent theoretically and empirically verifiable. The SDF approach, or the generalized method of moments, seeks to value an asset generally in terms of its future values using a stochastic discount factor. This development follows risk-neutral pricing, which justified a risk-free rate discounting process with respect to some probability measure, as we shall see later on (in Chapter 6).

Explicitly, the SDF approach for a single asset states that the price of an asset equals the expected value of the asset payoff, times a stochastic discount factor. This approach has the advantage that it leads to some of the classical results of financial economics and at the same time it can be used by applying financial statistics in asset pricing by postulating such a relationship. Define:

$p_t$  = asset price at time  $t$  that an investor may wish to buy

$\tilde{x}_t$  = asset returns, a random variable

$\tilde{m}_t$  = a stochastic discount factor to be defined below

The SDF postulates:

$$p_t = E(\tilde{m}_{t+1}\tilde{x}_{t+1}), \tilde{m}_{t+1} = \frac{1}{1 + \tilde{R}_{t+1}}$$

where  $\tilde{R}_{t+1}$  is a random discounting. The rationality for such a postulate is based on the expected utility of a consumption-based model. Say that an investor has a utility function for consumption  $u(\cdot)$ , which remains the same at times  $t$  and  $t + 1$ . We let the discount factor be  $\rho$ , expressing the subjective discount rate of the consumer. Current consumption is certain, while next period’s consumption is uncertain and discounted. Thus, in terms of expected utility we have:

$$U(c_t, c_{t+1}) = u(c_t) + \rho E_t u(c_{t+1})$$

where  $E_t$  is an expectation operator based on the information up to time  $t$ . Now assume that  $s_t$  is a consumer's salary at time  $t$ , part of which may be invested for future consumption in an asset whose price is  $p_t$  (for example, buying stocks). Let  $y$  be the quantity of an asset bought (say a stock). Current consumption left over after such an investment equals:  $c_t = s_t - yp_t$ . If the asset price one period hence is  $\tilde{x}_{t+1}$ , then the next period consumption is simply equal to the sum of the period's current income and the return from the investment, namely  $c_{t+1} = s_{t+1} + y\tilde{x}_{t+1}$ . As a result, the consumer problem over two periods is reduced to:

$$U(c_t, c_{t+1}) = u(s_t - yp_t) + \beta E_t u(s_{t+1} + y\tilde{x}_{t+1})$$

The optimal quantity to invest (i.e. the number of shares to buy), found by maximizing the expected utility with respect to  $y$ , leads to:

$$\begin{aligned} \frac{\partial U}{\partial y} &= -p_t u'(s_t - p_t y) + \beta E_t [\tilde{x}_{t+1} u'(s_{t+1} + \tilde{x}_{t+1} y)] \\ &= -p_t u'(c_t) + \beta E_t [\tilde{x}_{t+1} u'(c_{t+1})] = 0 \end{aligned}$$

which yields for an optimum portfolio:

$$p_t = E_t \left[ \beta \frac{u'(c_{t+1})}{u'(c_t)} \tilde{x}_{t+1} \right]$$

Thus, if we set the stochastic discount factor,  $\tilde{m}_{t+1}$ , expressing the inter-temporal substitution of current and future marginal utilities of consumption, then:

$$\tilde{m}_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)} \quad \text{and therefore} \quad \tilde{R}_{t+1} = \frac{1 - \beta \frac{u'(c_{t+1})}{u'(c_t)}}{\beta \frac{u'(c_{t+1})}{u'(c_t)}}$$

the pricing equation becomes  $p_t = E_t [\tilde{m}_{t+1} \tilde{x}_{t+1}]$ , which is the desired SDF asset-pricing equation. This equation is particularly robust and therefore it is also very appealing. For example, if the utility function is of the logarithmic type,  $u(c) = \ln(c)$ , then,  $u'(c) = 1/c$  and  $\tilde{m}_{t+1} = \beta c_t / c_{t+1}$ , or  $\tilde{R}_{t+1} = \beta [(1/\beta)c_{t+1} - c_t] / c_t$  and further,  $p_t / c_t = \beta E_t [\tilde{x}_{t+1} / c_{t+1}]$ . In other words, if we write  $\pi_t = p_t / c_t$ ;  $\tilde{\pi}_{t+1} = \tilde{x}_{t+1} / c_{t+1}$ , then we have:  $\pi_t = \beta E_t (\tilde{\pi}_{t+1})$ .

The results of such an equation can be applied to a broad a number of situations, which were summarized by Cochrane (2001) and are given in Table 3.1. This approach is extremely powerful and will be considered subsequently in Chapter 8, since it is applicable to a broad number of situations and financial products. For example, the valuation of a call option would be (following the information in Table 3.1) given as follows (where the time index is ignored):

$$C = E [m \text{Max}(S_T - K, 0)]$$

This approach can be generalized in many ways, notably by considering multiple periods and various agents (heterogeneous or not) interacting in financial markets



**Table 3.1** Selected examples.

	Price $p_t$	Payoff $x_{t+1}$
Dividend paying stock	$p_t$	$p_{t+1} + d_{t+1}$
Investment return	1	$R_{t+1}$
Price/dividend ratio	$\frac{p_t}{d_t}$	$\left(\frac{p_{t+1}}{d_{t+1}} + 1\right) \frac{d_{t+1}}{d_t}$
Managed portfolio	$z_t$	$z_t R_{t+1}$
Moment condition	$E(p_t z_t)$	$x_{t+1} z_t$
One-period bond	$p_t$	1
Risk-free rate	1	$R_f$
Option	C	$\text{Max}[S_T - K, 0]$

to buy, sell and transact financial assets. Cochrane (2001) in particular suggests many such situations. An inter-temporal framework uses the Euler conditions for optimality to generate an equilibrium discount factor and will be considered in Chapter 9.

### Example

Consider a one-dollar investment in a risk-free asset whose payoff is  $R_f$ . Thus,  $1/(1 + R_f) = E(\tilde{m} * 1)$  and therefore we have  $1/(1 + R_f) = E(\tilde{m})$  as expected.

## 3.6 INFORMATION ASYMMETRY

Uncertainty and information asymmetry have special importance because of their effects on decision-makers. These result also in markets being ‘incomplete’ since the basic assumptions regarding ‘fair competition’ are violated. In general, the presumption that information is commonly shared is also, often, violated. Some information may be truthful, some may not be. Truth-in-lending for example, is an important legislation passed to protect consumers, which is, in most cases, difficult to enforce. Courts are filled with litigation on claims and counter-claims, leading to a battle of experts on what the truth is and where it may lie. Environmental litigation has often led to a ‘battle of PhDs’ expounding alternative and partial pearls of knowledge. In addition, positive, negative, informative, partial, asymmetric, etc. information has different effects on both decision-makers and markets. For example, firms and funds are extremely sensitive to negative information regarding their stock, their products as well as their services. Pharmaceutical firms may be bankrupted upon adverse publicity, whether true or not, regarding one of their products. For example, the Food and Drug Administration warning on the content of benzene in Perrier’s sparkling water has more than tainted the company’s image, its bottom-line profits and, at a certain time, its future prospects. Of course, the tremendous gamble Perrier has taken to meet these claims (that were not entirely verified) is a sign of the importance Perrier attached to its reputation and to the effects of negative information.

Information asymmetry and uncertainty can open up the possibility of cheating, however. For example, some consumer journals may receive money in various forms (mostly advertising dollars) not to publish certain articles and thereby manage information in a way that does not benefit the public. For this reason, regulatory authorities are needed in certain areas. A used-car salesman may be tempted to sell a car with defects unknown to the prospective buyer. In some countries, importers are not required to inform clients of the origin and the quality (state) of the product and parts used in the product sold. As a result, a product claimed to be new by the seller may not in fact be new, opening up many possibilities for cheating legally. These questions arise on Wall Street in many ways. In an article in the *Wall Street Journal* the question was raised by Hugo Dixon with respect to analysts' claims and the conflict of interest when they act according to their own edicts:

Shouldn't analysts put their money where their mouth is? That is the contrarian response to Merrill Lynch's decision to ban its analysts from buying shares of companies they cover. It might be said that researchers will have an incentive to give better opinions if they stand to make money if they are right – and lose money if they are wrong. Clients might also be comforted to know that the analysts who are peppering them with 'buy' recommendations are following their own advice. Under this contrarian position, the fact that some analysts buy shares in companies they follow isn't a conflict of interest all. Quite the reverse! It is an alignment of interests . . . this contrarian view cannot be dismissed as a piece of errant nonsense. But it is nevertheless misconceived. There is a better way of aligning analyst's financial interests with those of their clients. And there are potential conflicts caused by an analyst trading stocks he or she covers!

These 'information problems' are the subject of extensive study, both for practical and theoretical reasons. A number of references are included at the end of this chapter. Below, we only consider some of the outstanding implications information asymmetry may create.

### **3.6.1 'The lemon phenomenon' or adverse selection**

In a seminal paper, Akerlof (1970) pointed out that goods of different qualities may be uniformly priced when buyers cannot realize that there are quality differences. This is also called 'adverse selection' because some of the information associated with the choice problem may be hidden. For example, one may buy a used car, not knowing its true state, and therefore be willing to pay a price that would not truly reflect the value of the car. In fact, we may pay an agreed-upon market price even though this may be a lemon. The used-car salesman may have such information but, for some obvious reason, he may not be amenable to revealing the true state of the car. In such situations, price is not an indicator of quality and informed sellers can resort to opportunistic behaviour (the used-car salesman phenomenon stated above). While Akerlof demonstrated that average quality might still be a function of price, individual units may not be priced at that level. By contrast, people who discovered in the 1980s that they had AIDS were very quick in taking

out very large life insurance (before insurance firms knew what it really entailed and therefore were at first less informed than the insured). Bonds or stocks of various qualities (but of equal ratings) are sometimes difficult to discern for an individual investor intending to buy. For this reason, rating agencies, tracking and following firms have an important role to play, compensating for the problems of information asymmetry and making markets more efficient. This role is not always properly played, as evidenced by the Enron debacle, where changing accounting practices had in fact hidden information from the public.

Information asymmetry and uncertainty can largely explain the desires of consumers to buy service or product warranties to protect themselves against failures or to favour firms who possess service organizations (in particular when the products are complex or involve some up-to-date technologies). Generally in transactions between producers and suppliers, uncertainty leads to constructing long-term trustworthy relationships and contractual engagements to assure that 'what is contracted is also delivered'. The potential for adverse selection may also be used to protect national markets. Anti-dumping laws, non-tariff trade barriers, national standards and approval of various sorts are some of the means used to manage problems of adverse selection on the one hand and to manage market entries to maintain a competitive advantage on the other. Finance and insurance are abound with applications and examples where asymmetry induces an uncertainty which has nothing to do with 'what nature does' but with 'what people do'.

Problems of adverse selection can sometimes be overcome by compulsory insurance regulation requiring all homeowners to insure their homes or requiring everyone to take out medical insurance, for example. Some employers insure all their employees as one package to avoid adverse selection problems. If everyone is insured, high-risk individuals will be better off (since they will be insured at a lower premium than justified by the risks they have in fact). Whether low-risk individuals are better off under this scheme depends on how risk-averse they are as the insurance they are offered is not actuarially fair.

### **3.6.2 'The moral hazard problem'**

For many situations, the cost of providing a product or a service depends on the behaviour of the purchaser. For example, the cost of insurance depends on the amount of travel done by the purchaser and by the care he takes in driving. Similarly, the cost of warranties depends on the care of the purchaser in using the commodity. Such behaviour cannot always be observed directly by the supplier/seller. As a result, the price cannot depend on the behaviour of the purchaser that affects costs. In this case, equilibrium cannot always be the first-best optimum and some intervention is required to reach the best solution. Questions are of course, how and how much. For example, should car insurance be obligatory? How do purchasers react after buying insurance? How do markets behave when there is moral hazard and how can it be compensated?

Imperfect monitoring of fund managers, for example, can lead to moral hazard. What does it mean? It implies that when the fund manager cannot be observed, there is a possibility that the provider, the fund manager, will use that fact to his

advantage and not deliver the right level of performance. Of course, if we contract the delivery of a given level of returns and if the fund manager knowingly does not maintain the terms of the contract, he would be cheating. We can deal with such problems with various sorts of controls combined with incentive contracts to create an incentive not to cheat or lie and to perform in the interest of the investor. If a fund manager were to cheat or lie, and if he were detected, he would then be penalized accordingly (following the terms agreed on by the contract).

For some, transparency (i.e. sharing information) is essential to provide a 'signal' that they operate with the best of intentions. For example, some restaurants might open their kitchen to their patrons to convey a message of truthfulness in so far as cleanliness is concerned. A supplier would let the buyer visit the manufacturing facilities as well as reveal procedures relating to quality, machining controls and the production process in general. A fund manager will provide regulators with truthful reports regarding the fund's state and strategy.

Moral hazard pervades some of the most excruciating problems of finance. The problem of deposit insurance and the 'too big to fail' syndrome encourages excessive risk taking. As a result of implicit governmental guarantee, banks enjoy a lower cost of capital, which leads to the consistent under-pricing of credit. Swings in economic cycles are thus accentuated. The Asian financial crisis of 1998 is a case in point. The extent of its moral hazard is difficult to measure but with each bail out by governments and the IMF, the trend for excessive risk taking is reinforced.

### 3.6.3 Examples of moral hazard

(1) An over-insured driver may drive recklessly. Thus, while the insured motorist is protected against any accident, this may induce him to behave in a nonrational manner and cause accidents that are costly to society.

(2) In 1998, the NYSE at last, belatedly (since the practice was acknowledged to have been going on since 1992), investigated charges against floor brokers for 'front-running' or 'flipping'. This is a practice in which the brokers used information obtained on the floor to trade and earn profits on their own behalf. One group of brokers was charged with making \$11 millions. (Financial Times, 20 February 2001).

(3) The de-responsibilization of workers in factories also induces a moral hazard. It is for this reason that incentives, performance indexation and responsabilization are so important and needed to minimize the risks from moral hazard (whether these are tangibles or intangibles). For example, decentralization of the workplace and getting people involved in their jobs may be a means to make them care a little more about their job and deliver the required performance in everything they do.

Throughout these examples, there are negative inducements to good performance. To control or reduce these risks, it is necessary to proceed in a number of ways. Today's concern for firms' organizational design, the management of traders and

their compensation packages, is a reflection of the need to construct relationships that do not induce counterproductive acts. Some of the steps that can be followed include:

- (1) Detecting signals of various forms and origins to reveal agents' behaviours, rationality and performance. A greater understanding of agents' behaviour can lead to a better design of the workplace and to appropriate inducements for all parties involved in the firm's business.
- (2) Managing and controlling the relationship between business partners, employees and workers. This means that no relationship can be taken for granted. Earlier, we saw that information asymmetry can lead to opportunistic behaviour such as cheating, lying and being counterproductive, just because there may be an advantage in doing so without having to sustain the consequences of such behaviour.
- (3) Developing an environment which is cooperative, honest and open, and which leads to a frank exchange of information and optimal performance.

All these actions are important. It is therefore not surprising that many of the concerns of managers deal with people, communication, simplification and the transparency of everything firms do.

#### **Example: Genetic testing and insurance**

In a *Financial Times* article of 7–8 November 1998, it was pointed out that genetic testing can give early warning of disease and that those results could have serious consequences for those seeking insurance. The problem at hand, therefore, indicates an important effect of information on insurance. How can such problems be resolved? Should genetic testing be a requirement imposed by insurers or should they not? If accurate genetic tests do become widely available, they could encourage two trends that would undermine the present economic basis of the insurance industry:

- (1) Adverse selection: people who know they are at high risk take out insurance. This drives up the price of premiums, so low-risk people are deterred from taking out policies and withdrawing from the insurance 'pool'.
- (2) Cherry-picking: insurers identify people at lower risk than average and offer them reduced premiums. If they join the preferred pool, this increases the average risk in the standard pool and premiums have to rise. For example, the insurance industry suffered from adverse selection in the 1980s when individuals who knew they had HIV/AIDS took out extra insurance cover without disclosing their HIV status. A more respectable name for cherry-picking is market segmentation, as applied in general insurance for house contents and motor vehicles, where policies favouring the (lesser) better risks are common. These are becoming increasingly important issues due to the improved databases available about insured and insurance firms ability to tap these databases.

### 3.6.4 Signalling and screening

In conditions of information asymmetry, one of the parties may have an incentive to reveal some of the information it has. The seller of an outstanding concept for a start-up to invest in will certainly have an interest in making his concept transparent to the potential VC (venture capitalist) investor. He may do so in a number of ways, such as pricing it high and therefore conveying the message to the potential VC that it is necessarily (at that price) a dream concept that will realize an extraordinary profit in an IPO (but then, the concept seller may also be cheating!). The seller may also spend heavily on advertising the concept, claiming that it is an outstanding one with special technology that it is hard to verify (but then, the seller may again be lying!). Claiming that a start-up concept is just 'great' may be insufficient. Not all VCs are gullible. They require and look for signals that reveal the true potential of a concept and its potential for making large profits. Pricing, warranties, advertising, are some of the means used by well-established firms selling a product to send signals. For example, the seller of a lemon with a warranty will eventually lose money. Similarly, a firm that wants to limit the entry of new competitors may signal that its costs are very low (and so if they decide to enter, they are likely to lose money in a price battle). Advertising heavily may be recuperated only through repeat purchase and, therefore, over-advertising may be used as a signal that the over-advertised products are of good quality. For start-ups, the game is quite different, VCs look for the signals leading to potential success, such as good management, proven results, patentable ideas and a huge potential combined with hefty growth rate in sales. Still, these are only signals, and more sophisticated investors actually get involved in the start-ups they invest in to reduce further the risks of surprise.

Uninformed parties, however, have an incentive to look for and obtain information. For example, shop and compare, search for a job etc. are instances of information-seeking by uninformed parties. Such activities are called *screening*. A life insurer requires a medical record history; a driver who has a poor accident record history is likely to pay a greater premium (if he can obtain insurance at all). If characteristics of customers are unobservable, firms can use *self-selection* constraints as an aid in screening to reveal private information. For example, consider the phenomenon of rising wage profiles where workers get paid an increasing wage over their careers. An explanation may be that firms are interested in hiring workers who will stay for a long time. Especially if workers get training or experience, which is valuable elsewhere, this is a valid concern. Then they will pay workers below the market level initially so that only 'loyal' workers will self-select to work for the firm. The classic example of 'signalling' was first analysed by Spence (1974) who pointed out that high-productivity individuals try to differentiate themselves from low-productivity ones by the amount of education they acquire. In other words, only the most productive workers invest in education. This is the case because the signalling cost to the productive workers is lower than to low-productivity workers and therefore firms can differentiate between these two types of workers because they make different choices.

Uninformed parties can screen by offering a menu of choices or possible contracts to prospective (informed) trading partners who ‘self-select’ one of these offerings. Such screening was pointed out in insurance economics for example, showing that, if the insurer offers a menu of insurance policies with different premiums and amounts of cover, the high-risk clients self-select into a policy with high cover. This can lead to insurance firms portfolios where bad risks crowd out the good ones. Insurance companies that are aware of these problems create risk groups and demand higher premium from members in the ‘bad’ risk portfolio, as well as introducing a number of clauses that will share responsibility for payments in case claims are made (Reyniers, 1999).

### 3.6.5 The principal–agent problem

Consider a business or economic situation involving two parties: a principal and an agent. For example the manager of a company may be the ‘agent’ for a stockholder who acts as a ‘principal’, trusting the manager to perform his job in the interest of stockholders. Similar situations arise between a fund and its traders. The fund ‘principal’ seeks to provide incentives motivating the traders – ‘agents’ – to perform in the interest of the fund. In these situations, the actions taken by the agent may be observed only imperfectly. That is, the performance observed by the principal is the outcome of the agent actions (known only by the agent) and some random variable, which may be known, or unknown by the agent at the time an action is assumed and taken. The principal – agent problem consists then in determining the rules for sharing the outcomes obtained through such an organization. This asymmetry of information leads of course to a situation of potential moral hazard. There are several approaches to this problem, which we consider below. For example, designing appropriate incentive systems is of great practical importance. CAR (capital adequacy requirements), health-care compensation etc. are only some of the tools that are used and widely practised to mitigate the effects of moral hazard through agency.

The principal–agent problem is well researched, and there are many research papers using assumptions leading to what we may call normative behaviours and normative compensations. Here we consider a simple example based on the *first-order approach*. Let  $\tilde{x}$  be a random variable, which represents the gross return, obtained by a hedge fund manager – the principal. The distribution of this return is influenced by the variable  $a$ , which is under the control of the trader and not observed by the principal. Now assume that the sharing rule is given by  $F(\tilde{x}, a)$  while the probability distribution of the outcome is  $f(\tilde{x}, a)$  which is independent statistically of  $a$ . The principal–agent problem consists in determining the amount transferred to the agent by the principal in order to compensate him for the efforts he performs on behalf of the principal. To do so, we assume that the agent utility is separable and given by:

$$V(y, a) = v(y) - w(a); v' > 0, v'' \leq 0, w' > 0, w'' > 0$$

In order to assure the agent’s participation, it is necessary to provide at least an

expected utility:

$$EV(y, a) \geq 0 \text{ or } E v(y) \geq w(a)$$

In this case, the utility of the principal is:

$$u(\tilde{x} - y), u' > 0, u'' \leq 0$$

The problems we formulate depends then on the information distribution between the principal and the agent.

Assume that the agent's effort 'a' is observable by the principal. In this case, the problem of the principal is formulated by optimizing both 'a' and of course the transfer. That is,

$$\text{Max}_{a, y(\cdot)} Eu(\tilde{x} - y(\tilde{x})) \text{ subject to: } Ev(y(\tilde{x}) - w(a)) \geq 0$$

By applying the conditions for optimality, the optimal solution is found to be:

$$\frac{u'(\tilde{x} - y(\tilde{x}))}{v'(y(\tilde{x}))} = \lambda$$

This yields a sharing rule based on the agent and the principal marginal utility functions, a necessary condition for Pareto optimal risk sharing. A differentiation of the sharing rule, indicates that:

$$\frac{dy}{dx} = \frac{u''/u'}{u''/u' + v''/v'}$$

For example, if we assume exponential utility functions given by:

$$u(w) = \frac{-e^{-aw_u}}{a}; v(w) = \frac{-e^{-bw_v}}{b}$$

where initial wealth is given by  $(w_u, w_v)$  then, we have a *linear risk-sharing rule*, given by:

$$\frac{e^{-a(w_u - (x - y(x)))}}{e^{-b(w_v - y(x))}} = \lambda \Rightarrow y(x) = \left( \frac{a}{a + b} \right) x + \left( \frac{bw_v - aw_u}{a + b} \right)$$

In other words, the share of the first party is proportional to the risk tolerance, which is given by  $\lambda/a$  and  $\lambda/b$  respectively.

Similarly, assume a fund manager can observe the trader's effort. Or, alternatively, consider a manager-trader relationship where there is a direct relationship between performance and effort. For example, for salesmen of financial products, there is a direct relationship between the performance of the salesmen (quantity of contracts sold) and his effective effort. Let  $e$  be the effort

and  $P(e)$  the profit function. The employee's cost is  $C(e)$  and the employee's reservation utility is  $u$ . There are a number of simple payment schemes that can motivate a trader/worker to work and provide the efficient amount of effort. These are payments based on effort, forcing contracts and franchises considered below (Reyniers (1999))

(1) *Payments based on effort*: The worker is paid based on his effort,  $e$ , according to the wage payment ( $we + K$ ), the manager's problem is then to solve:

$$\begin{aligned} \text{Max } \pi &= P(e) - (we + K) - C(e) \\ \text{Subject to: } & we + K - C(e) \geq u \end{aligned}$$

The inequality constraint is called the 'participation constraint' or the 'individual rationality constraint'. The employer has no motivation to give more money to the trader than his reservation utility. In this case, the effort selected by the manager will be at the level where marginal cost equals the marginal profit of effort, or  $P'(e^*) = C'(e^*)$ . The trader has to be encouraged to provide the optimal effort level that leads to the *incentive compatibility constraint*. In other words, the worker's net payoff should be maximized at the optimal effort or  $w = C'(e^*)$ . Thus, the trader is paid a wage per unit time equal to his marginal disutility of effort and a lump sum  $K$  that leaves him with his reservation utility.

(2) *Forcing contracts*: The manager could propose to pay the trader a lump sum  $L$  which gives him his reservation utility if he makes effort  $e^*$ , i.e.  $L = u + C(e^*)$  and zero otherwise. Clearly, the participation and incentive compatibility constraints are satisfied under this simple payment scheme. This arrangement is called a forcing contract because the trader is forced to make effort  $e^*$  (while above, the trader is left to select his effort level).

(3) *Franchises*: Now assume that the trader can keep the profits of his effort in return for a certain payment to the principal/manager. This can be interpreted as a franchise structure (similar in some ways to a fund of funds). To set the franchise fee, the trader proceeds as follows. First the trader maximizes  $P(e) - C(e) - F$  and therefore chooses the same optimal effort as before such that:  $P'(e^*) = C'(e^*)$ . The principal/manager can charge a franchise fee which leaves the trader with his reservation utility:  $F = P(e^*) - C(e^*) - u$ . When the effort cannot be observed, the problem is more difficult. In this case, payment based on effort is not possible. If we choose to pay based on output, then the employer would choose a franchise structure. However, if the employee is risk-averse, he will seek some payment to compensate the risk he is assuming. If the manager is risk-neutral, he may be willing to assume the trader's risk and therefore the franchise solution will not be possible in its current form!

## REFERENCES AND FURTHER READING

- Akerlof, G. (1970) The market for lemons: Quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, **84**, 488–500.
- Allais, M. (1953) Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine, *Econometrica*, **21**, 503–546.

- Allais, M. (1979) The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School, in M. Allais, and O. Hagen (Eds), *Expected Utility Hypothesis and the Allais Paradox*, D. Reidel, Dordrecht.
- Arrow, K.J. (1951) Alternative approaches to the theory of choice in risk-taking situations, *Econometrica*, October.
- Arrow, K.J. (1965) *Aspects of the Theory of Risk-Bearing*, Yrjö Jahnssonin Säätiö, Helsinki.
- Arrow, K.J. (1982) Risk perception in psychology and in economics, *Economics Inquiry*, January, 1–9.
- Bawa, V. (1978) Safety first, stochastic dominance and optimal portfolio choice, *Journal of Financial and Quantitative Analysis*, **13**(2), 255–271.
- Beard, R.E., T. Pentikainen and E. Pesonen (1979) *Risk Theory* (2nd edn), Methuen, London.
- Bell, D. (1982) Regret in decision making under uncertainty, *Operations Research*, **30**, 961–981.
- Bell, D. (1985) Disappointment in decision making under uncertainty, *Operations Research*, **33**, 1–27.
- Bell, D. (1995) Risk, return and utility, *Management Science*, **41**, 23–30.
- Bernoulli, D. (1954) Exposition of a new theory on the measurement of risk, *Econometrica*, January.
- Bierman, H., Jr (1989) The Allais paradox: A framing perspective, *Behavioral Science*, **34**, 46–52.
- Borch, K. (1968) *The Economics of Uncertainty*, Princeton University Press, Princeton, NJ.
- Borch, K. (1974) *The Mathematical Theory of Insurance*, Lexington Books, Lexington, MASS.
- Borch, K., and J. Mossin (1968) *Risk and Uncertainty*, Proceedings of the Conference on Risk and Uncertainty of the International Economic Association, Macmillan, London.
- Buhlmann, H. (1970) *Mathematical Methods in Risk Theory*, Springer-Verlag, Bonn.
- Chew, Soo H., and Larry G. Epstein (1989) The structure of preferences and attitudes towards the timing of the resolution of uncertainty, *International Economic Review*, **30**, 103–117.
- Christ, Marshall (2001) *Operational Risks*, John Wiley & Sons, Inc., New York.
- Cochrane, John H., (2001) *Asset Pricing*, Princeton University Press, Princeton, New Jersey.
- Dionne, G. (1981) Moral hazard and search activity, *Journal of Risk and Insurance*, **48**, 422–434.
- Dionne, G. (1983) Adverse selection and repeated insurance contracts, *Geneva Papers on Risk and Insurance*, **29**, 316–332.
- Dreze, Jacques, and Franco Modigliani (1966) Epargne et consommation en avenir aleatoire, *Cahiers du Seminaire d'Econometrie*.
- Dyer, J.S., and J. Jia (1997) Relative risk–value model, *European Journal of Operations Research*, **103**, 170–185.
- Eeckoudt, L., and M. Kimball (1991) Background risk prudence and the demand for insurance, in *Contributions to Insurance Economics*, G. Dionne (Ed.), Kluwer Academic Press, Boston, MA.
- Ellsberg, D. (1961) Risk, ambiguity and the Savage axioms, *Quarterly Journal of Economics*, November, 643–669.
- Epstein, Larry G., and Stanley E. Zin (1989) Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica*, **57**, 937–969.
- Epstein, Larry G., and Stanley E. Zin (1991) Substitution, risk aversion and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy*, **99**, 263–286.
- Fama, Eugene F. (1992) The cross-section of expected stock returns, *The Journal of Finance*, **47**, 427–465.
- Fama, Eugene F. (1996) The CAPM is wanted, dead or alive, *The Journal of Finance*, **51**, 1947.
- Fishburn, P.C. (1970) *Utility Theory for Decision Making*, John Wiley & Sons, Inc. New York.
- Fishburn, P.C. (1988) *Nonlinear Preference and Utility Theory*, The Johns Hopkins University Press, Baltimore, MD.

- Friedman, M., and L.J. Savage (1948) The utility analysis of choices involving risk, *Journal of Political Economy*, August.
- Friedman, M., and L.J. Savage (1952) The expected utility hypothesis and the measurability of utility, *Journal of Political Economy*, December.
- Grossman, S., and O. Hart (1983) An analysis of the principal agent model, *Econometrica*, **51**, 7–46.
- Gul, Faruk (1991) A theory of disappointment aversion, *Econometrica*, **59**, 667–686.
- Hadar, Josef, and William R. Russell (1969) Rules for ordering uncertain prospects, *American Economic Review*, **59**, 25–34.
- Holmstrom, B. (1979) Moral hazard and observability, *Bell Journal of Economics*, **10**, 74–91.
- Holmstrom, B. (1982) Moral hazard in teams, *Bell Journal of Economics*, **13**, 324–340.
- Hirschleifer, J. (1970) Where are we in the theory of information, *American Economic Review*, **63**, 31–39.
- Hirschleifer, J., and J.G. Riley (1979) The analysis of uncertainty and information: An expository survey, *Journal of Economic Literature*, **17**, 1375–1421.
- Holmstrom, B. (1979) Moral hazard and observability, *Bell Journal of Economics*, **10**, 74–91.
- Jacque, L., and C.S. Tapiero (1987) Premium valuation in international insurance, *Scandinavian Actuarial Journal*, 50–61.
- Jacque, L., and C.S. Tapiero (1988) Insurance premium allocation and loss prevention in a large firm: A principal–agency analysis, in M. Sarnat and G. Szego (Eds), *Studies in Banking and Finance*.
- Jia, J., J.S. Dyer and J.C. Butler (2001) Generalized disappointment models, *Journal of Risk and Uncertainty*, **22**, 159–178.
- Kahnemann, D., and A. Tversky (1979) Prospect theory: An analysis of decision under risk, *Econometrica*, March, 263–291.
- Kreps, D. (1979) A representation theorem for preference for flexibility, *Econometrica*, **47**, 565–577.
- Kimball, M. (1990) Precautionary saving in the small and in the large, *Econometrica*, **58**, 53–78.
- Knight, F.H. (1921) *Risk, Uncertainty and Profit*, Houghton Mifflin, New York.
- Kreps, David M., and Evan L. Porteus (1978) Temporal resolution of uncertainty and dynamic choice theory, *Econometrica*, **46**, 185–200.
- Kreps, David M., and Evan L. Porteus (1979) Dynamic choice theory and dynamic programming, *Econometrica*, **47**, 91–100.
- Laibson, David (1997) ‘Golden eggs and hyperbolic discounting’, *Quarterly Journal of Economics*, **112**, 443–477.
- Lintner, J. (1965) The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economic and Statistics*, **47**, 13–37.
- Lintner, J. (1965) Security prices, risk and maximum gain from diversification, *Journal of Finance*, **20**, 587–615.
- Loomes, Graham, and Robert Sugden (1986) Disappointment and dynamic consistency in choice under uncertainty, *Review of Economic Studies*, **53**, 271–282.
- Lucas, R.E. (1978) Asset prices in an exchange economy, *Econometrica*, **46**, 1429–1446.
- Machina, M.J. (1982) Expected utility analysis without the independence axiom, *Econometrica*, March, 277–323.
- Machina, M.J. (1987) Choice under uncertainty, problems solved and unsolved, *Economic Perspectives*, Summer, 121–154.
- Markowitz, Harry M. (1959) *Portfolio Selection; Efficient Diversification of Investments*, John Wiley & Sons, Inc., New York.
- Mossin, Jan (1969) A note on uncertainty and preferences in a temporal context, *American Economic Review*, **59**, 172–174.
- Pauly, M.V. (1974) Overinsurance and the public provision of insurance: The roles of moral hazard and adverse selection, *Quarterly Journal of Economics*, **88**, 44–74.
- Pratt, J.W. (1964) Risk aversion in the small and in the large, *Econometrica*, **32**, 122–136.

- Pratt, J.W. (1990) The logic of partial-risk aversion: Paradox lost, *Journal of Risk and Uncertainty*, **3**, 105–113.
- Quiggin, J. (1985) Subjective utility, anticipated utility and the Allais paradox, *Organizational Behavior and Human Decision Processes*, February, 94–101.
- Rabin, Matthew (1998) Psychology and economics, *Journal of Economic Literature*, **36** 11–46.
- Reyniers, D. (1999) Lecture Notes in Microeconomics, London School of Economics, London.
- Riley, J. (1975) Competitive signalling, *Journal of Economic Theory*, **10**, 174–186.
- Rogerson, W.P. (1985) The first order approach to principal agent problems, *Econometrica*, **53**, 1357–1367.
- Ross, S. (1973) On the economic theory of agency: The principal's problem, *American Economic Review*, **63**(2), 134–139.
- Ross, Stephen A. (1976) The arbitrage theory of capital asset pricing, *Journal of Monetary Economics*, **13**(3), 341–360.
- Samuelson, Paul A. (1963) Risk and uncertainty: A fallacy of large numbers, *Scientia*, **98**, 108–163.
- Sharpe, W.F. (1964) Capital asset prices: A theory of market equilibrium under risk, *The Journal of Finance*, **19**, 425–442.
- Siegel, Jeremy J., and Richard H. Thaler (1997) The Equity Premium Puzzle, *Journal of Economic Perspectives*, **11**, 191–200.
- Spence, M. (1974) *Market Signaling*, Harvard University Press, Cambridge, MA.
- Spence, Michael, and Richard Zeckhauser (1972) The effect of the timing of consumption decisions and the resolution of lotteries on the choice of lotteries, *Econometrica*, **40**, 401–403.
- Sugden, R. (1993) An axiomatic foundation of regret theory, *Journal of Economic Theory*, **60**, 150–180.
- Tapiero, C.S. (1983) The optimal control of a jump mutual insurance process, *Astin Bulletin*, **13**, 13–21.
- Tapiero, C.S. (1984) A mutual insurance diffusion stochastic control problem, *Journal of Economic Dynamics and Control*, **7**, 241–260.
- Tapiero, C.S. (1986) The systems approach to insurance company management, in *Developments of Control Theory for Economic Analysis*, C. Carraro and D. Sartore (Eds), Martinus Nijhoff, Dordrecht.
- Tapiero, C.S. (1988) *Applied Stochastic Models and Control in Management*, North Holland, Amsterdam.
- Tapiero, C.S., and L. Jacque (1987) The expected cost of ruin and insurance premiums in mutual insurance, *Journal of Risk and Insurance*, **54** (3), 594–602.
- Tapiero, C.S., and D. Zuckerman (1982) Optimum excess-loss reinsurance: A dynamic framework, *Stochastic Processes and Applications*, **12**, 85–96.
- Tapiero, C.S., and D. Zuckerman (1983) Optimal investment policy of an insurance firm, *Insurance Mathematics and Economics*, **2**, 103–112.
- Tobin, J. (1956) The interest elasticity of the transaction demand for cash, *Review of Economics and Statistics*, **38**, 241–247.
- Willasen, Y. (1981) Expected utility, Chebychev bounds, mean variance analysis, *Scandinavian Economic Journal*, **83**, 419–428.
- Willasen, Y. (1990) Best upper and lower Tchebycheff bounds on expected utility, *Review of Economic Studies*, **57**, 513–520.

# Probability and Finance

## 4.1 INTRODUCTION

Probability modelling in finance and economics provides a means to rationalize the unknown by imbedding it into a coherent framework, clearly distinguishing what we know and what we do not know. Yet, the assumption that we can formalize our lack of knowledge is both presumptuous and essential at the same time. To appreciate the problems of probability modelling it is essential to distinguish between randomness, uncertainty and chaos. These terms are central to an important polemic regarding ‘modelling cultures’ in probability, finance and economics. Kalman (1994) states that ‘the majority of observed phenomena of randomness in nature (always excluding games of chance) cannot and should not be explained by conventional probability theory; there is little or no experimental evidence in favour of (conventional) probability but there is massive, accumulating evidence that explanations and even descriptions should be sought outside the conventional framework’. This means that randomness might be defined without the use of probabilities. Kolmogorov, defined randomness in terms of non-uniqueness and non-regularity. For example, a die has six faces and therefore it has non-uniqueness. Further, the expansion of  $\sqrt{2}$  or of  $\pi$  provides an infinite string of numbers that appear irregularly, and can therefore be thought of as ‘random’. The Nobel Laureate, Born, in his 1954 inaugural address also stated that randomness occurs when ‘determinacy lapses into indeterminacy’ without any logical, mathematical, empirical or physical argumentation, preceding thereby an important research effort on chaos. Kalman, seeking to explain these approaches to modelling defined chaos as ‘randomness without probability’. Statements such as ‘we might have trouble forecasting the temperature of coffee one minute in advance, but we should have little difficulty in forecasting it an hour ahead’ by Edward Lorenz, a weather forecaster and one of the co-founders of chaos theory, reinforces the many dilemmas we must deal with in modelling uncertain phenomena. In weather modelling and forecasting for example, involving in many cases as many as 50 000 equations and more, it is presumed that if small models can predict well, it is only natural to expect that bigger and more sophisticated ones can do better. This turned out not to be the case, however. Bigger does not always turn out to be better; more sophisticated does not always mean improved accuracy.

In weather forecasting it soon became evident that no matter what the size and sophistication of the models used, forecasting accuracy decreased considerably beyond two to three days and provided no better predictions than using the average weather conditions of similar days of previous years to predict temperature, rainfall, or snow. What came to be known as the ‘butterfly effect’ (meaning in fact an insensitivity to initial conditions) or the effects of a flying butterfly exerting an unlikely and unpredictable critical influence on future weather patterns (Lorenz, 1966). In the short term too, the accuracy of weather forecasting could not improve much beyond the use of the naive approach which predicts that tomorrow’s or the next day’s weather will be exactly the same as today’s. Subsequent studies have amplified the importance of chaos in biology. Similar issues are raised in financial forecasting: the time scale of data, whether it is tickertape data or daily, weekly or monthly stock quotations, alters significantly the meaningfulness of forecasts.

In economic and business forecasting, the accuracy of predictions did not turn out to be any better than those of weather forecasting seen above. Further, accrued evidence points out that assumptions made by probability models are in practice violated. Long-run memory undermines the existence of martingales in finance. Further, can stock prices uncertainty or ‘noise’ be modelled by Brownian motion? This is one of the issues we must confront and deal with in financial modelling. The index of Hurst, entropy and chaos, which we shall discuss subsequently, are important concepts because they stimulate and highlight that there may be other approaches to be reckoned with and thereby stimulate economic and financial theoretical and empirical thinking.

The study of nonlinear financial time series and in particular chaos has assumed recently an added importance. Traditionally, it has been assumed for mathematical convenience that time series have a number of characteristics including:

- Existence of an equilibrium (or fixed point or a stationary state), or equivalently an insensitivity to initial conditions in the long run.
- Periodicity.
- Structural stability which allows the transformation of equations which are hard to study to some other forms which are stable and amenable to analysis.

There are a number of physical and economic phenomena that do not share all these properties. When this is the case, we call these series chaotic, implying that both indeterminacy and our inability to predict what the state of a system may be. Chaos can thus occur in both deterministic time series as well as in stochastic ones. For this reason, it has re-ignited the age-old confrontation of a deterministic versus a probabilistic view of nature and the world as well as mathematical modelling between externally and internally induced disturbances (which are the source of nonlinearities).

Commensurate analysis of nonlinear time series has also followed its course in finance. ARCH and GARCH type models used to estimate volatility are also nonlinear models expressed as a function (linear or not) of past variations in stocks.

Their analysis and estimation is the more difficult, the greater the nonlinearities assumed in representing the process. Current research is diverted towards the study of various nonlinear (non-Gaussian) and leptokurtic distributions, seeking to bridge a gap between traditional probability approaches in finance (based on the normal probability distribution) and systems exhibiting a chaotic behaviour and ‘fat tails’ in their distributions. For example, a great deal of research effort is devoted to explain why probability distributions are leptokurtic. Some approaches span herding behaviour in financial markets, reflecting the interaction of traders, imitation of investor groups and the following of gurus or opinion leaders. In such circumstances, collective behaviour can be ‘irrational’ leading to markets crashing. A paper ‘Turbulent cascades in foreign exchange markets’ by Ghashghaie *et al.*, published in a letter in *Nature* in 1996, has also pointed to the statistical observation that a similar behaviour is seen in financial exchange markets and hydrodynamic turbulence. Such behaviour clearly points to a ‘non-Gaussian-Normal noise’ and thereby to invalidating the assumption of ‘normal noise’ implicit in the underlying random walk models used in finance.

In general, in order to model uncertainty we seek to distinguish the known from the unknown and find some mechanisms (such as theories, common sense, metaphors and more often intuition) to reconcile our knowledge with our lack of it. For this reason, modelling uncertainty is not merely a collection of techniques but an art in blending the relevant aspects of a situation and its unforeseen consequences with a descriptive, yet theoretically justifiable and tractable, economic and mathematical methodology. Of course, we conveniently use probabilities to describe quantitatively the set of possible events that may unfold over time. Specification of these probabilities and their associated distributions are important and based on an understanding of the process at hand and the accrued evidence we can apply to estimate these probabilities. Any model is rationally bounded and also has its own sources of imperfections that we may (or may not) be aware of. However, ‘at the end of the day’, probabilities and their quantitative assessment, remain essential and necessary to provide a systematic approach to constructing a model of uncertainty. For this reason, it is important to know some of the assumptions we use in building probability models, as we shall briefly outline below. The approach we shall use is informal, however, emphasizing a study of models’ implications at the expense of formality.

## 4.2 UNCERTAINTY, GAMES OF CHANCE AND MARTINGALES

Games of chance, such as betting in Monte Carlo or any casino, are popular metaphors to represent the ongoing exchanges of stock markets, where money is thrown to chance. Its historical origins can be traced to Girolamo Cardano who proposed an elementary theory of gambling in 1565 (*Liber de Ludo Aleae* – The Book of Games of Chance). The notion of ‘fair game’ was clearly stated: ‘The most fundamental principle of all in gambling is simply equal conditions, e.g. of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. The extent to which you depart from that equality, if it is in your opponent’s

favour, you are a fool, and if in your own, you are unjust'. *This is the essence of the Martingale* (although Cardano did not use the word 'martingale'). It was in Bachelier's thesis in 1900 however that a mathematical model of a fair game, the martingale, was proposed. Subsequently J. Ville, P. Levy, J.L. Doob and others have constructed stochastic processes. The 'concept of a fair game' or martingale, in money terms, states that the expected profit at a given time given the total past capital is null with probability one. Gabor Szekely points out that a martingale is also a paradox. Explicitly,

If a share is expected to be profitable, it seems natural that the share is worth buying, and if it is not profitable, it is worth selling. It also seems natural to spend all one's money on shares which are expected to be the most profitable ones. Though this is true, in practice other strategies are followed, because while the expected value of our money may increase (our expected capital tends to infinity), our fortune itself tends to zero with probability one. So in Stock Exchange business, we have to be careful: shares that are expected to be profitable are sometimes worth selling.

Games of dice, blackjack, roulette and many other games, when they are fair, corrected for the bias each has, are thus martingales. 'Fundamental finance theory' subsumes as well that under certain probability measures, asset prices turn out to have the martingale property. Intuitively, what does a martingale assume?

- Tomorrow's price is today's best forecast.
- Non-overlapping price changes are uncorrelated at all leads and lags.

The martingale is considered to be a necessary condition for an efficient asset market, one in which the information contained in past prices is instantly, fully and perpetually reflected in the asset's current price. A technical definition of a martingale can be summarized as the presumption that each process event (such as a new price) is independent and can be summed (i.e. it is integrable) and has the property that its conditional expectation remains the same (i.e. it is time-invariant). That is, if  $\Phi_t = \{p_0, p_1, \dots, p_t\}$  are an asset price history at time  $t = 0, 1, 2, \dots$  expressing the relevant information we have at this time regarding the time series, also called the filtration. Then the expected next period price at time  $t + 1$  is equal to the current price

$$E(p_{t+1} | p_0, p_1, p_2, \dots, p_t) = p_t$$

which we also write as follows:

$$E(p_{t+1} | \Phi_t) = p_t \quad \text{for any time } t$$

If instead asset prices decrease (or increase) in expectation over time, we have a super-martingale (sub-martingale):

$$E(p_{t+1} | \Phi_t) \leq (\geq) p_t$$

Martingales may also be defined with respect to other processes. In particular, if  $\{p_t, t \geq 0\}$  and  $\{y_t, t \geq 0\}$  are two processes denoting, say, price and interest

rate processes, we can then say that  $\{p_t, t \geq 0\}$  is a martingale with respect to  $\{y_t, t \geq 0\}$  if:

$$E\{|p_t|\} < \infty \quad \text{and} \quad E(p_{t+1} | y_0, y_1, \dots, y_t) = p_t, \quad \forall t$$

Of course, by induction, it can be easily shown that a martingale implies an invariant mean:

$$E(p_{t+1}) = E(p_t) = \dots = E(p_0)$$

For example, given a stock and a bond process, the stock process may turn out to be a martingale with respect to the bond (a deflator) process, in which case the bond will serve as a numeraire facilitating our ability to compute the value of the stock.

Martingale techniques are routinely applied in financial mathematics and are used to prove many essential and theoretical results. For example, the first ‘fundamental theorem of asset pricing’, states that if there are no arbitrage opportunities, then properly normalized security prices are martingales under some probability measure. Furthermore, efficient markets are defined when the relevant information is reflected in market prices. This means that at any one time, the current price fully represents all the information, i.e. the expected future price  $p(t + T)$  conditioned by the current information and using a price process normalized to a martingale equals the current price. ‘The second fundamental theorem of asset pricing’ states in contrast that if markets are complete, then for each numeraire used there exists one and only one pricing function (which is the martingale measure). Martingales and our ability to construct price processes that have the martingale properties are thus extremely useful to price assets in theoretical finance as we shall see in Chapter 6.

Martingales provide the possibility of using a risk-neutral pricing framework for financial assets. Explicitly, when and if it can be used, it provides a mechanism for valuing assets ‘as if investors were risk neutral’. It is indeed extremely convenient, allowing the pricing of securities by using their expected returns valued at the risk-free rate. To do so, one must of course, find the probability measure, or equivalently find a discounting mechanism that renders the asset values a martingale. Equivalently, it requires that we determine the means to replicate the payoff of an uncertain stream by an equivalent ‘sure’ stream to which a risk-free discounting can be applied. Such a risk-neutral probability exists if there are no arbitrage opportunities. The martingale measures are therefore associated with a pricing of an asset which is unique only if markets are complete. This turns out to be the case when the assumptions made regarding market behaviours include:

- rational expectations,
- law of the single price,
- no long-term memory,
- no arbitrage.

The problem in applying rational expectations to financial valuation is that it may not be always right, however. The interaction of markets can lead to

instabilities due to very rapid and positive feedback or to expectations that are becoming trader- and market-dependent. Such situations lead to a growth of volatility, instabilities and perhaps, in some special cases, to bubbles and chaos. George Soros, the hedge fund financier has also brought attention to the concept of ‘reflexivity’ summarizing an environment where conventional traditional finance theory no longer holds and therefore theoretical finance does not apply. In these circumstances, ‘there is no hazard in uncertainty’. A trader’s ability to ‘identify a rational behaviour’ in what may seem irrational to others can provide great opportunities for profit making.

The ‘law of the single price’, claiming that two cash flows of identical characteristics must have, necessarily, the same price (otherwise there would be an opportunity for arbitrage) is not always satisfied as well. Information asymmetry, for example, may violate such an assumption. Any violation of these assumptions perturbs the basic assumptions of theoretical finance, leading to incomplete markets. In particular, we apply this ‘law’ in constructing portfolios that can replicate risky assets. By hedging, i.e. equating these portfolios to a riskless asset, it becomes possible to value the assets ‘as if they were riskless’. This approach will be developed here in greater detail and for a number of situations.

We shall attend to these issues at some length in subsequent chapters. At this point, we shall turn to defining terms often used in finance: random walks and stochastic processes.

### 4.3 UNCERTAINTY, RANDOM WALKS AND STOCHASTIC PROCESSES

A stochastic process is an indexed pair {events, time} expressed in terms of a function – a random variable indexed to time. This defines a sample path, i.e. a set of values that the process can assume over time. For example, it might be a stock price denoting events, indexed to a time scale. The study of stochastic processes has its origin in the study of the kinetic behaviour of molecules in gas by physicists in the nineteenth century. It was only in the twentieth century, following work by Einstein, Kolmogorov, Levy, Wiener and others, that stochastic processes were studied in some depth. In finance, however, Bachelier, in his dissertation in 1900, had already provided a study of stock exchange speculation using a fundamental stochastic process we call the ‘random walk’, establishing a connection between price fluctuations in the stock exchange and Brownian motion – a continuous-time expression of the random walk assumptions.

#### 4.3.1 The random walk

The random walk model of price change is based on two essential behavioural hypotheses.

- (1) In any given time interval, prices may increase with a known probability  $0 < p < 1$ , or decrease with probability  $1 - p$ .
- (2) Price changes from period to period are statistically independent.

Denote by  $\Delta\xi(t)$  the random event denoting the price change (of size  $\Delta x$ ) in a small time interval  $\Delta t$ :

$$\Delta\xi(t) = \begin{cases} +\Delta x & \text{w.p. } p \\ -\Delta x & \text{w.p. } 1-p \end{cases}$$

Thus, if  $x(t)$  is the price at the discrete time  $t$ , and if it is *only* a function of the last price  $x(t - \Delta t)$  and price changes  $\Delta\xi(t)$  in  $(t - \Delta t, t)$ , then an evolution of prices is given by:

$$x(t) = x(t - \Delta t) + \Delta\xi(t)$$

Prices are thus assuming values  $x(t)$  at times  $\dots, t - \Delta t, t, t + \Delta t, \dots$ . These values denote a stochastic process  $x(t)$  which is also written as  $\{x(t), t \geq 0\}$ . The price at time  $t$ ,  $x(t)$ , assumes in this case a binomial distribution since events are independent and of fixed probability, as we shall see next. Say that we start at a given price  $x_0$  at time  $t_0 = 0$ . At time  $t_1 = t_0 + \Delta t$ , either the price increases by  $\Delta x$  with probability  $0 < p < 1$  or it decreases with probability  $1 - p$ . Namely,

$$x(t_1) = x(t_1 - \Delta t) + \Delta\xi_1, \quad \text{or} \quad x(t_1) = x(t_0) + \Delta\xi_1.$$

We can also write this equation in terms of the number of times  $i_1$  the price increases. In our case, prices either increase or decrease in  $\Delta t$ , or

$$x(t_1) = x_0 + i_1\Delta x - (1 - i_1)\Delta x, \quad i_1 \sim B(1, p)$$

where  $i_1$  assumes two values  $i_1 = 0, 1$  given by the binomial probability distribution

$$B(1, p) = \binom{1}{i_1} p^{i_1} (1 - p)^{1-i_1}$$

$i_1 = 0, 1$  and parameter  $(1, p), 0 < p < 1$ . An instant of time later  $t_2 = t_1 + \Delta t = t_0 + 2\Delta t$ , we have:

$$x(t_2) = x(t_2 - \Delta t) + \Delta\xi_2 \text{ or } x(t_2) = x(t_1) + \Delta\xi_2 \text{ or } x(t_2) = x(t_0) + \Delta\xi_1 + \Delta\xi_2$$

which we can write as follows (see also Figure 4.1):

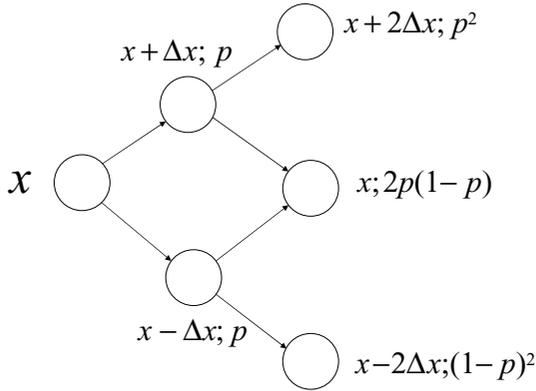
$$x(t_2) = x_0 + i_2\Delta x - (2 - i_2)\Delta x, \quad i_2 \sim B(2, p)$$

and generally, for  $n$  successive intervals of time ( $t_n = n\Delta t$ ), the price is defined by:

$$x(t_n) = x_0 + i_n\Delta x - (n - i_n)\Delta x, \quad i_n \sim B(n, p)$$

where:

$$B(n, p) = \binom{n}{i_n} p^{i_n} (1 - p)^{n-i_n}; \quad i_n = 0, 1, 2, \dots, n$$



**Figure 4.1** A two-period tree.

The price process can thus be written by:

$$x(t_n) = x(t_{n-1}) + \Delta\xi_n \quad \text{or} \quad x(t_n) = x(t_0) + \sum_{i=0}^n \Delta\xi_j$$

where  $x(t_n) - x(t_0)$  has the probability distribution of the sum  $\Delta\xi_j (j = 1, \dots, n)$ . Since price changes are of equal size, we can state that the number of times prices have increased is given by the binomial distribution  $B(n, p)$ . The expected price and its variance can now be calculated easily. The expected price at time  $t_n$  is:

$$E(x(t_n)) = x_0 + \Delta x E(i_n) - \Delta x E(n - i_n); \quad E(i_n) = np; \quad E(n - i_n) = n(1 - p)$$

Set  $d = [i_n - (n - i_n)] \Delta x$ . The mean distance and its variance, given by  $E(d)$  and  $\text{var}(d)$ , with  $q = 1 - p$  are then,

$$E(d) = n(p - q)\Delta x \quad \text{and} \quad \text{var}(d) = 4npq(\Delta x)^2$$

This is easily proved. Note that  $E(i) = np$  and  $\text{var}(i) = npq$ , with  $i$  replacing  $i_n$  for simplicity. Thus,

$$E(d) = E [i - (n - i)] \Delta x = E [2i - n] \Delta x = (2np - n)\Delta x = n(p - q)\Delta x$$

Also

$$\text{var}(d) = (\Delta x)^2 \text{var} [i - (n - i)] = (\Delta x)^2 \text{var} [2i - n] = 4npq(\Delta x)^2$$

The results above are expressed in terms of small distance (which we shall henceforth call states) increments  $\Delta x$  and small increments of time  $\Delta t$ . Letting these increments be very small, we can obtain continuous time and continuous state limits for the equation of motion. Explicitly, in a time interval  $[0, t]$ , let the number

of jumps be  $n$  and be given by  $n = [t/\Delta t]$ . When  $\Delta t$  is a small time increment, then (with  $t/\Delta t$  integer):

$$E(d) = \frac{t(p - q)\Delta x}{\Delta t}, \quad \text{var}(d) = \frac{4tpq(\Delta x)^2}{\Delta t}$$

For the problem to make sense, the limits of  $\Delta x/\Delta t$  and  $(\Delta x)^2/\Delta t$  as  $\Delta x \rightarrow 0$  must exist, however. In other words, we are specifying a priori that the stochastic process, has at the limit, finite mean and finite variance growth rates. If we let these limits be:

$$\lim_{\substack{\Delta t \rightarrow 0 \\ \Delta x \rightarrow 0}} \frac{\Delta x}{\Delta t} = 2C, \quad \lim_{\substack{\Delta t \rightarrow 0 \\ \Delta x \rightarrow 0}} \frac{(\Delta x)^2}{\Delta t} = 2D$$

It is also possible to express the probability of a price increase in terms of these parameters which we choose for convenience to be:

$$p = \frac{1}{2} + \frac{C}{2D}\Delta x,$$

Inserting this probability in the mean and variance equations, and moving to the limit, we obtain the mean and variance functions  $m(t)$  and  $\sigma^2(t)$  which are linear in time:

$$m(t) = 2Ct; \quad \sigma^2(t) = 2Dt$$

where  $C$  is called the ‘drift’ of the process expressing its tendency over time while  $D$  is its diffusion, expressing the process variability. The proof of these is simple to check. First note that:

$$E(d) = n(p - q)\Delta x = n\frac{C}{D}(\Delta x)^2$$

However,  $(\Delta x)^2 = 2D\Delta t$  and therefore,  $E(d) = 2Cn\Delta t = 2Ct$ . By the same token,

$$\text{var}(d) = 4npq(\Delta x)^2 = n\left[1 + \frac{C}{D}\Delta x\right]\left[1 - \frac{C}{D}\Delta x\right](\Delta x)^2$$

However, we also have  $(\Delta x)^2 = 4C^2(\Delta t)^2$ ,  $n = t/\Delta t$  which is inserted in the equation above to lead to:

$$\text{var}(d) = t\frac{(\Delta x)^2}{\Delta t}\left[1 - \frac{C^2}{D^2}(\Delta x)^2\right] = 2Dt\left[1 - \frac{2(\Delta x)^2(\Delta t)}{4(\Delta t)^2(\Delta x)^2}(\Delta x)^2\right]$$

since,

$$\left[1 - \frac{2(\Delta x)^2(\Delta t)}{4(\Delta t)^2(\Delta x)^2}(\Delta x)^2\right] = [1 - D\Delta t]$$

At the time limit, we obtain the variance  $\text{var}(d) = \sigma^2(t) = 2Dt$  stated above.

Since this limit results from limiting arguments to the underlying binomial process describing the random walk, we can conclude that the parameters  $(m(t), \sigma^2(t))$  are normally distributed, or:

$$f(x, t) = \frac{1}{\sqrt{2\pi}\sigma(t)} \exp \left\{ -\frac{1}{2} \frac{[x - m(t)]^2}{\sigma^2(t)} \right\}$$

This equation turns out to be also a particular solution of a partial differential equation expressing the continuous time–state evolution of the process probabilities and called the Fokker–Planck equation. Using the elementary observation that a linear transformation of normal random variables are also normal, we can write the price equation in terms of its drift  $2C$  and diffusion  $\sqrt{2D}$  (also called volatility), by:

$$\Delta x(t) = 2C\Delta t + \sqrt{2D}\Delta w(t), \quad \text{with } E(\Delta w(t)) = 0, \quad \text{var}(\Delta w(t)) = \Delta t$$

where  $\Delta w(t)$  is a normal probability distribution with zero mean and variance  $\Delta t$ . Such processes, in continuous time are called stochastic differential equations (SDEs) while the process  $\Delta x(t) = \Delta w(t)$  is called a Wiener (Levy) process. Finally, the integral of  $\Delta w(t)$ , or  $W(t)$ , is also known as Brownian motion which is essentially a zero mean normally distributed random variable with independent increments and a linear variance in time  $t$ . It is named after Robert Brown (1773–1858), a botanist who discovered the random motion of colloid-sized particles found in experiments performed in June–August 1825 with pollen. If we were to take a stock price, it would be interesting to estimate both the drift and the diffusion of the process. Would it fit? Would the residual error be indeed a normal probability distribution with mean zero, and a linear time variance with no correlation? Such a study would compare stock data taken every minute (tickertape), daily, weekly and monthly. Probably, results will differ according to the time scale taken for the estimation and thereby violate the assumptions of the model. Such studies are important in financial statistics when they seek to justify the assumption of ‘error normality’ in financial time series.

The Wiener process is of fundamental importance in mathematical finance because it is used to model the uncertainty associated with many economic processes. However, it is well known in finance that such a process underestimates the probability of the price not changing, and overestimates the mid-range value price fluctuations. Further, extreme price jumps are grossly underestimated by the Wiener (normal) process. The search for distributions that can truly reflect stock market behaviour has thus become an important preoccupation. Mandelbrot and Fama for example have suggested that we use Pareto–Levy distributions as well as leptokurtic distributions to describe the statistics of price fluctuations. Explicitly, say that a distribution has mean  $m$  and variance  $\sigma^2$  and define the following coefficients  $\zeta_1 = m_3/\sigma^3$  and  $\zeta_2 = m_4/\sigma^4 - 3$  where  $m_3$  and  $m_4$  are the third and the fourth moment respectively. The first index is an index of asymmetry pointing to leptokurtic distributions while the second is ‘an excess coefficient’ point to platokurtic distributions. For the Normal distribution we have  $\zeta_1 = 0$  and  $\zeta_2 = 0$ ,

thus any departure from these reference values will also indicate a departure from normality. Pareto–Levy stable distributions exhibit, however, an infinite variance, practically referred to as ‘fat tail distributions’ that also violate the underlying assumptions of ‘Normal–Wiener’ processes. When weekly or monthly data is used (rather than daily and intraday data), a smoothing of the data allows the use of the Normal distribution. This observation thus implies that the time scale we choose to characterize uncertainty is an important factor to deal with. When the time scale increases, the use of Normal distributions is justified because in such cases, we gradually move from leptokurtic to Normal distributions. What statistical distribution can one assume over different periods of consideration? The random walk is by far the most used and the easiest to work with and agrees well for larger periods of time. Other distributions are mathematically more challenging, especially since different results are seen for various assets. Part of the problem can be explained by the deviations from the *efficient markets hypothesis* and external influences on the market, as we shall see in subsequent chapters.

Formally, it is a Markov stochastic process  $x = \{x(t); t \geq 0\}$  whose non-overlapping increments  $\Delta x_t$  and  $\Delta x_s$

$$\Delta x(\tau) = x(\tau + \Delta t) - x(\tau); \tau = t, s$$

are *stationary, independently and normally distributed* with mean zero and variance  $\Delta t$ , i.e. with zero drift and volatility 1. In continuous time, this equation is often written as:

$$dx = 2C dt = \sqrt{2D} dw(t)$$

Such equations are known as *stochastic differential equations*. Generalization to far more complex movements can also be constructed by changing the modelling hypotheses regarding the drift and the diffusion processes. When the diffusion–volatility is also subject to uncertainty, this leads to processes we call stochastic volatility models, leading to incomplete markets (as will be seen in Chapter 5). In many cases, volatility can be a function of the process itself. For example, say that  $\sigma = \sigma(x)$ , then evidently,

$$\Delta x = \sigma(x)\Delta w$$

which need not lead, necessarily, to a Normal probability distribution for  $x$ . For example, in some cases, it is convenient to presume that rates of returns are Normal, meaning that  $\Delta x/x$  can be represented by a process with known drift (the expected rate of return) and known diffusion (the rates of returns volatility). Thus, the following hypothesis is stated:

$$\frac{\Delta x}{x} = \alpha \Delta t + \sigma \Delta w.$$

This is equivalent to stating that the log of return  $y = \ln(x)$  has a Normal probability distribution:

$$\Delta y = \alpha \Delta t + \sigma \Delta w, \quad y = \ln(x)$$

with mean  $\alpha t$  and variance  $\sigma^2 t$  and therefore,  $x$  has a lognormal probability distribution.

In many economic and financial applications stochastic processes are driven by a Wiener process leading to models of the form:

$$x(t + \Delta t) = x(t) + f(x, t)\Delta t + \sigma(x, t)\Delta w(t)$$

Of course, if the time interval is  $\Delta t = 1$ , this is reduced to a difference equation,

$$X_{t+1} = X_t + f_t(X) + \sigma_t(X)\varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad t = 0, 1, 2, \dots$$

where  $\varepsilon_t$  is a zero mean, unit variance and normally distributed random variable. When the time interval is infinitely small, in continuous time, we have a stochastic differential equation:

$$dx(t) = f(x, t) dt + \sigma(x, t) dw(t), \quad x(0) = x_0, \quad 0 \leq t \leq T$$

The variable  $x(t)$  is defined, however, only if the above equation is meaningful in a statistical sense. In general, existence of a solution for the stochastic differential equation cannot be taken for granted and conditions have to be imposed to guarantee that such a solution exists. Such conditions are provided by the Lipschitz conditions assuming that:  $f$ ,  $\sigma$  and the initial condition  $x(0)$  are real and continuous and satisfy the following hypotheses:

- $f$  and  $\sigma$  satisfy uniform Lipschitz conditions in  $x$ . That is, there is a  $K > 0$  such that for  $x_2$  and  $x_1$ ,

$$|f(x_2, t) - f(x_1, t)| \leq K |x_2 - x_1|$$

$$|\sigma(x_2, t) - \sigma(x_1, t)| \leq K |x_2 - x_1|$$

- $f$  and  $\sigma$  are continuous in  $t$  on  $[0, T]$ ,  $x(0)$  is any random variable with  $E(x(0))^2 < \infty$ , independent of the increment stochastic process. Then:
  - (1) The stochastic differential equation has, in the mean square limit sense, a solution on

$$t \in [0, T], \quad x(t) - x(0) = \int_0^t f(x, \tau) d\tau + \int_0^t \sigma(x, \tau) dw(\tau)$$

- (2)  $x(t)$  is mean square continuous on  $[0, T]$
- (3)  $E(x(0))^2 < M$ , for all  $t \in [0, T]$  and arbitrary  $M$ ,

$$\int_0^T E((x(t))^2) dt < \infty$$

- (4)  $x(t) - x(0)$  is independent of the stochastic process  $\{dw(\tau); \tau > t\}$  for  $t \in [0, T]$ .

The stochastic process  $x(t)$ ,  $t \in [0, T]$ , is then a Markov process and, in a mean square sense, is uniquely determined by the initial condition  $x(0)$ . The Lipschitz and the growth conditions, meaning  $(f(x, t))^2 + (\sigma(x, t))^2 \leq K^2(1 + |x|^2)$ , provide both a uniqueness and existence non-anticipating solution  $x(t)$  of the stochastic differential equation in the appropriate range  $[0, T]$ . In other words, if these conditions are not guaranteed, as is the case when the variance of processes increases infinitely, a solution to the stochastic differential equation cannot be assured.

Clearly, there is more than one way to conceive and formalize stochastic models of prices. In this approach, however, the evolution of prices was entirely independent of their past history. And further, a position at an instant of time depends only on the position at the previous instant of time. Such assumptions, compared to the real economic, financial and social processes we usually face, are extremely simplistic. They are, however, required for analytical tractability and we must therefore be aware of their limitations. The stringency of the assumptions required to construct stochastic processes, thus, point out that these can be useful to study systems which exhibit only small variations in time. Models with large and unpredictable variations must be based therefore on an intuitive understanding of the problem at hand or some other modelling techniques.

### 4.3.2 Properties of stochastic processes

The characteristics of time series are mostly expressed in terms of, ‘stationarity, ergodicity, correlation and independent increments’. These terms are often encountered in the study of financial time series and we ought therefore to understand them.

#### Stationarity

A time series is stationary when the evolution of its mean (drift) and variance (volatility–diffusion) are not a function of time. If  $f(x, t)$  is the probability distribution of  $x$  at time  $t$ , then:  $f(x, t) = f(x, t + \tau) = f(x)$  for all  $t$  and  $\tau$ . This property is called *strict stationarity*. In this case, for a two random variables process, we have:

$$f(x_1, x_2, t_1, t_2) = f(x_1, x_2, t_1, t_1 + \tau) = f(x_1, x_2, t_2 - t_1) = f(x_1, x_2, \tau)$$

That is, for the joint distribution of a strict stationary process, the distribution is a function of the time difference  $\tau$  of the two (prices) random variables. As a result, the correlation function  $B(t_1, t_2) = E(x(t_1)x(t_2))$ , describing the correlation between  $(x_1, x_2)$  at instants of time  $(t_1, t_2)$ , is a function of the time difference  $t_2 - t_1 = \tau$  only. The autocovariance function (the correlation function about the mean) is then given by  $K(t_1, t_2)$ , with  $K(t_1, t_2) = B(t_1, t_2) - E x_1(t_1) E x_2(t_2)$ . By the same token, the correlation coefficient  $R_1(\tau)$  of the random variable  $x_1$  is a

function of the time difference  $\tau$  only, or

$$R_1(\tau) = \frac{\text{cov}[x_1(t), x_1(t + \tau)]}{\text{var}[x_1(t)] \text{var}[x_1(t + \tau)]}$$

For stationary processes we have necessarily  $\text{var}[x(t)] = \text{var}[x(t + \tau)]$  and therefore the correlation coefficient is a function of the time difference only, or

$$R(\tau) = \frac{[B(\tau) - m]^2}{\text{var}[x(t)]} = \frac{K(\tau)}{K(0)}$$

### Independent increments

Increments  $\Delta x(t) = x(t + 1) - x(t)$  are stationary and independent if non-overlapping  $\Delta x(t)$  and  $\Delta x(s)$  are statistically, identically and independently distributed. This property leads to well-known processes such as the Poisson Jump and the Wiener process we saw earlier and can, sometimes, be necessary for the mathematical tractability of stochastic processes. The first two moments of non-overlapping independent and stationary increments point to a linear function of time (hence the term of linear finance, associated with using Brownian motion in financial model building). This is shown by the simple equalities:

$$\begin{aligned} E[X(t)] &= tE[X(1)] + (1 - t)E[X(0)]; \\ \text{var}[X(t)] &= t \text{var}[X(1)] + (1 - t) \text{var}[X(0)] \end{aligned}$$

The proof is straightforward and found by noting that if we set  $f(t) = E[X(t)] - E[X(0)]$ , then, non-overlapping stationary increments imply that:

$$\begin{aligned} f(t + s) &= E[X(t + s)] - E[X(0)] = E[X(t + s) - X(t)] + E[X(t) - X(0)] \\ &= E[X(s) - X(0)] + E[X(t) - X(0)] \\ &= f(t) + f(s) \end{aligned}$$

And the only solution is  $f(t) = tf(1)$ , which is used to prove the result for the expectation. The same technique applies to the variance.

## 4.4 STOCHASTIC CALCULUS

Financial and computational mathematics use stochastic processes extensively and thus we are called to manipulate equations of this sort. To do so, we mostly use Ito's stochastic calculus. The ideas of this calculus are simple and are based on the recognition that the magnitudes of second-order terms of asset prices are not negligible. Many texts deal with the rules of stochastic calculus, including Arnold (1974), Bensoussan (1982, 1985), Bismut (1976), Cox and Miller (1965), Elliot (1982), Ito (1961), Ito and McKean (1967), Malliaris and Brock (1982) and my own (Tapiero, 1988, 1998). For this reason, we shall consider here these rules in an intuitive manner and emphasize their application. Further, for simplicity, functions of time such as  $x(t)$  and  $y(t)$  are written by  $x$  and  $y$  except when the time specification differs.

The essential feature of Ito's calculus is Ito's Lemma. It is equivalent to the total differential rule in deterministic calculus. Explicitly, state that a functional relationship  $y = F(x, t)$ , continuous in  $x$  and time  $t$ , expresses the value of some economic variable  $y$  measured in terms of another  $x$  (for example, an option price measured in terms of the underlying stock price on which the option is written, the value of a bond measured as a function of the underlying stochastic interest rate process etc.) whose underlying process is known. We seek  $\Delta y = y(t + \Delta t) - y(t)$ . If  $x$  is deterministic, then application of the total differential rule in calculus, resulting from an application of Taylor series expansion of  $F(x, t)$ , provides the following relationship:

$$\Delta y = \frac{\partial F}{\partial t} \Delta t + \frac{\partial F}{\partial x} \Delta x$$

Of course, having higher-order terms in the Taylor series development yields:

$$\Delta y = \frac{\partial F}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 F}{\partial t^2} [\Delta t]^2 + \frac{\partial F}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} [\Delta x]^2 + \frac{\partial^2 F}{\partial t \partial x} [\Delta t \Delta x]$$

If the process  $\Delta x$  is deterministic, then obviously, terms of the order  $[\Delta t]^2$ ,  $[\Delta x]^2$  and  $[\Delta t \Delta x]$  are negligible relative to  $\Delta t$  and  $\Delta x$ , which leads us to the previous first-order development. However, when  $\Delta x$  is stochastic, with variance of order  $\Delta t$ , terms of the order  $[\Delta x]^2$  are non-negligible (since they are also of order  $\Delta t$ ). As a result, the appropriate development of  $F(x, t)$  leads to:

$$\Delta y = \frac{\partial F}{\partial t} \Delta t + \frac{\partial F}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} [\Delta x]^2$$

This is essentially Ito's differential rule (also known as Ito's Lemma), as we shall see below for continuous time and continuous state stochastic processes.

#### 4.4.1 Ito's Lemma

Let  $y = F(x, t)$  be a continuous, twice differentiable function in  $x$  and  $t$ , or  $\partial F/\partial t$ ,  $\partial F/\partial x$ ,  $\partial^2 F/\partial x^2$  and let  $\{x(t), t \geq 0\}$  be defined in terms of a stochastic differential equation with drift  $f(x, t)$  and volatility (diffusion)  $\sigma(x, t)$ ,

$$dx = f(x, t) dt + \sigma(x, t) dw, x(0) = x_0, 0 \leq t \leq T$$

then:

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x} dx + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dx)^2.$$

Or

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x} [f(x, t) dt + \sigma(x, t) dw] + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} [f(x, t) dt + \sigma(x, t) dw]^2$$

Neglecting terms of higher order than  $dt$ , we obtain Ito's Lemma:

$$dF = \left\{ \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x} f(x, t) + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 F}{\partial x^2} \right\} dt + \frac{\partial F}{\partial x} \sigma(x, t) dw$$

This rule is a ‘work horse’ of mathematical finance in continuous time. Note in particular, that when the function  $F(\cdot)$  is not linear, the volatility affects the process drift. Applications to this effect will be considered subsequently. Generalizing to multivariate processes is straightforward. For example, for a two-variable process,  $y = F(x_1, x_2, t)$  where  $\{x_1(t), x_2(t); t \geq 0\}$  are two stochastic processes while  $F$  admits first- and second-order partial derivatives, then the stochastic total differential yields:

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x_1} dx_1 + \frac{1}{2} \frac{\partial^2 F}{\partial x_1^2} (dx_1)^2 + \frac{\partial F}{\partial x_2} dx_2 + \frac{1}{2} \frac{\partial^2 F}{\partial x_2^2} (dx_2)^2 + \frac{\partial^2 F}{\partial x_1 x_2} (dx_1 dx_2)$$

in which case we introduce the appropriate processes  $\{x_1(t), x_2(t); t \geq 0\}$  and maintain all terms of order  $dt$ . For example, define  $y = x_1 x_2$ , then for this case:

$$\frac{\partial F}{\partial t} = 0; \frac{\partial F}{\partial x_1} = x_2; \frac{\partial^2 F}{\partial x_1^2} = 0; \frac{\partial F}{\partial x_2} = x_1; \frac{\partial^2 F}{\partial x_2^2} = 0; \frac{\partial^2 F}{\partial x_1 x_2} = 1$$

which means that:

$$dF = x_2 dx_1 + x_1 dx_2 + dx_1 dx_2$$

Other examples will be highlighted through application in this and subsequent chapters. Below, a number of applications in economics and finance are considered.

## 4.5 APPLICATIONS OF ITO’S LEMMA

### 4.5.1 Applications

The examples below can be read after Chapter 6, in particular the applications of the Girsanov Theorem and Girsanov and the binomial process.

#### (a) The Ornstein–Uhlenbeck process

The Ornstein-Uhlenbeck process is a process used in many circumstances to model mean returning processes. It is given by the following stochastic differential equation:

$$dx = -ax dt + \sigma dw(t), \quad a > 0$$

We shall show first that the process has a Normal probability distribution and solve the equation by an application of Ito’s Lemma. Let  $y(t) = e^{at} x(t)$  and apply Ito’s differential rule to lead to:

$$dy = \sigma e^{at} dw(t)$$

An integration of the above equation with substitution of  $y$  yields the solution:

$$x(t) = x(0) e^{-at} + \sigma \int_0^t e^{-a(t-\tau)} dw(\tau)$$

The meaning of this equation is that the process  $x(t)$  is an exponentially weighted function of past noise. Note that the transformed process  $y(t)$  has a constant mean since its mean growth rate is null, or  $E(dy) = E[\sigma e^{at} dw(t)] = \sigma e^{at} E[dw(t)] = 0$ . In other words, the exponential growth process  $y(t) = e^{at} x(t)$  is a constant mean process.

**(b) The wealth process of a portfolio of stocks**

Let  $x$  be the invested wealth of an investor at a given time  $t$  and suppose that in the time interval  $(t, t + dt)$ ,  $c dt$  is consumed while  $y dt$  is the investor's income from both investments and other sources. In this case, the rate of change in wealth is equal  $x(t + dt) = x + [y - c] dt$ . In order to represent this function in terms of investment assets, say that all our wealth is invested in stocks. The price of a stock, denoted by  $S_i$ ,  $i = 1, 2, \dots, n$  and the number of stocks  $N_i$  held of each type  $i$ , determines wealth as well as income. If income is measured only in terms of price changes (i.e. we do not include at this time borrowing costs, dividend payments for holding shares etc.), then income  $y dt$  in  $dt$  is necessarily:

$$y dt = \sum_{i=1}^n N_i dS_i$$

and therefore the investor's worth is:

$$dx = \sum_{i=1}^n N_i dS_i - c dt$$

For example, assume that prices are lognormal, given by:

$$\frac{d S_i}{S_i} = \alpha_i dt + \sigma_i dw_i, S_i(0) = S_{0,i} \quad \text{given, } i = 1, 2, \dots, n$$

where  $w_i(t)$  are standard Wiener processes (that may be independent or not, in which case they are assumed to be correlated). Inserting into the wealth process equation, we obtain:

$$\begin{aligned} dx &= \sum_{i=1}^n N_i [\alpha_i S_i dt + \sigma_i S_i dw_i] - c dt = \sum_{i=1}^n [\alpha_i N_i S_i - c] dt \\ &+ \sum_{i=1}^n [\sigma_i N_i S_i dw_i] \end{aligned}$$

This is of course, a linear stochastic differential equation. Simplifications to this equation can be reached, allowing a much simpler treatment. For example, say

that a proportion  $\theta_i$  of the investor wealth is invested in stock  $i$ . In other words,

$$N_i S_i = \theta_i x \text{ or } N_i = \theta_i x / S_i \quad \text{with} \quad \sum_{i=1}^n \theta_i = 1$$

Further, say that the Wiener processes are uncorrelated, in which case:

$$\sum_{i=1}^n [\sigma_i N_i S_i dw_i] = \sum_{i=1}^n [\sigma_i \theta_i x dw_i] = x \sqrt{\sum_{i=1}^n [\sigma_i^2 \theta_i^2]} dw$$

which leads to:

$$dx = \left[ \sum_{i=1}^n \alpha_i \theta_i x - c \right] dt + x \sqrt{\sum_{i=1}^n [\sigma_i^2 \theta_i^2]} dw, \quad \sum_{i=1}^n \theta_i = 1$$

Thus, by selecting trading and consumption strategies represented by  $\theta_i$ ,  $i = 1, 2, \dots, n$  we will, in fact, also determine the evolution of the (portfolio) wealth process.

#### 4.5.2 Time discretization of continuous-time finance models

When the underlying model is given in a continuous time and in a continuous state framework, it is often useful to use discrete models as an approximation. There are a number of approaches to doing so. Discretization can be reached by discretizing the state space, the time or both. Assume that we are given a stochastic differential equation (SDE). A time (process) discretization might lead to a simple stochastic difference equation or to a stochastic difference equation subject to multiple sources of risk, as we shall see below. A state discretization means that the underlying process is represented by discrete state probability models (such as a binomial random walk, a trinomial walk, or Markov chains and their like). These approaches will be considered below.

##### (a) Probability approximation (discretizing the states)

In computational finance, numerical techniques are sought that make it possible also to apply risk-neutral pricing, in other words approximate the process by using binomial models or other models with desirable mathematical characteristics that allow the application of fundamental finance theories. Approximations by binomial trees are particularly important since many results in fundamental finance are proved and explained using the binomial model. It would therefore be useful to define a sequence of binomial processes that converge weakly (at least) to diffusion–stochastic differential equation models. Nelson and Ramaswamy (1990) have suggested such approximations, which are in fact similar to a drift–volatility approximation. Let  $x$  be the current price of a stock and let its next price (in the discretized model with time intervals  $h$ ) be either  $X^+(x, t)$  or  $X^-(x, t)$ .

We denote by  $P(x, t)$ , the probability of transition to state  $X^+(x, t)$ . Or

$$X_t = x; X_{t+1} = \begin{cases} [X^+(x, t)] & \text{with Pr } P(x, t) \\ [X^-(x, t)] & \text{with Pr } 1 - P(x, t) \end{cases}$$

Thus,

$$X_{t+1} - X_t = \Delta x = \begin{cases} [X^+(x, t) - x] & \text{with Pr } P(x, t) \\ [X^-(x, t) - x] & \text{with Pr } 1 - P(x, t) \end{cases}$$

For a financial process defined by a stochastic differential equation with drift  $\mu(x, t)$  and volatility  $\sigma(x, t)$  we then have:

$$\begin{aligned} \mu(x, t)h &= P(x, t)[X^+(x, t) - x] + [1 - P(x, t)][X^-(x, t) - x] \\ \sigma^2(x, t)h &= P(x, t)[X^+(x, t) - x]^2 + [1 - P(x, t)][X^-(x, t) - x]^2 \end{aligned}$$

This is a system of two equations in the three values  $P(x, t)$ ,  $[X^-(x, t)]$  and  $[X^+(x, t)]$  that the discretized scheme requires. These values can be defined in several ways. Explicitly, we can write:

$$\begin{aligned} X^+(x, t) &\equiv x + \sigma\sqrt{h}; \quad X^-(x, t) \equiv x - \sigma\sqrt{h}; \\ P(x, t) &= 1/2 + \sqrt{h}\mu(x, t)/2\sigma(x, t) \end{aligned}$$

And, therefore, an approximate binomial tree can be written as follows:

$$\Delta x = \begin{cases} \sigma\sqrt{h} & \text{with Pr } [1/2 + \sqrt{h}\mu(x, t)/2\sigma(x, t)] \\ -\sigma\sqrt{h} & \text{with Pr } [1/2 - \sqrt{h}\mu(x, t)/2\sigma(x, t)] \end{cases}$$

For example, say that  $Hx = X^+(x, t)$  and  $Lx = X^-(x, t)$  as well as  $p = P(x, t)$ . Then:

$$X_{t+1} = \begin{cases} Hx & \text{with Pr } p \\ Lx & \text{with Pr } 1 - p \end{cases}; \quad X_t = x$$

Let  $i$  be the number of times the price (process) increases over a period of time  $T$ , then, the price distribution at time  $T$  is:

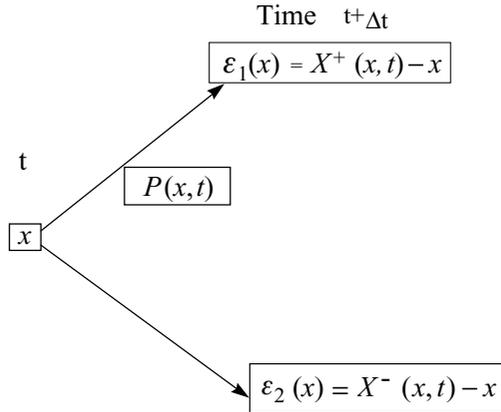
$$\frac{X_T}{x} = (H)^i(L)^{T-i} \text{ w.p. } P_i = \binom{T}{i} p^i(1-p)^{T-i} \quad i = 0, 1, 2, 3, \dots, T$$

Consider now the mean reverting (Ornstein–Uhlenbeck) process often used to model interest rate and volatility processes:

$$dx = \beta(\alpha - x) dt + \sigma dw, \quad x(0) = x_0 > 0, \quad \beta > 0, \quad x \in [0, 1]$$

Then, we define the following binomial process as an approximation:

$$\begin{aligned} X^+(x, t) &\equiv x + \sigma\sqrt{h}; \quad X^-(x, t) \equiv x - \sigma\sqrt{h}; \\ P(x, t) &= 1/2 + \sqrt{h}\mu(x, t)/2\sigma(x, t) \end{aligned}$$



**Figure 4.2** A binomial tree.

with the explicit transition probability given by:

$$P(x, t) = \begin{cases} 1/2 + \sqrt{h}\beta(\alpha - x)/2\sigma & \text{if } 0 \leq 1/2 + \sqrt{h}\beta(\alpha - x)/2\sigma \leq 1 \\ 0 & \text{if } 1/2 + \sqrt{h}\beta(\alpha - x)/2\sigma < 0 \\ 1 & \text{otherwise} \end{cases}$$

The probability  $P(x, t)$  is chosen to match the drift, it is censored if it falls outside the boundaries  $[0, 1]$ . As a result, the basic building block of the binomial process will be given as shown in Figure 4.2. In order to construct a simple (plain vanilla) binomial process it is essential that the volatility be constant, however. Otherwise the process will exhibit conditional heteroscedasticity. In the example above, this was the case and therefore we were able to define the states and the transition probability simply. When this is not the case, we can apply Ito's differential rule and find the proper transformation that will 'purge' this heteroscedasticity. Namely, we consider the transformation  $y(x, t)$  to which we apply Ito's Lemma:

$$dy = \left\{ \frac{\partial y}{\partial t} + \frac{\partial y}{\partial x} f(x, t) + \frac{1}{2} \sigma^2 \frac{\partial^2 y}{\partial x^2} \right\} dt + \frac{\partial y}{\partial x} \sigma(x, t) dw$$

and choose:

$$y(x, t) = \int^x \frac{dz}{\sigma(z, t)}$$

in which case, the term

$$\left( \sigma(x, t) \frac{\partial y(x, t)}{\partial x} \right) dw$$

is replaced by  $dw$  and the instantaneous volatility of the transformed process is constant. This allows us to obtain a computationally simple binomial tree. To see how to apply this technique, we consider another example. Consider the CEV

stock price:

$$dx = \mu x dt + \sigma x^\gamma dw; \quad 0 < \gamma < 1 \quad \text{and} \quad y(x, t) \equiv \sigma^{-1} \int^x z^{-\gamma} dz = \frac{x^{1-\gamma}}{\sigma(1-\gamma)}$$

which has the effect of reducing the stochastic differential equation to a constant volatility process, thus making it possible to transform it into a simple binomial process with an inverse transform given by:

$$x(y, t) = \begin{cases} [\sigma(1-\gamma)x]^{1/(1-\gamma)} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

**(b) The Donsker Theorem**

A justification of this approximation based on binomial trees can be made using the Donsker Theorem which is presented intuitively below. Given that the process we wish to represent is a trinomial random walk (or a simple random walk), we represent for convenience the transition probabilities as follows:

$$p_n = \frac{1-r}{2} + \frac{\alpha}{2n}; \quad q_n = \frac{1-r}{2} - \frac{\alpha}{2n}$$

with  $\alpha \in R, 0 \leq r < 1$  real numbers and  $n$  is a parameter representing the number of segments  $n^2$  used in dividing a given time interval, assumed to be large. Note that these (Markov) transition probabilities satisfy:  $p_n \geq 0, q_n \geq 0, r + p_n + q_n = 1$ . For each partitioning of the process, we associate the random walk  $(X_t^{(n)}; t \in T)$  which is a piecewise linear approximation of the stochastic process where the time interval  $[0, t]$  is divided into equal segments each of width  $1/n^2$ . At the  $k$ th segment, we have the following states:

$$X_{k/n^2}^{(n)} = \frac{1}{\sqrt{n^2}} X_k = X^{(k)}([n^2 t])$$

$$X_t^{(n)} = X_{k/n^2}^{(n)} + (n^2 t - [n^2 t])(X_{(k+1)/n^2}^{(n)} - X_{k/n^2}^{(n)}) \forall t \in \left[ \frac{k+1}{n^2}, \frac{k}{n^2} \right]$$

Note that the larger  $n$ , the greater the number of states and thereby, the more refined the approximation. The corresponding continuous (Brownian motion) process (a function of  $n$ ) is then assumed given by:

$$B^n(t) = \frac{1}{n} (X^{(n)}([n^2 t]) + (n^2 t - [n^2 t])\varepsilon_{([n^2 t+1])}); \quad t \geq 0$$

[.] denotes the integer value of its argument and  $\varepsilon_{(\cdot)}$  is the random walk with transition probabilities  $(p_n, q_n)$ . Note that by writing the continuous process in this form, we essentially divide the time scale in widths of  $1/n^2$ . Of course, when  $n$  is large, then time will become approximately continuous. The Donsker Theorem essential statement is that when  $n$  is large, the approximating random walk converge to a Brownian motion. Note that we can easily deduce that for  $n$

large, the following moments:

$$E \left[ \frac{1}{n} X^{(n)}([n^2 t]) \right] = \frac{\alpha}{n^2} [n^2 t];$$

$$\text{var} \left\{ \frac{1}{n} X^{(n)}([n^2 t]) \right\} = \frac{[n^2 t]}{n^2} \left( r(1-r) + (1-r)^2 - \frac{\alpha^2}{n^2} \right)$$

Consequently, for a fixed time, at the limit ( $[n^2 t] \underset{t \rightarrow \infty}{\approx} n^2 t$ ,  $t$  fixed):

$$\text{Lim}_{n \rightarrow +\infty} E \left[ \frac{1}{n} X^{(n)}([n^2 t]) \right] = \alpha t$$

$$\text{Lim}_{n \rightarrow \infty} \text{Var} \left\{ \frac{1}{n} X^{(n)}([n^2 t]) \right\} = t(r(1-r) + (1-r)^2) = (1-r)t$$

Furthermore, the last term in  $B^{(n)}(t)$  disappears since  $|(n^2 t - [n^2 t])\varepsilon_{([n^2 t]+1)}| \leq 1$ .

These allow the application of the Donsker Theorem stating that the process  $B^{(n)}$  converges in distribution to  $(\sqrt{1-r}B_t + \alpha t; t \geq 0)$  where  $(B_t; t \geq 0)$  designates a real and standardized Brownian motion starting at 0. This result, which is derived formally, justifies the previous continuous approximation for the random walk. Note that when  $r$  (the probability of remaining in the same state) is large, then we obtain a process with drift  $\alpha$  and zero variance. However, when  $n$  is small (close to zero or equal to zero) we obtain a Brownian motion with drift. Finally, we can proceed in a similar manner and calculate the empirical variance process of a trinomial random walk, providing therefore a volatility approximation as well. This is left as an exercise.

### (c) Discrete time approximations

In some cases, we approximate an underlying price stochastic differential equation by a stochastic difference equation. There are a number of ways to do so, however. As a result, a unique (price) valuation process based on a continuous-time model might no longer be unique in its 'difference' form. This will be seen below by using an approximation due to Milshtein (1974) which uses principles of stochastic integration in the relevant approximate discretized time interval. Thus, unlike application of the Donsker Theorem which justified the partitioning of the price (state) process, we construct below a discrete time process and apply Ito's differential rule within each time interval in order to estimate the evolution of prices (states) within each interval. The approximation within a time interval is achieved by a Taylor series approximation which can be linear, or of higher order. Integration using Ito's calculus provides then an estimate of the states in the discretized scheme. As we shall see below, this has the effect of introducing a process uncertainty which is not normal, thereby leading to stochastic volatility. The higher the order approximation the larger the number of uncertainty sources. To see how this is defined, we consider first the following Ito differential equation:

$$dx = f(x, t) dt + \sigma(x, t) dw(s)$$

Consider next two subsequent instants of time  $t$  and  $r, r > t$ , then a discretized process in the interval  $r - t, r > t$  can be written as follows:

$$x_r = x_t + \alpha(r, x_t, r - t) + \beta(r, x_t, r - t, w_r - w_t)$$

where  $\alpha(r, x_t, r - t), \beta(r, x_t, r - t, w_r - w_t)$  are a drift, a function of the beginning state and the end state as well as as a function of time, the time interval and the volatility which is, in addition, a function of the uncertainty in the time interval. Note the volatility function need not be a linear function in  $(w_r - w_t)$ . In a stochastic integral form, the evolution of states is:

$$x_t = x_s + \int_s^t f(r, x_r) dr + \int_s^t \sigma(x_r, r) dw_r$$

If we use in discrete time only a first-order (linear) approximation to the drift and the volatility, we have:

$$x_r \cong x_t + f(s, x_t) \Delta t + \sigma(t, x_t) \Delta w_t, \\ \Delta t = (r - t); \Delta w_t = (w_r - w_t), \alpha(.) = f(s, x_t) \Delta t; \beta(.) = \sigma(s, x_s) \Delta w_t$$

When we take a second-order approximation, an additional source of uncertainty is added. To see how this occurs, consider a more refined approximation based on a Taylor series expansion of the first two terms for the functions  $(f, \sigma)$ . Then in the time interval  $(t, s; t > s)$  :

$$x_t = x_s + \int_s^t \left[ f(s, x_s) + \frac{\partial f(s, x_s)}{\partial t}(r - s) + \frac{\partial f(s, x_s)}{\partial x} (f(s, x_s)(r - s) + \sigma(s, x_s)(w_r - w_s)) \right] dr + \int_s^t \left[ \sigma(s, x_s) + \frac{\partial \sigma(s, x_s)}{\partial t}(r - s) + \frac{\partial \sigma(s, x_s)}{\partial x} (f(s, x_s)(r - s) + \sigma(s, x_s)(w_r - w_s)) \right] dw_r$$

Since the stochastic integral is given by:

$$x_t = x_s + \int_s^t [w_r - w_s] dw_r = \frac{1}{2}[(w_t - w_s)^2 - (t - s)]$$

Including terms of order  $(t - s)$  only, we obtain:

$$x_t = x_s + f(s, x_s)(t - s) + \sigma(s, x_s)(w_t - w_s) + \frac{1}{2} \frac{\partial \sigma(s, x_s)}{\partial x} [(w_t - w_s)^2 - (t - s)]$$

or:

$$x_t = x_s + f(s, x_s) \Delta t + \sigma(s, x_s) \Delta w_s + \frac{1}{2} \frac{\partial \sigma(s, x_s)}{\partial x} [(\Delta w_s)^2 - \Delta t]$$

Note that in this case, there are two sources of uncertainty. First we have a normal term  $(w_t - w_s)$  of mean zero and variance  $t - s$ , while we also have a chi-square term defined by  $(w_t - w_s)^2$ . Their sum produces, of course, a nonlinear (stochastic) model.

An improvement of this approximation can further be reached if we take a higher-order Taylor series approximation. Although this is cumbersome, we summarize the final result here that uses the following stochastic integral relations (which are given without their development):

$$\begin{aligned} \int_s^t (r-s) dr &= \int_0^{t-s} u du = \frac{1}{2}(t-s)^2 = \frac{1}{2}(\Delta t)^2 \\ \int_s^t (r-s)^2 dr &= \int_0^{t-s} u^2 du = \frac{(t-s)^3}{3} = \frac{(\Delta t)^3}{3} \\ \int_s^t (r-s)(w_r - w_s) dr &= \int_0^{t-s} \tau(w_r - w_s) d\tau = \int_0^{t-s} \tau \int_0^v dw(v) d\tau \\ &= \int_0^{t-s} dw(v) \int_v^{t-s} \tau d\tau = \frac{1}{2} \int_0^{t-s} (\Delta t^2 - v^2) dw(v) \end{aligned}$$

which has Normal probability distribution with mean zero and variance:

$$\begin{aligned} \text{var}(x(t)) &= \frac{1}{4} \int_0^{t-s} [\Delta t^2 - u^2]^2 du = \frac{1}{4} \int_0^{t-s} [\Delta t^4 + u^4 - 2\Delta t^2 u^2] du = \frac{\Delta t^5}{6} \\ \int_s^t w(\tau) dw(\tau) &= \frac{1}{2}(w(t)^2 - w(s)^2) - \frac{1}{2}(t-s) \\ \int_s^t w(\tau)^2 dw(\tau) &= \frac{1}{3}(w(t)^3 - w(s)^3) - \int_s^t w(\tau) d\tau \\ &= \frac{1}{3}(w(t)^3 - w(s)^3) - \frac{1}{2}(w(t)^2 - w(s)^2) + \frac{1}{2}(t-s) \end{aligned}$$

And therefore we have at last:

$$\begin{aligned}
 x_t &= x_s + [f(s, x_s) + \sigma(s, x_s)] \Delta t + \\
 &+ \frac{1}{2} \left[ \frac{\partial f(s, x_s)}{\partial t} + \frac{\partial \sigma(s, x_s)}{\partial t} + \right. \\
 &\quad \left. \frac{\partial f(s, x_s)}{\partial x} f(s, x_s) + \frac{\partial \sigma(s, x_s)}{\partial x} f(s, x_s) \right] (\Delta t)^2 \\
 &+ \frac{1}{2.3} \left[ \frac{\partial^2 f(s, x_s)}{\partial t^2} + \frac{\partial^2 \sigma(s, x_s)}{\partial t^2} + \frac{\partial^2 \sigma(s, x_s)}{\partial x^2} f(s, x_s) \right] (\Delta t)^3 \\
 &+ \frac{1}{2} \left[ \frac{\partial f(s, x_s)}{\partial x} + \frac{\partial \sigma(s, x_s)}{\partial x} \right] \sigma(s, x_s) [(\Delta w_s)^2 - \Delta t] \\
 &+ \frac{1}{2.3} \left[ \frac{\partial^2 f(s, x_s)}{\partial x^2} \sigma^2(s, x_s) + \right. \\
 &\quad \left. \frac{\partial^2 \sigma(s, x_s)}{\partial x^2} \sigma^2(s, x_s) \right] \{ (w_t^3 - w_s^3) - 3(w_t^2 - w_s^2) + 3\Delta t \} \\
 &+ \frac{\partial^2 \sigma(s, x_s)}{\partial x^2} f(s, x_s) \sigma(s, x_s) \frac{(\Delta t)^{5/2}}{\sqrt{6}} \Delta w_s
 \end{aligned}$$

This scheme allows the numerical approximation of the stochastic differential equation and clearly involves multiple sources of risk.

For the lognormal price process:

$$dx = \alpha x dt + \beta x dw$$

The transformation of the Ito stochastic differential equation becomes:

$$d(\log x) = \left[ \alpha - \frac{\beta^2}{2} \right] dt + \beta dw_s$$

And therefore, a finite differencing, based on integration over a unit interval yields:

$$\log x_t - \log x_{t-1} = \left[ \alpha - \frac{\beta^2}{2} \right] + \beta \varepsilon_t; \quad \varepsilon_t \equiv (W_t - W_{t-1})$$

which can be used now to estimate the model parameters using standard statistical techniques. Interestingly, if we consider other intervals, such as smaller length intervals (as would be expected in intraday data), then the finite difference model would be instead:

$$\Delta_\tau = \log x_t - \log x_{t-\tau} = \left[ \alpha - \frac{\beta^2}{2} \right] + \beta \sqrt{\tau} \varepsilon_t^\tau; \quad \varepsilon_t^\tau \sim N(0, 1)$$

Of course, the estimators of the model parameters will be affected by this discretization. Higher-order discretization can be used as well to derive more precise results, albeit these results might not allow the application of standard fundamental finance results.

### 4.5.3 The Girsanov Theorem and martingales\*

The Girsanov Theorem is important to many applications in finance. It defines a ‘discounting process’ which transforms a given price process into a martingale. A martingale essentially means, as we saw it earlier, that any trade or transaction will be ‘fair’ in the sense that the expected value of any such transaction is null. And therefore its ‘transformed price’ remains the same. In particular, define the transformation:

$$L = \exp \left\{ \int_0^t \sigma(s) dW(s) - \frac{1}{2} \int_0^t \sigma^2(s) ds \right\}$$

where  $\sigma(s)$ ,  $0 \leq s \leq T$  is bounded and is the unique solution of:

$$\frac{dL}{L} = \sigma dW; L(0) = 1; E(L) = 1, \forall t \in [0, T]$$

and  $L(t)$  is a martingale. The proof can be found, by an application of Ito’s differential rule with  $y = \log L$ , or  $dy = dL/L - (1/2L^2)(dL)^2 = \sigma dW - (1/2)\sigma^2 dt$  whose integration leads directly to:

$$y - y_0 = \ln(L) - \ln(L_0) = \int_0^t \sigma(s) dW(s) - \frac{1}{2} \int_0^t \sigma^2(s) ds$$

And therefore,

$$L(0) = L(t) \exp \left\{ - \int_0^t \sigma(s) dW(s) + \frac{1}{2} \int_0^t \sigma^2(s) ds \right\}, L(0) = 1$$

In this case, note that  $y$  is a process with drift while that of  $L$  has no drift and thus  $L$  is a martingale. For example, for the lognormal stock price process:

$$\frac{dS}{S} = \alpha dt + \sigma dW, S(0) = S_0$$

Its solution at time  $t$  is simply:

$$S(t) = S(0) e^{\alpha t} \exp \left[ -\frac{\sigma^2}{2} t + \sigma W(t) \right], W(t) = \int_0^t dW(t)$$

which we can write in terms of  $L(t)$  as follows:

$$S(t) = S(0) e^{\alpha t} \frac{L(t)}{L(0)} \text{ or } \frac{S(t)}{L(t)} = e^{\alpha t} \frac{S(0)}{L(0)} \text{ or } \frac{S(0)}{L(0)} = e^{-\alpha t} \frac{S(t)}{L(t)}$$

Note that the term  $S(t)/L(t)$  is no longer stochastic and therefore no expectation is taken (alternatively, a financial manager would claim that, since it is deterministic,

its future value ought to be at the risk-free rate). By the same token, consider both a stock price and a risk-free bond process given by:

$$\frac{dS}{S} = \alpha dt + \sigma dW, \quad S(0) = S_0; \quad \frac{dB}{B} = R_f dt, \quad B(0) = B_0$$

If the bond is defined as the numeraire, then we seek a transformation such that  $S(t)/B(t)$  is a martingale. For this to be the case, we will see that the stock price is then given by:

$$\frac{dS}{S} = R_f dt + \sigma dW^*, \quad S(0) = S_0$$

where  $dW^*$  is the martingale measure which is explicitly given by:

$$dW^* = dW + \frac{\alpha - R_f}{\sigma} dt$$

In this case, under such a transformation, we can apply the risk-neutral pricing formula:

$$S(0) = e^{-R_f t} E^*(S(t))$$

where  $E^*$  denotes expectation with respect to the martingale measure.

### Problem

Let  $y = x_1/x_2$  where  $\{x_1(t), x_2(t); t \geq 0\}$  are two stochastic processes, each with known drift and known diffusion, then show that the stochastic total differential yields:

$$dy = \frac{1}{x_2} dx_1 - \frac{x_1}{x_2^2} dx_2 + \frac{x_1}{x_2^3} (dx_2)^2 - \frac{1}{x_2^2} (dx_1 dx_2)$$

Further, show that if we use the stock price and the bond process, it is reduced to:

$$dy = \frac{dS}{B(t)} - \frac{S(t)}{B^2(t)} dB + \frac{S(t)}{B^3(t)} (dB)^2 - \frac{dS dB}{B^2(t)} = \frac{dS}{B(t)} - \frac{S(t)}{B^2(t)} dB$$

And as a result,

$$\frac{dy}{y} = (\alpha - R_f) dt + \sigma dW$$

Finally, replace  $dW$  by

$$-\frac{\alpha - R_f}{\sigma} dt + dW^* = dW \text{ and obtain: } \frac{dy}{y} = \sigma dW^*$$

which is of course a martingale under the transformed measure.

**Martingale examples\***

(1). Let the stock price process be defined in terms of a Bernoulli event where stock prices grow from period to period at rates  $a > 1$  and  $b < 1$  with probabilities defined below:

$$S_t = \begin{cases} aS_{t-1} & \text{w.p. } \frac{1-b}{a-b} \\ bS_{t-1} & \text{w.p. } \frac{a-1}{a-b} \end{cases}$$

This process is a martingale. First it can be summed and, further, we have to show that this a constant mean process with:

$$E(S_{t+1} | S_t, S_{t-1}, \dots, S_0) = E(S_{t+1} | S_t) = S_t$$

Explicitly, we have:

$$E(S_{t+1} | S_t) = aS_t \frac{1-b}{a-b} + bS_t \frac{a-1}{a-b} = S_t \frac{b(a-1) + a(1-b)}{a-b} = S_t$$

By the same token, we can show that there are many other processes that have the martingale property. Consider the trinomial (birth–death) random walk defined by:

$$\begin{cases} S_{t+1} = S_t + \varepsilon_t, \quad S_0 = 0 \\ \varepsilon_t = \begin{cases} +1 & \text{w.p. } p \\ 0 & \text{w.p. } r \\ -1 & \text{w.p. } q \end{cases} \end{cases} ; p \geq 0, q \geq 0, r \geq 0, p + q + r = 1$$

Then it is easy to show that  $[S_t - t(p - q); t \geq 0]$  is a martingale. To verify this assertion, note that:

$$\begin{aligned} E[S_{t+1} - (t+1)(p - q) | \Phi_t] &= E[S_t + \varepsilon_t - (t+1)(p - q) | \Phi_t] \\ &= S_t - (t+1)(p - q) + E(\varepsilon_t) = S_t - (t+1)(p - q) + (p - q) \\ &= S_t - t(p - q) \end{aligned}$$

where  $\Phi_t \equiv (S_0, S_1, S_2, \dots, S_t)$  resumes the information set available at time  $t$ . Further:

$$[S_t^2 - t(p - q); t \geq 0], \{[S_t^2 - t(p - q)]^2 + t[(p - q)^2 - (p + q)]; t \geq 0\}$$

and  $\{\lambda^{x_t}, \lambda = q/p, t \geq 0\}$ ,  $p > 0$  are also martingales. The proof is straightforward and a few cases are treated below. By the same token, we can also consider processes that are not martingales and then find a transformation or another process that will render the original process a martingale.

(2) *The Wiener process is a martingale\**. The Wiener process  $\{w(t), t \geq 0\}$  is a Markov process and as we shall see below a martingale. Let  $\mathbf{F}(t)$  be its filtration, in other words, it defines the information set available at time  $t$  on which a conditional expectation is calculated (and on the basis of which financial calculations are

assumed to be made). Then, we can state that the Wiener process is a martingale with respect to its filtration. The proof is straightforward since

$$E \{w(t + \Delta t) | \mathbf{F}(t)\} = E \{w(t + \Delta t) - w(t) | \mathbf{F}(t)\} + E \{w(t) | \mathbf{F}(t)\}$$

The first term is null while the independent conditional increments of the Wiener process imply:

$$E \{w(t + \Delta t) - w(t) | \mathbf{F}(t)\} = E \{w(t + \Delta t) - w(t)\} = 0$$

since  $\{w(t + \Delta t) - w(t)\}$  is independent of  $w(s)$  for  $s \leq t$ . Thus by the law of conditional probabilities which are independent, we note that the Wiener process is a martingale.

(3) *The process  $x(t) = w(t)^2 - t$  is a martingale.* This assertion can be proved by showing that  $E [x(t + s) | \mathbf{F}(t)] = x(t)$ . By definition we have:

$$\begin{aligned} E[w(t + s) | \mathbf{F}(t)] &= E[w(t + s)^2 | \mathbf{F}(t)] - (t + s) \\ &= E[\{w(t + s) - w(t)\}^2 + 2w(t + s)w(t) - w(t)^2 | \mathbf{F}(t)] - (t + s) \\ &= E[\{w(t + s) - w(t)\}^2 | \mathbf{F}(t)] + 2E[w(t + s) | \mathbf{F}(t)] - E[w(t)^2 | \mathbf{F}(t)] - (t + s) \end{aligned}$$

Independence of the non-overlapping increments implies that  $w(t + s) - w(t)$  is independent of  $\mathbf{F}(t)$  which makes it possible to write:

$$E[\{w(t + s) - w(t)\}^2 | \mathbf{F}(t)] = E[\{w(t + s) - w(t)\}^2] = s$$

Conditional expectation and using the fact that  $w(t)$  is a martingale with respect to its filtration imply:

$$E [\{w(t + s)w(t)\} | \mathbf{F}(t)] = w(t)^2$$

and since  $E[w(t)^2 | \mathbf{F}(t)] = w(t)^2$ . We note therefore that:

$$E [x(t + s) | \mathbf{F}(t)] = s + 2w(t)^2 - w(t)^2 - (t + s) = w(t)^2 - t = x(t)$$

which proves that the process is a martingale.

(4) *The process*

$$x(t) = \exp \left( \alpha w(t) - \frac{\alpha^2 t}{2} \right)$$

where  $\alpha$  is any real number is a martingale. The proof follows the procedure above. We have to show that  $E [x(t + s) | \mathbf{F}(t)] = x(t)$ , or

$$\begin{aligned} E [x(t + s) | \mathbf{F}(t)] &= E \left[ \exp \left( \alpha w(t + s) - \frac{\alpha^2(t + s)}{2} \right) | \mathbf{F}(t) \right] \\ &= E \left[ x(t) \exp \left( \alpha [w(t + s) - w(t)] - \frac{\alpha^2 s}{2} \right) | \mathbf{F}(t) \right] \end{aligned}$$

Independence and conditional expectations make it possible to write:

$$\begin{aligned} E[x(t+s) | \mathbf{F}(t)] &= x(t) E \left[ x(t) \exp \left( \alpha [w(t+s) - w(t)] - \frac{\alpha^2 s}{2} \right) | \mathbf{F}(t) \right] \\ &= x(t) \exp \left( -\frac{\alpha^2 s}{2} \right) E[\exp \{ \alpha [w(t+s) - w(t)] \}] \end{aligned}$$

The term  $\alpha [w(t+s) - w(t)]$  has a Normal probability distribution with zero mean and variance  $\alpha^2 s$ . Consequently, the term  $\exp \{ \alpha [w(t+s) - w(t)] \}$  has a lognormal probability distribution with expectation  $\exp(\alpha^2 s/2)$  which leads to  $E[x(t+s) | \mathbf{F}(t)] = x(t)$  and proves that the process is a martingale.

## REFERENCES AND FURTHER READING

- Arnold, L. (1974) *Stochastic Differential Equations*, John Wiley & Sons, Inc., New York.
- Bachelier, L. (1900) *Théorie de la spéculation, thèse de mathématique*, Paris.
- Barrois, T. (1834) Essai sur l'application du calcul des probabilités aux assurances contre l'incendie, *Mem. Soc. Sci. de Lille*, 85–282.
- Bensoussan, A. (1982) *Stochastic Control by Functional Analysis Method*, North Holland, Amsterdam.
- Bernstein, P.L. (1998) *Against the Gods*, John Wiley & Sons, Inc., New York.
- Bibby, Martin, and M. Sorenson (1997) A hyperbolic diffusion model for stock prices, *Finance and Stochastics*, **1**, 25–41.
- Bismut, J.M. (1976) Théorie Probabiliste du Contrôle des Diffusions, *Memoirs of the American Mathematical Society*, **4**, no. 167.
- Born, M. (1954) Nobel Lecture, published in *Les Prix Nobel*, Nobel Foundation, Stockholm.
- Brock, W.A., and P.J. de Lima (1996) Nonlinear time series, complexity theory and finance, in G. Maddala and C. Rao (Eds), *Handbook of Statistics*, Vol. 14, *Statistical Methods in Finance*, North Holland, Amsterdam.
- Brock, W.A., D.A. Hsieh and D. LeBaron (1991) *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, MIT Press, Boston, MA.
- Cinlar, E. (1975) *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, NJ.
- Cox, D.R., and H.D. Miller (1965) *The Theory of Stochastic Processes*, Chapman & Hall, London.
- Cramer, H. (1955) *Collective Risk Theory*, Jubilee Volume, Skandia Insurance Company.
- Doob, J.L. (1953) *Stochastic Processes*, John Wiley & Sons, Inc., New York.
- Elliot, R.J. (1982) *Stochastic Calculus and Applications*, Springer Verlag, Berlin.
- Feller, W. (1957) *An Introduction to Probability Theory and its Applications*, Vols. I and II, John Wiley & Sons, Inc., New York (second edition in 1966).
- Gardiner, C.W. (1990) *Handbook of Stochastic Methods*, (2nd edn), Springer Verlag, Berlin.
- Gerber, H.U. (1979) *An Introduction to Mathematical Risk Theory*, Monograph no. 8, Huebner Foundation, University of Pennsylvania, Philadelphia, PA.
- Geske, R. and K. Shastri (1985) Valuation by approximation: A comparison of alternative option valuation techniques, *Journal of Financial and Quantitative Analysis*, **20**, 45–71.
- Ghahgshaie S., W. Breyman, J. Peinke, P. Talkner and Y. Dodge (1996) Turbulent cascades in foreign exchange markets, *Nature*, **381**, 767.
- Gihman, I.I., and A.V. Skorohod (1970) *Stochastic Differential Equations*, Springer Verlag, New York.
- Harrison, J.M., and D.M. Kreps (1979) Martingales and arbitrage in multiperiod security markets, *Journal of Economic Theory*, **20**, 381–408.

- Harrison, J.M., and S.R. Pliska (1981) Martingales and stochastic integrals with theory of continuous trading, *Stochastic Processes and Applications*, **11**, 261–271.
- Iglehart, D.L. (1969) Diffusion approximations in collective risk theory, *Journal of Applied Probability*, **6**, 285–292.
- Ito, K. (1961) *Lectures on Stochastic Processes*, Lecture Notes, Tata Institute of Fundamental Research, Bombay, India.
- Ito, K., and H.P. McKean (1967) *Diffusion Processes and their Sample Paths*, Academic Press, New York.
- Judd, K. (1998) *Numerical Methods in Economics*, MIT Press, Cambridge, MA.
- Kalman, R.E. (1994) Randomness reexamined, *modeling, Identification and Control*, **15**(3), 141–151.
- Karatzas, I., J. Lehocsky, S. Shreve and G.L. Xu (1991) Martingale and duality methods for utility maximization in an incomplete market, *SIAM Journal on Control and Optimization*, **29**, 702–730.
- Karlin, S., and H.M. Taylor (1981) *A Second Course in Stochastic Processes*, Academic Press, San Diego, CA.
- Kushner, H.J. (1990) Numerical methods for stochastic control problems in continuous time, *SIAM Journal on Control and Optimization*, **28**, 999–1948.
- Levy, P. (1948) *Processus Stochastiques et Mouvement Brownien*, Paris.
- Lorenz, E. (1966) Large-scale motions of the atmosphere circulation, in P.M. Hurley (Ed.), *Advances in Earth Science*, MIT Press, Cambridge, MA.
- Lundberg, F. (1909) *Zur Theorie der Ruckversicherung* Verhandlungskongress fur Versicherungsmathematik, Vienna.
- Malliariis, A.G., and W.A. Brock (1982) *Stochastic Methods in Economics and Finance*, North Holland, Amsterdam.
- Mandelbrot, B. (1972) Statistical methodology for non-periodic cycles: From the covariance to R/S analysis, *Annals of Economic and Social Measurement*, **1**, 259–290.
- Mandelbrot, B., and M. Taqqu (1979) Robust R/S analysis of long run serial correlation, *Bulletin of the International Statistical Institute*, **48**, book 2, 59–104.
- Mandelbrot, B., and J. Van Ness (1968) Fractional Brownian motion, fractional noises and applications, *SIAM Review*, **10**, 422–437.
- McKean, H.P. (1969) *Stochastic Integrals*, Academic Press, New York.
- Milshtein, G.N. (1974) Approximate integration of stochastic differential equations, *Theory of Probability and Applications*, **19**, 557–562.
- Milshtein, G.N. (1985) Weak approximation of solutions of systems of stochastic differential equations, *Theory of Probability and Applications*, **30**, 750–206.
- Nelson, Daniel, and K. Ramaswamy (1990) Simple binomial processes as diffusion approximations in financial models, *The Review of Financial Studies*, **3**(3), 393–430.
- Peter, Edgar E. (1995) *Chaos and Order in Capital Markets*, John Wiley & Sons, Inc., New York.
- Pliska, S. (1986) A stochastic calculus model of continuous trading: Optimal portfolios, *Mathematics of Operations Research*, **11**, 371–382.
- Révész, Pal (1994) *Random Walk in Random and Non-Random Environments*, World Scientific, Singapore.
- Samuelson, P.A. (1965) Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review*, **6**, 41–49.
- Seal, H.L. (1969) *Stochastic Theory of a Risk Business*, John Wiley & Sons, Inc., New York.
- Snyder, D.L. (1975) *Random Point Processes*, John Wiley & Sons, Inc., New York.
- Spitzer, F. (1965) *Principles of Random Walk*, Van Nostrand, New York.
- Stratonovich, R.L. (1968) *Conditional Markov Processes and their Applications to the Theory of Optimal Control*, American Elsevier, New York.
- Sulem, A., and C.S. Tapiero (1994) Computational aspects in applied stochastic control, *Computational Economics*, **7**, 109–146.
- Taylor, S. (1986) *modeling Financial Time Series*, John Wiley & Sons, Inc., New York.

## CHAPTER 5

# Derivatives Finance

### 5.1 EQUILIBRIUM VALUATION AND RATIONAL EXPECTATIONS

Fundamental notions such as rational expectations, risk-neutral pricing, complete and incomplete markets, underlie the market's valuation of risk and its pricing of derivatives assets. Both economics and mathematical finance use these concepts for the valuation of options and other financial instruments. Rational expectations presume that current prices reflect future uncertainties and that decision makers are rational, preferring more to less. It also means that current prices are based on the unbiased, minimum variance mean estimate of future prices. This property seems at first to be simple, but it turns out to be of great importance. It provides the means to 'value assets and securities', although, in this approach, bubbles are not possible, as they seem to imply a persistent error or bias in forecasting. This property also will not allow investors to earn above-average returns without taking above-average risks, leading to market efficiency and no arbitrage. In such circumstances, arbitrageurs, those 'smart investors' who use financial theory to identify returns that have no risk and yet provide a return, will not be able to profit without assuming risks.

The concept of rational expectation is due to John Muth (1961) who formulated it as a decision-making hypothesis in which agents are informed, constructing a model of the economic environment and using all the relevant and appropriate information at the time the decision is made (see also Magill and Quinzii, 1996, p. 23):

I would like to suggest that expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory... We call such expectations 'rational'... The hypothesis can be rephrased a little more precisely as follows: that expectations... (or more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (the objective probability of outcomes).

In other words, if investors are 'smart' and base their decisions on informed and calculated predictions, then, prices equal their discounted expectations. This hypothesis is essentially an equilibrium concept for the valuation of asset prices stating that under the 'subjective probability distribution' the asset price equals the

expectation of the asset's future prices. In other words, it implies that investors' subjective beliefs are the same as those of the real world – they are neither pessimistic nor optimistic. When this is the case, and the 'rational expectations equilibrium' holds, we say that markets are complete or efficient. Samuelson had already pointed out this notion in 1965 as the martingale property, leading Fama (1970), Fama and Miller (1972) and Lucas (1972) to characterize markets with such properties as markets efficiency.

Lucas used a concept of rational expectations similar to Muth to confirm Milton Friedman's 1968 hypothesis of the long-run neutrality of the monetary policy. Specifically, Lucas (1972, 1978) and Sargent (1979) have shown that economic agents alter both their expectations and their decisions to neutralize the effects of monetary policy. From a practical point of view it means that an investor must take into account human reactions when making a decision since they will react in their best interest and not necessarily the investor's.

Martingales and the concept of market efficiency are intimately connected. If prices have the martingale property, then only the information available today is relevant to make a prediction on future prices. In other words, the present price has all relevant information embedding investors' expectations. This means that in practice (the weak form efficiency) past prices should be of no help in predicting future prices or, equivalently, prices have no memory. In this case, arbitrage is not possible and there is always a party to take on a risk, irrespective of how high it is. Hence, risk can be perfectly diversified away and made to disappear. In such a world without risk, all assets behave as if they are risk-free and therefore prices can be discounted at a risk-free rate. This is also what we have called risk-neutral pricing (RNP). It breaks down, however, if any of the previous hypotheses (martingale, rationality, no arbitrage, and absence of transaction costs) are invalid. In such cases, prices might not be valued uniquely, as we shall see subsequently.

There is a confrontation between economists, some of whom believe that markets are efficient and some of whom do not. Market efficiency is 'under siege' from both facts and new dogmas. Some of its critics claim that it fails to account for market anomalies such as bubbles and bursts, firms' performance and their relationship to size etc. As a result, an alternative 'behavioural finance' has sought to provide an alternative dogma (based on psychology) to explain the behaviour of financial markets and traders. Whether these dogmas will converge back together as classical and Keynesian economics have, remains yet to be seen. In summary, however, some believe that the current price imbeds all future information. And some presume that past prices and behaviour can be used (through technical analysis) to predict future prices. If the 'test is to make money', then the verdict is far from having been reached. Richard Roll, a financial economist and money manager argues:

'I have personally tried to invest money, my clients' and my own, in every single anomaly and predictive result that academics have dreamed up. And I have yet to make a nickel on these supposed market inefficiencies. An inefficiency ought to be an exploitable opportunity. If there is nothing that investors can exploit in a systematic way, time in and time out, then it's very hard to say that information is not being properly incorporated into stock prices. Real money investment strategies do not produce the results that academic papers say they

should' ... but there are some exceptions including long term performers that have over the years systematically beaten the market. (Burton Malkiel, *Wall Street Journal*, 28 December 2000)

Rational expectation models in finance may be applied wrongly. There are many situations where this is the case. Information asymmetries, insider trading and advantages of various sorts can provide an edge to individual investors, and thereby violate the basic tenets of market efficiency, and an opportunity for the lucky ones to make money. Further, the interaction of markets can lead to instabilities due to very rapid and positive feedback or to expectations that are becoming trader- and market-dependent. Such situations lead to a growth of volatility, instabilities and perhaps, in some special cases, to bubbles and chaos. Nonetheless, whether it is fully right or wrong, it seems to work sometimes. Thus, although rational expectations are an important hypothesis and an important equilibrium pillar of modern finance, they should be used carefully for making money. It is, however, undoubtedly in theoretical finance where it is used with simple models for the valuation of options and for valuing derivatives in general – albeit this valuation depends on a riskless interest rate, usually assumed known (i.e. mostly assumed exogenous). Thus, although the arbitrage-free hypothesis (or rational expectations) assumes that decision makers are acting intelligently and rationally, it still requires the risk-free rate to be supplied. In contrast, economic equilibrium theory, based on the clearing of markets by equating ‘supply’ to ‘demand’ for all financial assets, provides an equilibrium where interest rates are endogenous. It assumes, however, that beliefs are homogeneous, markets are frictionless (with no transaction costs, no taxes, no restriction on short sales and divisible assets) as well as competitive markets (in other words, investors are price takers) and finally it also assumes no arbitrage. Thus, general equilibrium is more elaborate than rational expectations (and arbitrage-free pricing) and provides more explicit results regarding market reactions and prices (Lucas, 1972). The problem is particularly acute when we turn to incomplete markets or markets where pricing cannot be uniquely defined under the rational expectations hypothesis. In this case, a decision makers’ rationality is needed to determine asset prices. This was done in Chapter 3 when we introduced the SDF (stochastic discount factor) approach used to complement the no arbitrage hypothesis by a rationality that is sensitive to decision makers’ utility of consumption. In Chapter 9, we shall return to this approach in an inter-temporal setting. For the present, we introduce the financial instruments that we will attempt to value in the next chapters.

## 5.2 FINANCIAL INSTRUMENTS<sup>1</sup>

There are a variety of financial instruments that may be used for multiple purposes, such as hedging, speculating, investing, and ‘money multiplying’ or leveraging. Their development and use require the ingenuity of financial engineers and the

<sup>1</sup> This section is partly based on a paper written by students at ESSEC, Bernardo Dominguez, Cédric Lespiau and Philippe Pages in the Master of Finance programme. Their help is gratefully acknowledged.

care of practising investors. Financial instruments are essentially contracts of various denominations and conditions on financial assets. Contracts by definition, however, are an agreement between two or more parties that involves an exchange. The terms of the contract depend on the purpose of contracting, the contractees, the environment and the information available to each of the parties. Examples of contracts abound in business, and more generally in society. For example, one theory holds that a firm is nothing more than a nexus of contracts both internal and external in nature.

Financial contracts establish the terms of exchange between parties mostly for the purpose of managing contractors' and contract holders' risks. Derivative assets or derivative contracts are special forms of contract that derive their value from an underlying asset. Such assets are also called contingent claim assets, as their price is dependent upon the state of the underlying asset. For example, warrants, convertible bonds, convertible preferred stocks, options and forward contracts, etc. are some well-known derivatives. They are not the only ones, however. The intrinsic value of these assets depends on the objectives and the needs of the buyer and the seller as well as the right and obligations these assets confer on each of the parties. When the number of buyers and (or) sellers is very large, these contingent assets are often standardized to allow their free trading on an open market. Many derivatives remain over-the-counter (OTC) and are either not traded on a secondary market or are in general less traded and hence less liquid than their market counterparts. The demand for such trades has led to the creation of special stock exchanges (such as the Chicago, London, and Philadelphia commodities and currency exchanges) that manage the transactions of such assets. A number of such contingent assets and financial instruments are defined next.

### 5.2.1 Forward and futures contracts

A futures contract gives one side, the holder of the contract, the obligation to buy or sell a commodity, a foreign currency etc. at some specified future time at a specified price, place, quantity, location and quality, according to the contract specification. The buyer or long side has at the end of the contract, called the maturity, the option to buy the underlying asset at a predetermined price and sell it back at the market price if he wishes to do so. The seller or short side (provider), however, has the obligation to sell the underlying asset at the predetermined price. In futures contracts, the exchange of the underlying asset at a predetermined price is between anonymous parties which is not the case in OTC forward contracts. Financial futures are used essentially for trading, hedging and arbitrage.

Futures contracts can be traded on the CBOT (the Chicago Board of Trade) and the CME (the Chicago Mercantile Exchange), as well as on many trading floors in the world. Further, many commodities, currencies, stocks etc. are traded daily in staggering amounts (hundreds of billions of dollars). A futures price at time  $t$  with delivery at time  $T$  can be written by  $F(t, T)$ . If  $S(t)$  is the spot price, then clearly if  $t = T$ , we have by definition  $F(t, t) = S(t)$  and  $S(t) \geq F(t, T)$ ,  $T \geq t$ .

The difference between the spot assets to be pledged in a future contract and its futures price is often called the ‘basis risk’ and is given by  $b(t, T) = S(t) - F(t, T)$ . It is the risk one suffers when reversing a futures contracts position. Imagine we need to buy in 3-month pork for a food chain. We may buy futures contracts today that deliver the asset at a predetermined price in 6 months. After 3 months, we reverse or sell our futures position. The payoff is thus the change in the futures price less the price paid for the underlying asset, or  $F(3, 6) - F(0, 6) - S(3)$ . If we were at maturity, only  $-F(0, 6)$  would remain. That is, the price of the underlying asset is set by a delayed physical transaction using futures contracts. However, if there remains a basis risk in the payoff, then  $-F(0, 6) + b(3, 6)$  would remain. If the futures contract does not closely match the price of the underlying asset then the effectiveness of our hedging strategy will be reduced.

Futures contracts like forwards can be highly speculative instruments because they require no down payment since no financial exchange occurs before either maturity or the reversal of the position. Traders in the underlying assets can therefore use these markets to enhance their positions in the underlying asset either short or long. Unsurprisingly, a position in these contracts is considered levered or a borrowed position in the underlying asset, as the price of a forward and futures contract is nothing more than an arbitrage with the asset bought today using borrowed funds and delivered at maturity. There are differences between futures and forwards involving liquidity, marking to market, collaterals and delivery options, but these differences are generally glossed over.

The leverage implied in a futures contract explains why collaterals are required for forwards and marking to market for futures. In their absence, defaults would be much more likely to happen. For example, for a short futures contract, when prices fall, the investor is making a virtual loss since he would have to sell at a higher price than he started with (should he terminate his contract) and take an offsetting position by buying a futures contract. This is reflected in a ‘futures market’ when the bank adjusts the collateral account of the trader, called the margin. The margin starts at an initial level in, generally, the form of Treasury bills. It is adjusted every day to reflect the day’s gains or losses. Should the margin fall below a maintenance level, the trader will ask the investor to add funds to meet margin requirements. If the investor fails to meet such requirements, the trader cuts his losses by reversing the position.

A *forward rate agreement (FRA)* is an agreement made between two parties seeking generally to protect or hedge themselves against a future interest-rate or price movement, for a specific hedging period, by fixing the future interest rate or price at which they will buy or sell for a specific principal sum in a specified currency. It requires that settlement be effected between the parties in accordance with an established formula. Typically, forward contracts, unlike futures contracts are not traded and can therefore be tailored to specific needs. This means that contracts tend to be much higher in size, far less liquid and less competitively priced, but suffer from no basis risk. The price at time  $t$  of a forward contract at time  $T$  in the future can be written by  $p(t, T)$  or by  $p(t, t + x)$ ,  $x = T - t$  and is defined by the (delivery) price for which the contract value is null at delivery

under risk neutral pricing.

$$E(\text{Future spot rate} - \text{Forward rate}) = 0$$

Of course,  $p(T, T) = 1$  and therefore the derivative of the price with respect to  $T$  (or  $x$ ) is necessarily negative, reflecting the lower value of the asset in the future compared to the same asset in the present.

The relationship between forward rates and spot prices is a matter of intensive research and theories. For example, the theory of rational expectations suggests that we equate the expected future spot rate to the current forward rate, that is (see also the next chapter):

$$\text{Forward rate} = E(\text{Future spot rate})$$

For example, if  $s_t$  is the logarithm of the spot price of a currency at time  $t$  and  $f_t$  is the logarithm of the 1 month forward price, the expectation hypothesis means that:

$$f_t = E(s_{t+1})$$

Note that if  $S_t$  is the spot price at time  $t$ ,  $\Delta S_t = S_{t+1} - S_t$ , then the rate of change, expressing the rates of return  $\Delta S_t/S_t$  is given by  $\Delta(\log S_t)$  with  $s_t = \log S_t$ . Empirical research has shown, at least for currencies forward, that it is misleading and therefore additional and alternative theories are often devised which introduce concepts of risk premium as well as the expected rate of depreciation to explain the incoherence between spot and forward market values and risk-neutral pricing.

Forward and futures contracts are not only used in financial and commodities markets. For example, a transport futures exchange has been set up on the Internet to help solve forward-planning problems faced by truckers and companies shipping around the world. The futures exchange enables companies to purchase transport futures, helping them to plan their freight requirements and shipments by road, rail and, possibly, barge. The exchange allows truckers and manufacturers to match transport capacity to their shipments and to match their spot requirements, buy and sell forward, and speculate on future movements of the market. This market completes other markets where one can buy and sell space on ocean-going ships. For example, London's Baltic Exchange handles spot trades in dry cargo carriers and tankers.

## 5.2.2 Options

Options are instruments that let the buyer of the option (the long side) the *right to exercise*, for a price, called the *premium*, the delivery of a commodity, a stock, a foreign currency etc. at a given price, called the *strike price*, at (*within*) a given time period, also called the *exercise date*. Such an option is called a *European (American) CALL* for the buyer. The seller of such an option (the short side), has by contrast the *obligation* to sell the option at the stated strike and exercise date. A *PUT* option (the long side) provides an option to sell while for the short seller

this is an obligation to buy. There are many types of options, however. Below are a selected few (in the next two chapters we shall consider a far larger number of option contracts):

- *Call option* (long) (on foreign exchange (FX), deposit or futures etc.): an option contract that gives the holder the right to buy a specified amount of the commodity, stock or foreign currency for a premium on or before an expiration date as stated above. A call option (short), however, is an obligation to maintain the terms of this contract.
- *Put option* (long) (on FX, commodity etc.): gives the right to sell a specified amount at the strike price on (or before, for an American option) a specific expiration date. The short side of such a contract is an obligation, however, to meet the terms of this contract.
- *Swaps* (for interest rates, currency and cross-currency swaps, for example): transactions between two unrelated and independent borrowers, borrowing identical principal amounts for the same period from different lenders and with an interest rate calculated on a different basis. The borrowers agree to make payments to each other based on the interest cost of the other's borrowing. It is used both for arbitrage and to manage firm's liabilities. It can facilitate access of funding in a particular currency, provide export credits or other credits in a particular currency, provide access to various capital markets etc. These contracts are used intensively by banks and traders and will be discussed at length in the next chapter.
- *Caps*: a contract in which a seller pays a buyer predetermined payments at prespecified dates, with an interest (cap) rate calculated at later dates. If the rate of reference (the variable rate) is superior to a guaranteed rate, then the cap rate becomes effective, meaning that the largest interest rate is applied.
- *Floors*: products consisting in buying a cap and at the same time selling another product at a price compensating exactly the buying price of the cap. In this case, the floor is a contract in which the seller pays to the buyer for a predetermined period with a rate calculated at the fictive date. If the reference rate (the variable rate) is inferior to the guaranteed rate by the floor (rate), then the lower rate is applied.

### **Options again**

Trading in options and other derivatives is not new. Derivative products were used by Japanese farmers and traders in the Middle Ages, who effectively bought and sold rice contracts. European financial markets have traded equity options since the seventeenth century. In the USA, derivative contracts initially started to trade in the CBOT (Chicago Board of Trade) in 1973. Derivatives were thus used for a long time without stirring up much controversy. It is not the idea that is new, it is the volume of trade, the large variety of instruments and the significant and growing number of users trading in financial markets that has made derivatives a topic that attracts permanent attention.

Today, the most active derivative market is the CBOT, while the CME (Mercantile Stock Exchange) ranks second. Other active exchanges are the CBOE,

PHLX, AMEX, NYSE and TSE (Toronto Stock Exchange). In Montreal a stock exchange devoted to derivatives was also started in 2001. In Europe the most active markets are LIFFE (London International Financial Futures Exchange), MATIF (Marché à Terme International de France), DTB (Deutsche Terminbörse), and the EOE (Amsterdam's European Options Exchange). The most voluminous markets in East Asia include TIFFE (Tokyo International Financial Futures Exchange), the Hong Kong Futures Exchange and SIMEX (Singapore International).

Options contracts in particular are traded on many trading floors and, mostly, they are defined in a standard manner. Nevertheless, there are also 'over-the-counter options' which are not traded in specific markets but are used in some contracts to fit specific needs. For example, there are 'Bermudan and Asian options'. The former option provides the right to exercise the option at several specific dates during the option lifetime while the latter defines the exercise price of the option as an average of the value attained over a certain time interval. Of course, each option, defined in a different way, will lead to alternative valuation formulas. More generally, there can be options on real assets, which are not traded but used to define a contract between two parties. For example, an airline company contracts the acquisition (or the option to acquire) a new (technology) plane at some future time. The contract may involve a stream or a lump sum payment to the contractor (Boeing or Airbus) in exchange for the delivery of the plane at a specified time. Since payments are often made prior to the delivery of the plane, a number of clauses are added in the contract to manage the risks sustained by each of the parties if any of the parties were to deviate from the contract stated terms (for example, late deliveries, technological obsolescence etc.). Similarly, a manufacturer can enter into binding bilateral agreements with a supplier by which agreed (contracted) exchange terms are used as a substitute for the free market mechanism. This can involve future contractual prices, delivery rates at specific times (to reduce inventory holding costs) and, of course, a set of clauses intended to protect each party against possible failures by the other in fulfilling the terms of the contract. Throughout the above cases the advantage resulting from negotiating a contract is to reduce, for one or both parties, the uncertainty concerning future exchange operating and financial conditions. In this manner, the manufacturer will be eager to secure long-term sources of supplies, and their timely availability, while the investor, buyer of options, would seek to avoid too large a loss implied by the acquisition of a risky asset, currency or commodity, etc. Since for each contract there, necessarily, needs to be one (or many) buyer and one (or many) seller, the price of the contract can be interpreted as the outcome of a negotiation process where both parties have an inducement to enter into a contractual agreement. For example, the buyer and the seller of an option can be conceived of as being involved in a game, the benefits of which for each of the players are deduced from premium and risk transfer. Note that the utility of entering into a contractual agreement is always positive *ex-ante* for all parties; otherwise there would not be any contractual agreement (unless such a contract were to be imposed on one of the parties!). When the number of buyers and sellers of such contracts becomes extremely large, transactions become 'impersonal' and it is the 'market price' that defines the value of the contract. Strategic

behaviours tend to break down the larger the group and prices tend to become more efficient.

### **Making decisions with options**

We shall see in Chapter 7, ‘Options and Practice’, some approaches using options in hedging and in speculating. Decisions involving options are numerous, e.g.:

- Buy and sell; on the basis of the stock price and the remaining time to its exercise.
- Buy and sell; on the basis of estimated volatility of the underlying or related statistics.
- Use options to hedge downside risk.
- Use stock options to motivate management and employees.
- Use options and stock options for tax purposes.
- Use options to raise money for investments.

These problems clearly require a competent understanding of options theory and financial markets and generally the ability to construct and compound assets, options and other contracts into a portfolio of desirable characteristics. This is also called financial engineering and is also presented in the next chapter.

We shall use a theoretical valuation of options based on ‘risk-neutral probabilities’. ‘Uncertainty’, defined by ‘risk-neutral probabilities’, unlike traditional (historical) probabilities, determined by interacting market forces, reflects the market resolution of demand and supply (equilibrium) for assets of various risks. This difference contrasts two cultures. It is due to economic and financial assumptions that current market prices ‘endogenize’ future prices (states and their best forecast based on available information). If this is the case, and it is so in markets we call *complete markets*, the current price ought to be determined by an appropriate discounting of expected future values. In other words, *it is the market that determines prices and not uncertainty*. We shall calculate explicitly these ‘probabilities’ in the next chapter when we turn to the technical valuation of options.

## **5.3 HEDGING AND INSTITUTIONS**

Financial instruments are used in many ways to reduce risk (*hedging*), make money through *speculation* (which means that the trader takes a position, short or long, in the market) or through *arbitrage*. Arbitrage consists of taking positions in two or more markets so that a riskless profit is made (i.e. providing an infinite rate of return since money can be made without committing any investment). The number of ways to hedge is practically limitless. There are therefore many trading strategies financial managers and insurers can adopt to protect their wealth or to make money. Firms can use options on currency, commodities and other assets to protect their assets from unexpected variations. Financial institutions (such as banks and lending institutions) by contrast, use options to cover their risk

exposure and immunize their investment portfolios. Insurance firms, however, use options to seek protection against excessive uncontrollable events, to diversify risks and to spread out risks with insured clients as we have outlined earlier. Generally, hedging strategies can be ‘specialized services’, tailored to individual and collective needs.

### 5.3.1 Hedging and hedge funds

Hedging is big business, with many financial firms providing a broad range of services for protecting investments and whatnot. The traditional approach, based on portfolio theory, optimizes a portfolio holding on the basis of risk return substitution (measured by the mean and variance of returns as we saw in Chapter 3). Hedging, however, proceeds to eliminate a particular risk in a portfolio through a trade or a series of trades (called hedges). While in portfolio management the investor seeks the largest returns given a risk level, in hedging – also used in the valuation of derivatives – a portfolio is constructed to eliminate completely the risk associated with the derivative (the option for example). In other words, a hedging portfolio is constructed, replicating the derivative security. If this can be done, then the derivative security and the replicating portfolio should have the same value (since they have exactly the same return properties). Otherwise, there would be a potential for arbitrage.

Hedge funds, however, may be a misnomer. They attract much attention because of the medias’ fascination with their extremes – huge gains and losses. They were implicated in the 1992 crisis that led to major exchange rate realignments in the European Monetary System, and again in 1994 after a period of turbulence in international bond markets. Concerns mounted in 1997 in the wake of the financial upheavals in Asia. And they were amplified in 1998, with allegations of large hedge fund transactions in various Asian currency markets and with the near-collapse of Long-Term Capital Management (LTCM). Government officials, fearing this new threat to world financial markets, stepped in to coordinate a successful but controversial private–public sector rescue of LTCM. Yet, for all this attention, little concrete information is available about the extent of hedge funds’ activities and how they operate. Despite a plethora of suggestions for reforms, no consensus exists on the implications of hedge fund activity for financial stability, or on how policy should be adapted.

The financial community defines a hedge fund as any limited partnership, exempted from certain laws (due to its legal location, shareholders features, etc.), whose main objective is to manage funds and profits. The term ‘hedge fund’ was coined in the 1960s when it was used to refer to investment partnerships that used sophisticated arbitrage techniques to invest in equity markets. Federal regulation of financial instruments and market participants in the USA is based on Acts of Congress seeking to protect individual market investors. However, by accepting investments only from institutional investors, companies, or high-net-worth individuals, hedge funds are exempt from most of investor protection and regulations. Consequently, hedge funds and their operators are generally not registered and are not required to publicly disclose data regarding their financial performance

or transactions. Hence they have been accused of being speculative vehicles for financial institutions that are constrained by costly prudential regulations.

Hedge funds can also be eclectic investment pools, typically organized as private partnerships and often located offshore for tax and regulatory reasons. Their managers, who are paid on a fee-for-performance basis, may be free to use a variety of investment techniques, including short positions and leverage, to raise returns and cushion risk. While hedge funds are a rapidly growing part of the financial industry, the fact that they operate through private placements and restrict share ownership to rich individuals and institutions frees them from most disclosure and regulation requirements applied to mutual funds and banks. Further, funds legally domiciled outside the main financial markets and main trading countries are generally subject to even less regulation.

Hedge funds operate today as both speculators and hedgers, using a broad spectrum of risk-management tools. *Macro funds*, for example, base their investment strategies on the use of perceived discrepancies in the economic fundamentals of macroeconomic policies. Macro funds, may take large directional (unhedged) positions in national markets based on top-down analysis of macroeconomic and financial conditions, including current accounts, the inflation rate, the real exchange rate, etc. As a result, they necessarily are very sensitive to countries risk, global and national politics, economics and finance. Macro hedge funds may be classified into two essential categories:

- Arbitrage-based investment strategies
- Macro specific funds strategies

*Arbitrage-based strategies* seek to profit from current price discrepancies in two instruments (or portfolios) that will, at their maturity, have the same value. However, hedge funds that call themselves arbitrage-type use analytical models to profit from the discrepancy between their valuation ‘model’ and the actual market price. The arbitrage always involves two transactions: the purchase of an undervalued asset and the sale of an overvalued asset. Some commonly utilized arbitrage strategies include:

- Trade an instrument (cash instruments, index of equity securities, currency spot price, etc.) against its futures counterpart.
- Misalignments in prices of cash market fixed-income securities. A hedge fund might have a model for the levels of yield representing a number of bonds with various maturities. If the yield curve models differ from the yields of some bonds there is an opportunity for arbitrage.
- Misalignments because of the credit quality of two instruments. These sorts of trades are routinely executed by hedge funds examining the differences in the creditworthiness of various US corporate securities relative to the US Treasury bond yield spread.
- Convertible arbitrage involves purchasing convertible securities, mostly fixed-income bonds that can be converted into equity under certain circumstances. A portion of the equity risk embedded in the bond is hedged by selling short

the underlying equity. Sometimes the strategy will also involve an interest rate hedge to protect against general fluctuations in the yield curve. Thus, this trade would be designed to profit from mis-pricing of the equity associated with the convertible bond.

- Misalignments between options or other features imbedded in mortgage-backed securities. Often, complicated structures can be decomposed into various components that have market counterparts, permitting hedge funds to profit from deviations in prices of the underlying components and the structured product. The prepayment risk – the risk that the mortgage holder will prepay the mortgage prior to its maturity – is such an example.

Many of the determinants of a viable strategy are not specific to hedge funds, but are common to many types of investors. Virtually all hedge funds calculate whether the all-in return more than compensates for the risk undertaken. Three elements are taken into account in these calculations: (1) examining the market risk, which usually includes some type of ‘stress test’ to assess the downside risks of the proposed strategy; (2) examining the liquidity risk, that is, to see whether the hedge fund can enter and exit markets without extra costs in both normal times and in periods of market distress; (3) examining the timing and the cost of financing the position. If the expected duration of the trade is too long, with a prohibitive financing cost, the position will not be assumed.

*Macro specific funds strategies* are based on information regarding economic fundamentals. They seek incongruent relationships between the level of prices and the country’s fundamentals – both economically and psychologically. Macro hedge funds are universally known for their ‘top-down’ global approach to investments, combining knowledge of economics, politics and history into a coherent view of things to come. In currency markets, a macro fund strategy might examine countries maintaining a pegged exchange rate to the dollar but having little economic reason for using the dollar for the peg. Some funds use rather detailed macroeconomic modelling techniques; others use less quantitative techniques, examining historical relationships among the various variables of interest. They may examine the safety and soundness of the banking sector and its connections to other parts of the financial sector. Excess liquidity and credit growth within the banking sector are often cited by funds as leading indicators of subsequent banking problems. Extensive use of unhedged foreign-currency-denominated debt of banks is also a tip-off for hedge funds. A pattern of high and fast appreciation of various assets is also used as a signal for a financial sector awaiting a downturn. Political risk and the probability that government’s strategy may, or may not, be implemented are also used as signals on the basis of which positions may be taken. However, market funds are very sensitive to the potential for market exit (and thus liquidity) in the case where events are delayed or do not confirm expectations.

Risk management in a hedge fund is often planned and integrated across products and markets (related through correlation analysis). Scenario analysis and stress tests are common diagnostic techniques. Further, some trading risks are managed by limiting the types, the number and the market exposure of trades.

The criteria used are varied, such as the recent track record of the trader, the relative portfolio risk of the trade, and market liquidity.

### 5.3.2 Other hedge funds and investment strategies

There are of course, many strategies for hedge fund management and trading. We can only refer to a few.

*Market hedge funds* focus on either equity or debt markets of developing or emerging countries. In general, they are classified by geographical areas and combine arbitrage and macro hedge fund strategies. Since many emerging markets are underdeveloped and illiquid, we note three points. (1) The size of transactions is relatively small. (2) Pricing of various securities abounds. These are inefficient markets for a number of reasons such as a basic misunderstanding of their operation, due to selling agents behaviour governed by liquidity needs rather than by ‘market rationality’. (3) Bets on political events may cause important differences in valuations. Political risk receives special attention for emerging market hedge funds compared to the Group of Ten leading countries where economic considerations are prominent.

*Event-related funds* focus on securities of firms undergoing a structural change (mergers, acquisitions, or reorganizations), seeking to profit from increases or decreases in both stock prices – before a merger or when valuation of the merged firm is altered appreciably. These funds may estimate the time to complete the merger and the annualized return on the investment if undertaken. Annualized returns includes the purchase and sale of the equity of the two merging companies and the cost of executing the short position, any dividends gained or lost and commissions. Using such returns calculations, the fund can assess the probability that a deal will be consummated. If annualized returns, including the probability that the merger will come through is greater than a ‘baseline’, the fund may execute the deal.

*Value investing funds* have a strategy close to mutual funds (portfolio) strategies seeking to profit from undervalued companies. Hedge funds are probably more likely to use hedging methodologies designed to offset industry risk and reduce market volatility, however.

*Short-selling funds* use short-selling strategies. They involve limited partnerships and offshore funds sponsored by wealthy individuals. In short sales, the investor sells short a stock at the current market price while the capital is invested in US Treasury securities with the same holding period. The amount of capital is then adjusted daily to reflect the change in the stock price. If the stock price decreases, free cash is released; when the stock price increases, the capital must be increased. Losses on a short position are unlimited since they must be paid in real time. As a result, the short seller may run out of capital, making the depth of the short sellers’ pocket and the timing of trades important determinants of success of the fund.

*Sector funds* combine strategies described above but applied to the ‘sector’ the fund focuses on and in which it trades. A sector may have specific characteristics, recognized and capitalized on for making greater profits.

Hedge funds have raised concerns due to their often speculative and destabilizing character. For this reason, financial regulation agencies have devoted special attention to regulating funds. Further, hedge funds often use stabilizing strategies. Two such strategies are employed: ‘counter’ strategies and arbitrage strategies. Counter strategies involve buying when prices are thought to be too low and selling when they are thought to be high, countering current market movements. It is an obvious strategy when prices are naturally pushed back to their perceived fair value, thereby stabilizing prices. Arbitrage strategies are neither stabilizing nor destabilizing since the arbitrageur’s action simply links one market to another. However, studies have shown that arbitrage activity on stock indices are in fact stabilizing, in the sense of reducing volatility of the underlying stocks.

In contrast, *destabilizing strategies* can be divided into two essential groups: (1) strategies that use existing prices and (2) strategies that use positions of other market participants for trading decisions. The first group is often called ‘positive feedback trading’; if there are no offsetting forces, these participants can cause prices to ‘overshoot’ their equilibrium value, adding volatility relative to that determined by fundamental information. It can arise under a variety of circumstances, some of which are related to institutional features of markets. These include dynamic hedging, stop loss orders, and collateral or margin calls. On a simpler level, positive feedback strategies also incorporate general trend-following behaviour where investors use various technical rules to determine trends, reinforced by buying and selling on the trend. Among strategies inducing a positive feedback type behaviour, the most complex is dynamic hedging. Options sellers, for example (using a put protective strategy, see Chapter 7), sell the underlying asset as its price decreases in order to dynamically hedge to replicate put options. Thus, to hedge, they would be required to sell the underlying asset in a falling market to maintain a hedged position, potentially exacerbating the original movement. In general, hedge funds are typically buyers of options (not sellers) and do not need to hedge themselves; but dealers that sell those options to hedge funds do need to hedge. Other institutional features like collateral calls or margin calls can also lead to a positive feedback response. Collateral holders may require additional collateral from their customers when prices fall and losses are incurred. Often, the collateral is obtained by selling any number of instruments, causing further price declines and losses. Some intermediaries, providing margins to hedge funds can keep these funds on a very tight leash, requiring margin calls more than once a day if necessary.

A second group of trading behaviours that destabilize hedge funds results from herding – taking similar positions to other market participants, rather than basing decisions explicitly on prices. Positions can be mimicked directly by observing what other participants do or indirectly by using the same information, analysis and tools as other participants. Often, fund managers have an incentive to mimic other participants’ behaviour to hide their own incompetence. There may be then a temptation to ignore private information and realign their performance on others. Since hedge fund managers have most of their wealth invested in the fund and are compensated on total absolute returns rather than on relative benchmarks, they are less inclined than other fund managers to ‘herd’, directly mimicking others.

However, many hedge funds probably hold the same analytical tools and have access to the same information, arriving necessarily, at similar assessments and at approximately the same time, creating an appearance of collusion. Further, even if hedge funds do not herd (directly or indirectly), other investors may herd with them or follow their lead into various markets.

Hedge funds, like other institutional investors, are potentially subject to three general types of prudential regulations: (1) those intended to protect investors, (2) those designed to ensure the integrity of markets, and (3) those meant to contain systemic risk. Investor protection regulation is employed when authorities perceive a lack of sophistication on the part of investors, for example, lacking the information needed to properly evaluate their investments. Then, regulations can either ensure that sufficient information is properly disclosed or exclude certain types of investors from participating in certain investments.

Regulation to protect market integrity seeks to ensure that markets are designed so that price discovery is reasonably efficient, that market power is not easily concentrated in ways that allow manipulation, and that pertinent information is available to potential investors.

Systemic risk is often the most visible element in the regulation of financial markets because it often requires coordination across markets and across regulatory and geographical boundaries. Regulation to protect market integrity and/or limit systemic risk, which includes capital requirements, exposure limits, and margin requirements, seeks to ensure that financial markets are sufficiently robust to withstand the failure of even the largest participants.

### **5.3.3 Investor protection rules**

Shares in hedge funds are securities but, since they are issued through private placements,<sup>2</sup> they are exempt from making extensive disclosure and commitments in the detailed prospectuses required of registered investment funds. They must still provide investors with all material information about their securities and will generally do so in an offering memorandum. Non-accredited investors are generally not accepted by hedge funds, because they would have to be given essentially the same information that would have been provided as a registered offer. However, most hedge fund operators are likely to be subject to regulation under the Commodity Exchange Act, because of their activity as commodity pool operators and/or as large traders in the exchange-traded futures markets. Requirements for commodity pools and commodity pool operators (CPOs) are mainly relative to (1) personal records and exams to get registered, (2) disclosure and reporting on issues as risks relevant to the pool, historical performance, fees incurred by participants, business backgrounds of CPOs, any possible conflict of interest on the part of the CPOs, and (3) maintenance of detailed records at the head office.

<sup>2</sup> A private placement consists of an offering of securities made to investors on an individual (bilateral) basis rather than through broader advertising. It is not allowed to offer for sale the securities by any form of general solicitation or advertising.

*Market integrity protection rules:* Although hedge funds can opt out of many of the registration and disclosure requirements of the securities laws, they are subject to all the laws enacted to protect market integrity. The essential purpose of such laws is to minimize the potential of market manipulation by increasing transparency and limiting the size of positions that a single participant may establish in a particular market. Many of these regulations also help in containing the spillovers across markets and hence in mitigating systemic risks.

The Treasury monitors all 'large' participants in the derivatives markets. Weekly and monthly reports are required of large participants, defined as players with more than US\$50 billion equivalent in contracts at the end of any quarter during the previous year. The Treasury puts out the aggregate data in its monthly bulletin but the desegregated data by participant are not published or revealed to the public. For government securities, the US Treasury is allowed to impose reporting requirements on entities having large positions in to-be-issued or recently issued Treasury securities. Such information is deemed necessary for monitoring large positions in Treasury securities and making sure that players are not squeezing other participants. The Security Exchange Act (SEA) also requires the reporting of sizeable investments in registered securities. It obliges any person who, directly or indirectly, acquires more than 5 % of the shares of a registered security to notify the SEC within 10 days of such acquisition. In overseeing the futures markets, the CFTC attempts to identify large traders in each market, their positions, interaction of related accounts, and, sometimes, even their trading intentions. Also, to reinforce the surveillance, each exchange is required to have its own system for identifying large traders. For example, the Chicago Mercantile Exchange requires position reports for all traders with more than 100 S&P 500 contracts. The regulators have the authority to take emergency action if they suspect manipulation, cornering of a market, or any hindrance to the operation of supply and demand forces.

*Systemic risk reduction rules:* The key systemic question is to what extent are large, and possibly leveraged, investors, including hedge funds, a source of risk to the financial institutions that provide them with credit and to the intermediaries, such as broker-dealers, who help them implement their investment strategies. Banks provide many services to hedge funds and accept hedge funds as profitable customers with associated risks controllable. They examine the structure of the collective investment vehicle, the disclosure documents submitted to regulators and those offered to clients, the financial statements, and the fund's performance history. Further, generally, a large proportion of the credit extended by banks to hedge funds is collateralized.

The SEC also monitors brokers' and dealers' credit risk exposure. The net capital rule fortifies a broker-dealer against defaults by setting minimum net capital standards and requiring it to deduct from its net worth the value of loans that have not been fully collateralized by liquid assets. Further, reporting rules enable a periodic assessment and, at times, continuous monitoring of the risks posed to broker-dealers by their material affiliates, including those involved in over-the-counter. Along with the bank and broker-dealer credit structures that protect

against excessively large uncollateralized positions, the Treasury and CFTC large position and/or large trader reporting requirements, by automatically soliciting information, provide continuous monitoring of large players in key markets and hence allow early detection of stresses in the system.

*Mutual funds regulation*, however, is strict, protecting shareholders, by:

- *Regulatory requirements* to ensure that investors are provided with timely and accurate information about management, holdings, fees, and expenses and to protect the integrity of the fund's assets. Therefore, mutual fund holdings and strategies are also regulated. In contrast, hedge funds are free to choose the composition of their portfolios and the nature of their investment strategies.
- *Fees*. Federal law requires a detailed disclosure, a standardized reporting and imposes limits to mutual fund fees and expenses. Hedge fund fees need not be disclosed and there are no imposed limits, which generally are between 15 and 20 % of returns and between about 1 and 2 % of net assets.
- *Leverage* practices and derivative products are used to enhance returns or reduce risks and have a restricted usage in mutual funds, while hedge funds have no restrictions other than their own internal strategies or partnership agreements.
- *Pricing and liquidity*. Mutual funds are required to price their shares daily and to allow shareholders to redeem shares also on a daily basis. Hedge funds, however, have no rules about pricing their own shares and redemption of shares may be restricted by the partnership agreement if wanted.
- *Investors*. The minimum initial investment to enter a mutual fund is about US\$1000–2500. To own shares in a hedge fund, it is commonly required to make a commitment of US\$1 million. Such measures are designed to restrict share ownership and, in consequence, to fall in a much weaker investor protection rules environment.

## REFERENCES AND ADDITIONAL READING

- Asness, C., R. Kraib and J. Liew (2001) Do hedge funds hedge?, *Journal of Portfolio Management*, Fall, 6–19.
- Fama, E.F. (1970) Efficient capital markets: A review of theory and empirical work, *The Journal of Finance*, **25**, 383–417.
- Fama, E.F., and M.H. Miller (1972) *The Theory of Finance*, Holt, Rinehart & Winston, New York.
- Fothergill, M. and C. Coke (2001) Funds of hedge funds: An introduction to multi-manager funds, *Journal of Alternative Investments*, Fall, 7–16.
- Henker, T. (1998) Naïve diversification for hedge funds, *Journal of Alternative Investments*, Winter, 32–42.
- Liang, B. (2001) Hedge funds performance: 1990–1999, *Financial Analysts Journal*, Jan/Feb., **57**, 11–18.
- Lucas, R.E. (1972), Expectations and the Neutrality of Money, *Journal of Economic Theory*, **4**(2), 103–124.
- Lucas, R.E. (1978) Asset prices in an exchange economy, *Econometrica*, **46**, 1429–1446.

- Magill, M., and M. Quinzii (1996) *Theory of Incomplete Markets*, Vol 1, MIT Press, Boston, MA.
- Muth, J. (1961) Rational expectations and the theory of price movements, *Econometrica*, **29**, 315–335.
- Sargent, T.J. (1979) *Macroeconomic Theory*, Academic Press, New York.

PART II

# Mathematical and Computational Finance

## CHAPTER 6

# Options and Derivatives Finance Mathematics

### 6.1 INTRODUCTION TO CALL OPTIONS VALUATION

Options are some of the building blocks of modern corporate finance and financial economics. Their mathematical study is in general difficult, however. In this chapter and in the following one, we consider the valuation of options and their use in practice. Terms such as a trading strategy, risk-neutral pricing, rational expectations, etc. will be elucidated in simple mathematical terms. To value an option it is important to define first, and clearly, a number of terms. This is what we do next.

We begin by defining *wealth* at a given time  $t$ ,  $W(t)$ . This is the amount of money an investor has either currently invested or available for investment. Investments can be made in a number of assets, some of which may be risky, providing uncertain returns, while others may provide a risk-free rate of return (as would be achieved by investing in a riskless bond) which we denote by  $R_f$ . A risky investment is assumed for simplicity to consist of an investment in securities. Let  $N_0$  be the number of bonds we invest in, say zero coupon of \$1 denomination, bearing a risk-free rate of return  $R_f$  one period hence. Thus, at a given time, our investment in bonds equals  $N_0B(t, t + 1)$  with  $B(t + 1, t + 1) = 1$ . This means that one period hence, this investment will be worth  $B(t, t + 1)N_0(1 + R_f) = N_0(1 + R_f)$  for sure. We can also invest in risky assets consisting of  $m$  securities each bearing a known price  $S_i(t)$ ,  $i = 1, \dots, m$  at time  $t$ . The investment in securities is defined by the number of shares  $N_1, N_2, \dots, N_m$  bought of each security at time  $t$ . Thus, a trading strategy at this time is given by the portfolio composition  $(N_0, N_1, N_2, \dots, N_m)$ . The total portfolio investment at time  $t$ , is thus given by:

$$W(t) = N_0B(t, t + 1) + N_1S_1(t) + N_2S_2(t) + \dots + N_mS_m(t)$$

For example, for a portfolio consisting of a bond and in a stock, we have:

$$W(t) = \begin{cases} \$N_0 & \text{invested in a riskless bond} \\ \$N_1 S_1(t) & \text{invested in a risky asset, a stock} \end{cases}$$

$$W(t) = N_0 + N_1 S_1(t)$$

A period later, the bond is cashed while security prices may change in an uncertain manner. That is to say, the price in the next period of a security  $i$  is a random variable that we specify by a 'tilde', or  $\tilde{S}_i(t+1)$ . The gain (loss) is thus the random variable:

$$\Delta S_i(t) = \tilde{S}_i(t+1) - S_i(t), i = 1$$

Usually, one attempts to predict the gain (loss) by constructing a stochastic process for  $\Delta S_i(t)$ . The wealth gain (loss) over one period is:

$$\Delta W(t) = \tilde{W}(t+1) - W(t)$$

where,

$$\tilde{W}(t+1) = N_0(1 + R_f) + N_1 \tilde{S}_1(t+1)$$

Thus, the net gain (loss) in the time interval  $(t, t+1)$ , is:

$$\Delta W(t) = N_0 R_f + N_1 \Delta S_1(t)$$

In general, a portfolio consists of multiple assets such as bonds of various denominations and maturities, stocks, options, contracts of various sorts and assets that may be more or less liquid (such as real estate or transaction-cost-prone assets). We restrict ourselves for the moment to an investment in a simple binomial stock and a bond.

Over two periods, future security prices assume two values only, one high  $S_H$  (the security price increases), the other low  $S_L$  (the security price decreases) with  $0 < S_L < S_H$  as well as  $S_L/S \leq 1 + R_f \leq S_H/S$ . These conditions will exclude arbitrage opportunities as we shall see later on. Thus stock prices at  $t$  and at  $t+1$  are (see Figure 6.1):

$$S(t) \quad \text{and} \quad \tilde{S}(t+1) = \begin{cases} S_H \\ S_L \end{cases}$$

This results in a portfolio that assumes two possible values at time  $t+1$ :

$$W(t) = \begin{cases} N_0 \\ N_1 S \end{cases} \quad \text{and} \quad \tilde{W}(t+1) = \begin{cases} N_0(1 + R_f) + N_1 S_H \\ N_0(1 + R_f) + N_1 S_L \end{cases}$$

In other words, at time  $t$  the current time, the price of a stock is known and given by  $S = S(t)$ . An instant of time later, at  $(t+1)$ , its price is uncertain and assumes the two values  $(S_H, S_L)$ , with  $S_H > S_L$ . As a result, if at  $t = 0$ , wealth is invested in a bond and in a security, we have the investment process given by (see Figure 6.2):

$$W(0) = N_0 + N_1 S_1(0) \quad \text{and} \quad \tilde{W}(1) = N_0(1 + R_f) + N_1 \tilde{S}_1(1)$$

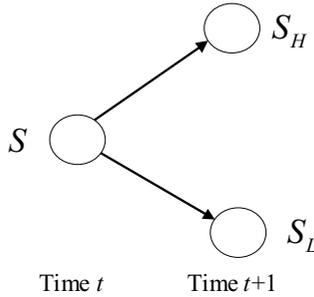


Figure 6.1

where in period 1, wealth can assume two values only since future prices are equal to either of  $(S_H, S_L)$ ,  $S_H > S_L$  and the trading strategy is defined by  $(N_0, N_1)$ . In this specific case, the price process is predictable, assuming two values only. This predictability is an essential assumption to obtain a unique value for the derivative asset, as we shall see subsequently.

For example, say that a stock has a current value of \$100 and say that a period hence (say a year), it can assume two possible values of \$140 and \$70. That is:

$$S(t) = S(=100) \quad \text{and} \quad \tilde{S}(t+1) = \begin{cases} S_H(=140) \\ S_L(=70) \end{cases}$$

The risk-free yearly interest rate is 12%, i.e.  $R_f = 0.12$ . Thus, if we construct a portfolio of  $N_0$  units of a bond worth each \$1 and  $N_1$  shares of the stock, then the portfolio investment and its future value one period hence are:

$$W(0) = N_0(1 + 0.12) + 100N_1 \quad \text{and} \quad \tilde{W}(1) = N_0(1 + 0.12) + \begin{cases} 140N_1 \\ 70N_1 \end{cases}$$

Now assume that we want to estimate the value of an option derived from such a security. Namely, consider a *call option* stating that at time  $t = 1$ , the strike time, the buyer of the option has the right to buy the security at a price of  $K$ , the exercise or strike price, with, for convenience,  $S_H \geq K \geq S_L$ . If the price is high, then the gain for the buyer of the option is  $S_H - K > 0$  and the option is exercised

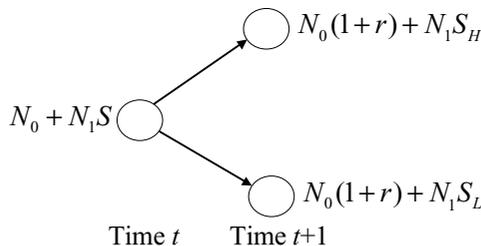


Figure 6.2

while the short seller of the option has a loss, which is  $K - S_L$ . If the price is low (below the strike price) then there is no gain and the only loss to the buyer of the call option is the premium paid for it initially. The problem we are faced with concerns the value/price of such a derived (option) contract. In other words, how much money would the (long) buyer of the option be willing to pay for this right. To find out, we proceed as follows. First, we note the possible payoffs of the option over one period and denote it by  $\tilde{C}(1)$ . Then we construct a portfolio replicating the exact cash flow associated to the option. Let the portfolio worth at the strike time be  $\tilde{W}(1)$ :

$$\tilde{W}(1) = N_0(1 + R_f) + N_1\tilde{S}(1) = \begin{cases} N_0(1 + R_f) + N_1S_H \\ N_0(1 + R_f) + N_1S_L \end{cases} \quad \text{and} \quad \tilde{W}(1) \equiv \tilde{C}(1)$$

To determine this equivalence, the portfolio composition  $N_0, N_1$  has to be determined uniquely. If it were not possible to replicate the option cash flow uniquely by a portfolio, then we would not be able to determine a unique price for the option and we would be in a situation we call incomplete. This conclusion is based on the economic hypothesis that two equivalent and identical cash flows have necessarily the same economic value (or cost). If this were not the case, there may be more than one price or no price at all for the derivative asset. Our ability to replicate a risky asset by a portfolio uniquely underlies the notion of the ‘no arbitrage’ assumption, which implies in turn the ‘law of the single price’. Thus, by constructing portfolios that have exactly the same returns with the same risks, their value ought to be the same. If this were not the case, then one of the two assets would be dominated and therefore their value could not be the same. Further, there would be an opportunity for profits that can be made with no investment – or equivalently, an opportunity for infinite rates of returns (assuming perfect liquidity of markets) that cannot be sustained (and therefore not maintain a state of equilibrium). Thus, to derive the option price, it is sufficient to estimate the replicating portfolio initial value. This is done next. Say that, for a call option, its value one period hence is:

$$\tilde{C}(1) = \begin{cases} S_H - K & \text{if the security price rises} \\ 0 & \text{if the security price decreases} \end{cases}$$

where  $S_L < K < S_H$ . A replicating portfolio investment equivalent to an option would thus be:

$$\tilde{W}(1) = \tilde{C}(1)$$

Or, equivalently,

$$\tilde{W}(1) = \tilde{C}(1) \Leftrightarrow \begin{cases} N_0(1 + R_f) + N_1S_H = S_H - K \\ N_0(1 + R_f) + N_1S_L = 0 \end{cases}$$

Note that these are two linear equations in two unknowns and have therefore a unique solution for the replicating portfolio:

$$N_1 = \frac{S_H - K}{(S_H - S_L)}, \quad N_0 = -\frac{S_L(S_H - K)}{(1 + R_f)(S_H - S_L)}$$

The procedure followed is summarized below.

$$\begin{array}{ccc} W(0) & \Leftarrow & \tilde{W}(1) \\ \Downarrow & & \Updownarrow \\ C(0) & & \tilde{C}(1) \end{array}$$

The call option's payoff is replicated by holding short bonds to invest in a stock ( $N_0 < 0$ ,  $N_1 > 0$ ). As the stock price increases, the portfolio is shifted from bonds to stocks. As a result, calling upon the 'no arbitrage' assumption, the option price and the replicating portfolio must be the same since they have identical cash flows. That is, as stated above:

$$\tilde{W}(1) = \tilde{C}(1) \Leftrightarrow W(0) = C(0) \quad \text{and} \quad \text{since: } W(0) = N_0 + N_1 S(0)$$

We insert the values for  $(N_0, N_1)$  calculated above and obtain the *call option price*:

$$C(0) = \frac{(S(1 + R_f) - S_L)(S_H - K)}{(1 + R_f)(S_H - S_L)}$$

Thus, if we return to our portfolio, and assume that the option has a strike price of \$120, then the replicating portfolio is:

$$\begin{aligned} N_1 &= \frac{S_H - K}{(S_H - S_L)} = \frac{140 - 120}{140 - 70} = \frac{2}{7} \\ N_0 &= -\frac{S_L(S_H - K)}{(1 + R_f)(S_H - S_L)} = -\frac{70(140 - 120)}{(1 + 0.12)(140 - 70)} = -\frac{20}{1.12} \end{aligned}$$

and further, the option price is:

$$W(0) = N_0 + N_1 S(0) = -\frac{20}{1.12} + \frac{200}{7} = 10.72$$

which can be calculated directly from the formula above:

$$C(0) = \frac{[S(1 + R_f) - S_L](S_H - K)}{(1 + R_f)(S_H - S_L)} = \frac{(100(1.12) - 70)(140 - 120)}{(1 + 0.12)(140 - 70)} = 10.72$$

By the same token, say that the current price of a stock is  $S = \$100$  while the price a period hence (at which time the option may be exercised) is either  $S_H = \$120$  or  $S_L = \$70$ . The strike price is  $K = \$110$  while the discount rate over the relevant period is 0.03. Thus, a call option taken for the period on such a stock has a price, which is given by:

$$C(0) = \frac{(100 - 70)(120 - 110)}{(1 + 0.03)(120 - 70)} = \$5.825$$

### 6.1.1 Option valuation and rational expectations

The rational expectations hypothesis claims that an expectation over 'future prices' determines current prices (see Figure 6.3). That is to say, assuming that

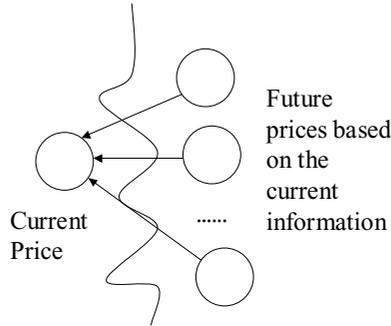


Figure 6.3

rational expectations hold, there is a probability measure that values the option in terms of its expected discounted value at the risk-free rate, or

$$C(0) = \frac{1}{1 + R_f} E^* \tilde{C}(1)$$

where  $E^*$  is an expectation taken over the appropriate probability measure assumed to exist (in our current case it is given by  $[p^{*,1} - p^*]$ ) and therefore:

$$C(0) = \frac{1}{1 + R_f} [p^* C(1|S_H) + (1 - p^*) C(1|S_L)]$$

where  $C(1|S_H) = S_H - K$ ,  $C(1|S_L) = 0$  are the option value at the exercise time and  $p^*$  denotes a 'risk-neutral probability'. *This probability is not, however, a historical probability of the stock moving up or down but a 'risk-neutral probability', making it possible to value the asset under a risk-neutrality assumption.* In this case, the option's price is the discounted (at a risk-free rate) expected value of the option,

$$C(0) = \frac{1}{1 + R_f} [p^*(S_H - K) + (1 - p^*)(0)]$$

And, using the value of the option found earlier, we have:

$$0 \leq p^* = \frac{1}{S_H - S_L} [(1 + R_f)S - S_L] \leq 1$$

In our previous example, we have:

$$0 \leq p^* = \frac{1}{120 - 70} [(1 + 0.03)100 - 70] = 0.66$$

By the same token, we can verify that:

$$S = \frac{1}{1 + .03} [(0.66)(120) + (1 - 0.66)(70)] = 100$$

$$C(0) = \frac{1}{1 + .03} [(0.66)(120 - 110)] = \frac{(0.66)(10)}{1.03} = 5.825$$

with  $p^* = 0.66$ . This ‘risk-neutral probability’ is determined in fact by traders in financial markets interacting with others in developing the financial market equilibrium – where profits without risk cannot be realized. For this reason, ‘risk-neutral pricing’ is ‘determined by the market and provides the appropriate discount mechanism to value the asset in the following form (see also Chapter 3 and our discussion on the stochastic discount factor):

$$C(0) = E\{m_1 \tilde{C}(1)\}; \quad m_1 = \frac{1}{1 + R_f}$$

Risk-neutral probabilities, as we have just seen, allow a linear valuation of the option which hinges on the assumption of no arbitrage. Nonetheless, the existence of risk-neutral probabilities do not mean that we can use linear valuation, for to do so requires markets completeness (expressed by the fact in this section that we were able to replicate by portfolio the option value and derive a unique price of the option). In subsequent chapters, we shall be concerned with market incompleteness and see that this is not always the case. These situations will complicate the valuation of financial assets in general.

### 6.1.2 Risk-neutral pricing

The importance of risk-neutral pricing justifies our considering it in greater depth. In many instances, security prices can be conveniently measured with respect to a given process – in particular, a growing process called the *numeraire*, expressing the value of money (money market), a bond or some other asset. That is, allowing us to write (see also Chapter 3):

$$V(S(t)) = \frac{1}{1 + R_f} E^*(V(\tilde{S}(t + 1))) = \frac{1}{1 + R_f} [p^* V(S_H) + (1 - p^*) V(S_L)]$$

$p^*$  is said to be a ‘risk-neutral probability’ and  $R_f$  is a risk-free discount rate. And for an option (since  $R_f$  has a fixed value):

$$C(t) = E^* \left[ \frac{1}{1 + R_f} \tilde{C}(t + 1) \right] = \frac{1}{1 + R_f} E^* [\tilde{C}(t + 1)]$$

In general, for any value (whether it is an option or not)  $\tilde{V}_i$  at time  $i$  with a risk-free rate  $R_f$ , we have, over one period:  $V_0 = E^* \left( \frac{1}{1 + R_f} \tilde{V}_1 \right)$

By iterated expectations, we have as well:

$$V_1 = E^* \left( \frac{1}{1 + R_f} \tilde{V}_2 \right) \quad \text{and}$$

$$V_0 = \frac{1}{(1 + R_f)} E^* \left( E^* \left( \frac{1}{1 + R_f} \tilde{V}_2 \right) \right) = E^* \left( \frac{1}{(1 + R_f)^2} \tilde{V}_2 \right)$$

and therefore, over  $n$  periods:

$$V_0 = E^* \left( \frac{1}{(1 + R_f)^n} \tilde{V}_n \right) = \frac{1}{(1 + R_f)^n} E^*(\tilde{V}_n)$$

If we set  $\Phi_0$ , the information regarding the process initially, then we write

$$V_0(1 + R_f)^n = E^*(\tilde{V}_n | \Phi_0)$$

Further, application of iterated expectations has shown that this discounting process defines a martingale. Namely, we have:

$$V(S_0) = (1 + R_f)^{-k} V(S_k) = E^* \{ (1 + R_f)^{-(k+n)} V(S_{k+n}) | \Phi_k \};$$

$$k = 0, 1, 2, \dots \quad \text{and} \quad n = 1, 2, 3, \dots$$

or, equivalently,

$$V(S_k) = E^* \{ (1 + R_f)^{-n} V(S_{k+n}) | \Phi_k \}; \quad k = 0, 1, 2, \dots \quad \text{and} \quad n = 1, 2, 3, \dots$$

This result can be verified next using our binomial model. Set the unit one period risk-free bond,  $B(t) = B(t, t + 1)$  for notational convenience, then discounting a security price with respect to the risk-free bond yields:

$$S^*(t) = \frac{S(t)}{B(t)} \quad \text{or} \quad S^*(t) = \frac{S(t)}{(1 + R_f)^t}$$

and  $S^*(t)$  is a martingale. Generally, under the risk-neutral measure,  $\mathbf{P}^*$  the discounted process

$$\{ (1 + R_f)^{-k} S_k | \Phi_k \}, \quad k = 0, 1, 2, \dots$$

is, as we saw earlier, a martingale. Here again, the proof is simple since:

$$E^* \{ (1 + R_f)^{-(k+1)} S_{k+1} | \Phi_k \} = (1 + R_f)^{-k} S_k$$

and

$$E^* \{ (1 + R_f)^{-(k+1)} S_{k+1} | \Phi_k \} = (1 + R_f)^{-(k+1)} \left[ p^* \frac{S_H}{S_k} + q^* \frac{S_L}{S_k} \right] S_k$$

$$= (1 + R_f)^{-(k+1)} S_k [(1 + R_f)] = (1 + R_f)^{-k} S_k$$

This procedure remains valid if we consider a portfolio which consists of a bond and  $m$  stocks. In this case, dropping for simplicity the tilde over random variables, we have:

$$W^*(t) = N_0 + N_1 S_1^*(t) + N_2 S_2^*(t) + \dots + N_m S_m^*(t)$$

$$W^*(t + 1) = N_0 + N_1 S_1^*(t + 1) + N_2 S_2^*(t + 1) + \dots + N_m S_m^*(t + 1)$$

and

$$\Delta W^*(t) = N_1 \Delta S_1^*(t) + N_2 \Delta S_2^*(t) + \dots + N_m \Delta S_m^*(t)$$

Equating these to the value of some derived asset, a period hence:

$$W^*(t + 1) = C^*(t + 1)$$

leads to a solution for  $(N_0, N_1, N_2, \dots, N_m)$  where  $C^*(t + 1)$  is a vector of assets we use to construct a riskless hedge and replicate the derivative product we wish to estimate (Pliska (1997) and Shreve *et al.* (1997) for example).

### Example: Options and portfolios holding cost

Consider now the problem of valuing the price of a call option on a stock when the alternative portfolio consists in holding a risky asset (a stock) and a bond, for which there is a 'holding cost'. This cost is usually the charge a bank may require for maintaining in its books an investor's portfolio. In this case, the hedging portfolio is given by equating:

$$\tilde{W}(1) = \begin{cases} N_0(1 + R_f - c_B) + N_1(S_H - c_S) \\ N_0(1 + R_f - c_B) + N_1(S_L - c_S) \end{cases}$$

where  $c_B$  is the bond holding cost and  $c_S$  is the stock holding cost. The option's cash flow is:

$$\tilde{C}(1) = \begin{cases} S_H - K & \text{if the security price rises} \\ 0 & \text{if the security price decreases} \end{cases}$$

This leads to:

$$\begin{cases} N_0(1 + R_f - c_B) + N_1(S_H - c_S) = S_H - K \\ N_0(1 + R_f - c_B) + N_1(S_L - c_S) = 0 \end{cases}$$

and

$$N_1 = \frac{S_H - K}{S_H - S_L}, N_0 = -\frac{(S_H - K)(S_L - c_S)}{(S_H - S_L)(1 + R_f - c_B)}$$

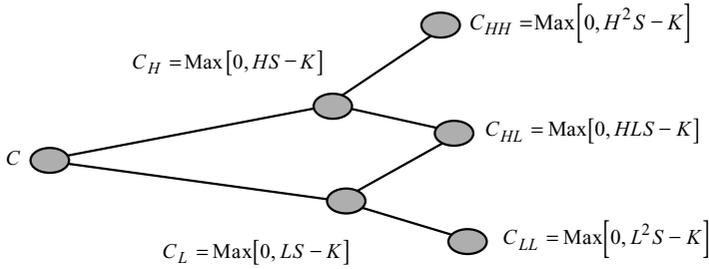
Therefore, the option price is equal instead to:

$$C(0) = N_1 S + N_0 = S \frac{S_H - K}{S_H - S_L} + \frac{(S_H - K)(S_L - c_S)}{(S_H - S_L)(1 + R_f - c_B)}$$

For example, if we use the data used in the previous option's example with  $S = 100$ ,  $S_H = 140$ ,  $S_L = 70$ ,  $K = 120$ ,  $R_f = 0.12$  and the 'holding costs' are:  $c_S = 5$ ,  $c_B = 0.02$ , then

$$\begin{aligned} C(0) &= 100 \frac{140 - 120}{140 - 70} - \frac{(140 - 120)(70 - 5)}{(140 - 70)(1 + 0.12 - 0.02)} \text{ or} \\ C(0) &= \frac{200}{7} - \frac{(20)(65)}{(70)(1.1)} = 28.57 - 16.88 = 11.68 \end{aligned}$$

which compares to a price of 10.64 without the holding cost. In this sense, holding costs will increase the price of acquiring the option. A general approach to this problem is treated by Bensoussan and Julien (2000) in continuous-time models. The costs of holding, denoted friction costs, are, however, far more complex, leading to incompleteness.



**Figure 6.4** A two-period binomial tree.

### 6.1.3 Multiple periods with binomial trees

Over two or more periods, the problem remains the same. For one period, we saw that the price of a call option is:  $C_H = \text{Max}[0, S_H - K]$ ;  $C_L = \text{Max}[0, S_L - K]$  and by risk-neutral pricing,

$$C(0) = \frac{1}{1 + R_f} E^* \tilde{C}(1) = \frac{1}{1 + R_f} [p^* C_H + (1 - p^*) C_L]$$

Over two periods, we have:

$$C(0) = \frac{1}{(1 + R_f)^2} E^* \tilde{C}(2)$$

$$C_H = \frac{1}{1 + R_f} [p^* C_{HH} + (1 - p^*) C_{HL}]; \quad C_L = \frac{1}{1 + R_f} [p^* C_{HL} + (1 - p^*) C_{LL}]$$

which we insert in the previous equation, to obtain the option price for two periods (see Figure 6.4). Explicitly, we have the following calculations:

$$C(0) = \frac{1}{(1 + R_f)} E^* \tilde{C}(1) = \frac{1}{(1 + R_f)} E^* \left\{ \frac{1}{(1 + R_f)} E^* \tilde{C}(2) \right\} \text{ or}$$

$$C(0) = \left( \frac{1}{1 + R_f} \right)^2 [p^{*2} C_{HH} + 2p^*(1 - p^*) C_{HL} + (1 - p^{*2}) C_{LL}]$$

Generally, the price of a call option at time  $t$  whose strike price is  $K$  at time  $T$  can be calculated recursively by:

$$C(t) = E^* \left( \frac{1}{1 + R_f} \tilde{C}(t + 1) \right); \quad C(T) = \text{Max}[0, S(T) - K]$$

Explicitly, if we set,  $S_H = HS$ ,  $S_L = LS$ , we have :

$$C(0) = \frac{1}{(1 + R_f)^2} \left\{ \begin{array}{l} p^{*2}(H^2S - K)^{++} \\ 2p^*(1 - p^*)(HLS - K)^{++} \\ (1 - p^*)^2(H^2S - K)^+ \end{array} \right\} = \frac{1}{(1 + R_f)^2} \sum_{j=0}^2 \binom{2}{j} p^{*j} (1 - p^*)^{2-j} \{H^j L^{2-j} S - K\}^+$$

We generalize to  $n$  periods and obtain by induction:

$$C(0) = \frac{1}{(1 + R_f)^n} \left\{ \sum_{j=0}^n \binom{n}{j} p^{*j} (1 - p^*)^{n-j} (H^j L^{n-j} S - K)^+ \right\}$$

We can write this expression in still another form:

$$C_n = \frac{1}{(1 + R_f)^n} E\{(S_n - K)^+\} = \frac{1}{(1 + R_f)^n} \sum_{j=0}^n P_j (\tilde{S}_j - K)^+$$

where

$$P_j = P(S_n = H^j L^{n-j} S) = \binom{n}{j} p^{*j} (1 - p^*)^{n-j}$$

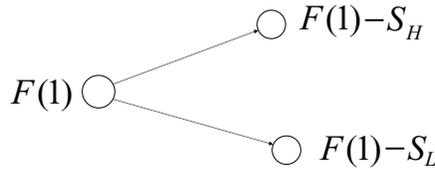
are the risk-neutral probabilities. This expression is of course valid only under the assumption of no arbitrage. This mechanism for pricing options is generally applicable to other types of options, however, such as American, Look-Back, Asiatic, esoteric and other options, as we shall see later on.

The option considered so far is European since exercise of the option is possible only at the option's maturity. American options, unlike European ones, give the buyer the right to exercise the option before maturity. The buyer must therefore take into account to optimal timing of his exercise. An option exercised too early may forgo future opportunities, while exercised too late it may lose past opportunities. The optimal exercise time will be that time that balances the live value of the option versus its 'dead' or exercise value. The recursive solution of the European call option can be easily modified for the exercise feature of the American option. Proceeding backward from maturity, the option will be exercised when its 'dead' value is larger than its 'live' one. Technically, the exercise time is a stopping time, as we shall see subsequently. Note that early exercise of the option is optimal only if the option value diminishes. For a call option (and in the absence of dividends), it does not diminish over time and therefore it will never pay to exercise an option early. For this reason we note that the price of a European and an American call are equal. For a put option, the present value of the payoff is a decreasing function of time hence, early exercise is possible irrespective of the existence of dividend payments.

## 6.2 FORWARD AND FUTURES CONTRACTS

A forward contract is an agreement to buy or sell an asset at a fixed date for a price determined today. The buyer agrees to buy the asset at the price  $F$  and sell it at the market price at maturity for a payoff  $S - F$ . The seller takes the opposite position and sells at the market price  $F$  and buys the asset at the market price  $S$  at maturity.

Forward contracts are thus an agreement between two parties or traders regarding the price, the *delivery price*, of a stock, a commodity or any another asset, settled at some future time – the *maturity*. Unlike options, forward contracts are



**Figure 6.5** Forward contract valuation.

an obligation to be maintained by the buyer and the seller at maturity. The party that has agreed to buy the forward contract is said to assume a long position while the party that agrees to sell is said to assume the short position. Such contracts allow for the parties to exchange the price risk at maturity. For example, a wheat farmer may be exposed to a fall of the wheat price when he brings it to market. He can then enter in a forward contract to sell his wheat today at the fixed price  $F$ . At maturity, he may sell wheat at a predetermined price and buy it at the spot rate  $S$  from the buyer (say, the baker) of the forward for a payoff of  $(F - S)$ . The buyer (baker) takes the opposite position for a payoff of  $(S - F)$ . Both sell and buy are in the market and their position is  $[(F - S) + S = F]$  and  $[(S - F) - S = F]$  respectively. The parties have therefore perfectly eliminated their wheat price risk as their payoffs are determined at the initiation of the contract. In this example, we evolved into a world where risk can be completely shifted away, which is also the risk-neutral world that conveniently discounts risky payoffs at the risk-free rate (under an appropriately defined probability measure). This transformation to the 'risk-neutral world' breaks down when a seller cannot find a buyer with the exact opposite hedging needs and vice versa. In this case, speculators are needed to take on the risk and a risk neutral world will no longer exist. Depending on whether excess hedging is in long or short forwards, the pressure will be upward or downward compared to the risk-neutral price.

To calculate the forward price at times  $t = 1$  and  $t = 2$ , say  $F(1)$  and  $F(2)$  we proceed as follows. Consider the first period only, at which the gain can be either  $F(1) - S_H$  in case of a price increase or  $F(1) - S_L$  in case of a price decrease (see Figure 6.5). Initially nothing is spent and therefore, initially we also get nothing. At present it is thus worth nothing. Assuming no arbitrage (otherwise we would not be able to use the risk-neutral probability), and proceeding as in the previous section, we have:

$$0 = \frac{1}{1 + R_f} [p^*(F(1) - S_H) + q^*(F(1) - S_L)]; p^* + q^* = 1$$

which is an one equation in one unknown and where  $R_f$  is an effective risk-free annual rate. The forward price  $F(1)$  resulting from the solution of the equation above is therefore:

$$F(1) = [p^* S_H + q^* S_L] = S(1 + R_f)$$

In other words, *the one period forward price equals the discounted current spot price*. For two periods we note equivalently that when the spot price is  $S_H$  or  $S_L$ ,

then (from period 1 to 2):

$$\tilde{F}(2) = \begin{cases} p^* S_{HH} + q^* S_{HL} = (1 + R_f) S_H & \text{w.p. } p^* \\ p^* S_{HL} + q^* S_{LL} = (1 + R_f) S_L & \text{w.p. } q^* \end{cases}$$

As a result,  $F(2) = E^* \tilde{F}(2) = p^*(1 + R_f) S_H + q^*(1 + R_f) S_L$  and therefore  $F(2) = (1 + R_f)^2 S$  and obviously:

$$F(n) = (1 + R_f)^n S$$

This means that the  $n$  periods forward price equals the  $n$  periods discounted current spot price (see also Figure 6.4). Of course, using the risk-neutral reasoning, since there is no initial expenditure at the time the forward contract is signed, while at time  $t$ , the profit realized equals the difference between the current price and the forward (agreed) on price at time zero which we write by  $F(n)$ , we have:

$$0 = \frac{1}{(1 + R_f)^n} E^* [S_n - F(n)] \quad \text{and} \quad F(n) = E^* [S_n]$$

Since under risk-neutral pricing,

$$S_0 = \frac{1}{(1 + R_f)^n} E^* (S_n) \rightarrow E^* (S_n) = S_0(1 + R_f)^n$$

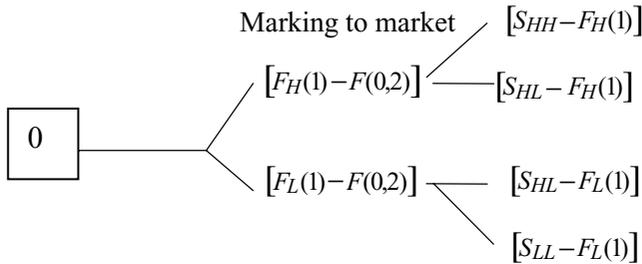
we obtain at last the general forward price:

$$F(n) = S_0(1 + R_f)^n$$

In practice, there may be some problems because decision makers may use forward prices to revalue the spot price. Feedback between these markets can induce an opportunity for arbitrage. Further, it is also necessary to remember that we have assumed a risk-neutral world. As a result, when traders use historical data, there may again be some problems, leading to a potential for arbitrage since the fundamental assumption of rational expectations is violated. For example, if the spot price of silver is \$50, while the delivery price is \$53 with maturity in one year, while interest rates equal 0.08, then the no arbitrage price is:  $50(1 + 0.08) = \$54$ . This provides an arbitrage opportunity since in one year there is an arbitrage profit of  $\$1 (= 54 - 53)$  that can be realized.

A *futures contract* differs from a forward contract in that it is standardized, openly traded and marked to market. Marking to market involves adjusting an investor's initial margin deposit by the change in the futures contract price each day. If the investor's margin account falls below the maintenance margin, the trader asks the investor to fill the margin account back to the initial margin, posted in the form of interest-bearing T-bonds.

A *futures price* is determined as follows. The futures price one period hence  $F(0, 1)$  at time  $t = 1$  is set equal to the forward price for that time, since no cost is incurred. In other words, we have,  $F(0, 1) = F(1)$ . Now consider the futures price in two periods,  $F(0, 2)$ . If the spot price increases to  $S_H$ , the futures price turns out to equal the one-period forward price, or  $F_H(1)$  (since only one more period is left till the exercise time). Similarly, if the spot price decreases to  $S_L$ ,



**Figure 6.6** Future price valuation.

the future price is now  $F_L(1)$ . As a result, cash flow payments at the first and second periods are given by Figure 6.6. Initially, *the value of these flows is worth nothing, since nothing is spent and nothing is gained*. Thus, an expectation of futures flows is worth nothing today. That is,

$$0 = \frac{1}{1 + R_f} [p^* (F_H(1) - F(0, 2)) + q^* (F_L(1) - F(0, 2))] \\ + \frac{1}{(1 + R_f)^2} \left[ \begin{aligned} & p^{*2} (S_{HH} - F_H(1)) + p^* q^* (S_{HL} - F_H(1)) + \\ & p^* q^* (S_{HL} - F_L(1)) + q^{*2} (S_{LL} - F_L(1)) \end{aligned} \right]$$

which is one equation and three unknowns. However, noting that for the one-period futures (forward) price, we have:

$$F_H(1) = (1 + R_f)S_H, \quad F_L(1) = (1 + R_f)S_L$$

Inserting these results into our equation, we obtain the futures price:  $F(0, 2) = (1 + R_f)^2 S_0$  which is equal to the forward price. This is the case, however, because the discount interest rate is deterministic. In a stochastic interest rate framework, this would not be the case. A generalization to  $n$  periods yields:

$$F(0, n) = (1 + R_f)^n S_0$$

Futures contracts are stated often in terms of a *basis*, measuring the difference between the spot and the futures price. The basis may be mis-priced, however, because of mismatching of assets (cross-hedged), because of maturity (forward versus futures) and the quality of related assets (options). There are some fundamental differences between forward and futures contracts that we summarize in Table 6.1. These relate to the hedging quality of these financial products, their barriers to entry, etc. Further, although under risk-neutral pricing they have the same price, in practice (when interest rates are stochastic as stated above) they can differ appreciably. In many cases, futures contracts are preferred to forward contracts simply because they are more liquid and thereby more ‘tradable’.

### Example

We compare the consequences of forward and futures contracts on a volume of 100 Dax shares each worth 77E over say five periods. We obtained the following

**Table 6.1** Forward and futures contracts: contrasts.

	Forward	Futures
Market	OTC (Private)	Exchange markets
Standard contract	No	Yes
Barrier to entry	Substantial	Weak
Security	Individual	Margin system
Daily controls	No	Yes
Flexibility	Inverse contract	Long–short
Hedge quality	Best	Problematic

results, pointing to differences in cash flow (Table 6.2). Calculations are performed as follows. The cash flows associated with a forward contract of five periods (denoted by \*) and associated with a futures contract at period 2 (denoted by \*\*) are given in Table 6.2.

Further, note that the sum of payments of mark to market are equal to the sum of payments of the forward since their initial prices (investment) were the same.

**Table 6.2**

T	1	2	3	4	5
DAX	7700	7770	7680	7650	7730
Price forward	7800	—	—	—	7730
Price future	7800	7845	7730	7675	7730
Cash flow forward	—	—	—	—	-7000(*)
Cash flow future	—	+4500(**)	-11500	-5500	+5500

(\*) =  $(F[1;T] - F[0;T]) * \text{volume} = (7730 - 7800) * 100 = 7000$

(\*\*) =  $(7845 - 7800) * 100 = 4500$

### Example

Say that  $K$  is the forward delivery price with maturity  $T$  while  $F$  is the current forward price. The value of the long forward contract is then equal to the present value of their difference at the risk-free rate  $R_f$ , or  $P_L = (F - K) e^{-R_f T}$ . Similarly, the value of the short forward contract is  $P_S = -P_L = (K - F) e^{-R_f T}$ .

### Example: Futures on currencies

Let  $S$  be the dollar value of a euro and let  $(R_\$, R_E)$  be the risk-free rate of the local (dollar) and the foreign currency (euro). Then the relative rate is  $(R_\$ - R_E)$  and the future euro price  $T$  periods hence is:  $F = S e^{(R_\$ - R_E)T}$ .

## 6.3 RISK-NEUTRAL PROBABILITIES AGAIN

Risk-neutral probabilities, conveniently, allow linear pricing measures. These probabilities are defined in terms of market parameters (although their existence

hinges importantly on a risk-free rate,  $R_f$ , and rational traders) and differ markedly from historical probabilities. This difference, contrasting two cultures, is due to economic assumptions that the market price of a traded asset ‘internalizes’ all the past, future states and information that such an asset can be subjected to. If this is the case, and it is so in markets we call *complete markets*, then the current price ought to be determined by its future values as we have shown here. In other words, *the market determines the price and not historical (probability) uncertainty! If there is no unique set of risk-neutral pricing measures, then market prices are not unique and we are in a situation of market incompleteness, unable to value the asset price uniquely.*

It is therefore important to establish conditions for market completeness. Our ability to construct a unique set of risk neutral probabilities for the valuation of the stock at period 1 or the value of buying an option depends on a number of assumptions that are of critical importance in finance and must be maintained theoretically and practically. Pliska (1997) for example, emphasizes the importance of these assumptions and their implications for risk-neutral probabilities. Namely, there can be a linear pricing measure if and only if there are no dominating trading strategies. Further, if there are no dominant trading strategies, then the law of the single price holds, albeit the converse need not necessarily be true. And finally, if there were a dominating trading strategy, then there exists an arbitrage opportunity, but the converse is not necessarily true. Thus, risk-neutral pricing requires, as stated earlier:

- No arbitrage opportunities.
- No dominant trading strategies.
- The law of the single price.

When the assumption of market completeness is violated, it is no longer possible to obtain a unique set of *risk-neutral probabilities*. This means that one cannot duplicate the option with a portfolio or price it uniquely. In this case, an appropriate portfolio is optimized for the purpose of selecting risk-neutral probabilities. Such an optimization problem can be based on the best mean forecast as we shall outline below. These probabilities will, however, be a function of a number of parameters implied by the portfolio and decision makers’ preferences and of course the information available to the decision maker. When this is not possible, we can, for a given set of parameters bound the relevant option prices.

### 6.3.1 Rational expectations and optimal forecasts

Rational expectations mean that economic agents can forecast the ‘mean’ price (since risk-neutral probabilities imply that an expected value is used to value the asset). In this case, a mean forecast can be selected by minimizing the forecast error (in which case the mean error is null). Explicitly, say that  $\{x\} = \{x_1, x_2, \dots, x_T\}$  stands for an information set (a time series, a stock price record, financial variables etc.). A forecast is thus an estimate based on the information set  $\{x\}$  written for convenience by the function  $f(\cdot)$  such that  $\bar{y} = f(x)$  whose error forecast is

$\varepsilon = y - \bar{y}$  where  $y$  is the actual record of the series investigated and its forecast is obtained by minimizing the least squares errors. Assume that the forecast is unbiased, that is, based on all the relevant information available,  $I$ ; the forecast equals the conditional expectation, or  $\bar{y} = E(y | I)$  whose error is  $\varepsilon = y - \bar{y} = y - E(y | I)$ . In this case, rational expectations exist when the expected errors are both null and uncorrelated with its forecast as well as with any observation in the information set. This is summarized by the following three conditions of rational expectations:

$$E(\varepsilon) = 0; E(\varepsilon \bar{y}) = E(\varepsilon E(y | I)) = 0; E(\varepsilon x) = \text{cov}(\varepsilon, x) = 0, \forall x \in I$$

Of course, there can be various information sets as well as various mechanisms that can be used to generate rational expectations. However, it is essential to note that the behaviour of forecast residual errors determine whether these forecasts are rational expectations forecasts or not.

## 6.4 THE BLACK-SCHOLES OPTION FORMULA

In continuous time and continuous state, the procedure for pricing options remains the same but its derivation is based on stochastic calculus. We shall demonstrate how to proceed by developing the Black-Scholes model for the valuation of a call option. Let  $S(t)$  be a security-stock price at time  $t$  and let  $W$  be the value of an asset derived from this stock which we can write by the following function  $W = f(S, t)$  assumed to be differentiable with respect to time and the security-stock  $S(t)$ . For simplicity, let the security price be given by a lognormal process:

$$dS/S = \alpha dt + \sigma dw, \quad S(0) = S_0$$

The procedure we follow consists in a number of steps:

- (1) We calculate  $dW$  by applying Ito's differential rule to  $W = f(S, t)$
- (2) We construct a portfolio  $P$  that consists of a risk-free bond  $B$  and number 'a' of shares  $S$ . Thus,  $P = B + aS$  or  $B = P - aS$ .
- (3) A perfect hedge is constructed by setting:  $dB = dP - a dS$ . Since,  $dB = R_f B dt$ , this allows determination of the stockholding 'a' in the replicating portfolio.
- (4) We equate the portfolio and option value processes and apply the 'law of the single price to determine the option price. Thus, setting  $dP = dW$ , we obtain a second-order partial differential equation with appropriate boundary conditions and constraints, providing the solution to the option price.

Each of these steps is translated into mathematical manipulations. First, by an application of Ito's differential rule we obtain the option price:

$$\begin{aligned} dW &= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial S} dS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} (dS)^2 = \\ &= \left( \frac{\partial f}{\partial t} + \alpha S \frac{\partial f}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2} \right) dt + \left( \sigma S \frac{\partial f}{\partial S} \right) dw \end{aligned}$$

Next, we construct a replicating riskless portfolio by assuming that the same amount of money is invested in a bond whose riskless (continuously compounded) rate of return is  $R_f$ . In other words, instead of a riskless investment, say that we sell  $a$  units of the asset at its price  $S$  (at therefore at the cost of  $aS$ ) and buy an option whose value is  $W$ . The return on such a transaction is  $W - aS$ . In a small time interval, this will be equal:  $dW - a dS$ . To replicate the bond's rate,  $R_f$ , we establish an equality between  $dB$  and  $dW - a dS$ , thus:  $dB \equiv dW - a dS$ . This argument implies no arbitrage, i.e. the risk-free and the 'risky' market rates should yield an equivalent return, or

$$dB = dW - a dS = dW - a(\alpha S dt + \sigma S dw)$$

Inserting  $dW$ , found above by application of Ito's Lemma, we obtain the following stochastic differential equation:

$$dB = \mu dt + \left( \frac{\partial f}{\partial S} - a \right) S \sigma dw; \quad \mu = \alpha S \frac{\partial f}{\partial S} - a \alpha S + \frac{\partial f}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2}$$

Since  $dB = R_f B dt$ , this is equivalent to:  $\mu dt = R_f B dt$  and

$$\left( \frac{\partial f}{\partial S} - a \right) S \sigma = 0$$

These two equalities lead to the following conditions:

$$a = \frac{\partial f}{\partial S} \quad \text{and} \quad R_f S = \alpha S \frac{\partial f}{\partial S} - a \alpha S + \frac{\partial f}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2}$$

Inserting the value of  $a$ , we obtain:

$$R_f f = \frac{\partial f}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2}$$

Since  $B = f - aS$ , then inserting in the above equation we obtain the following second-order differential equation in  $f(S, t)$ , the price of the derived asset:

$$-\frac{\partial f}{\partial t} = R_f S \frac{\partial f}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2} - R_f f$$

To obtain an explicit solution, it is necessary to specify a boundary condition. If the derived asset is a European call option, then there are no cash flows arising from the European option until maturity. If  $T$  is the exercise date, then clearly,

$$f(0, t) = 0, \quad \forall t \in [0, T]$$

At time  $T$ , the asset price is  $S(T)$ . If the strike (exercise price) is  $K$ , then if  $S(T) > K$ , the value of the call option at this time is  $f(S, T) = S(T) - K$  (since the investor can exercise his option and sell back the asset at its market price at time  $T$ ). If  $S(T) \leq K$ , the value of the option is null since it will not be worth exercising. In other words,

$$f(S, T) = \text{Max} [0, S(T) - K]$$

This final condition, together with the asset price partial differential equation, can be solved providing thereby a valuation of the option, or the option price. Black and Scholes, in particular have shown that the solution is given by:

$$W = f(S, t) = S\Phi(d_1) - Ke^{-R_f t}\Phi(d_2)$$

where

$$\Phi(y) = (2\pi)^{-1/2} \int_{-\infty}^y e^{-u^2/2} du; \quad d_1 = \left[ \frac{\log(S/K) + (T-t)(R_f + \sigma^2/2)}{\sigma\sqrt{T-t}} \right];$$

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

This result is remarkably robust and holds under very broad price processes. Further, it can be estimated by simulation very simply. There are many computer programs that compute these options prices as well as their sensitivities to a number of parameters.

The price of a put option is calculated in a similar manner and is therefore left as an exercise. For an American option, the value of the call equals that of a European call (as we have shown earlier). While for the American put, calculations are much more difficult, although we shall demonstrate at the end of this chapter how such calculations are made.

Properties of the European call are easily calculated using the explicit equation for the option value. It is simple to show that the option price has the following properties:

$$\frac{\partial W}{\partial S} \geq 0, \quad \frac{\partial W}{\partial T} > 0, \quad \frac{\partial W}{\partial K} < 0, \quad \frac{\partial W}{\partial R_f} > 0$$

They express the option's sensitivity. Intuitively, the price of a call option is the discounted expected value (with risk-neutral probabilities) of the payoff  $f(S, T) = \text{Max}[0, S(T) - K]$ . The greater the stock price at maturity (or the lower the strike price), the greater the option price. The higher the interest rate, the greater the discounting of the terminal payoff and thus the stock price increases as it grows at the risk-free rate in the risk-neutral world. The net effect is an increase in the call option price. The longer the option's time to maturity the larger the chances of being in the money and therefore the greater the option price. The call option is therefore an increasing function of time. The higher the stock price volatility, the larger the stock option price. Because of the correspondence between the option price and the stock price volatility, traders often talk of volatility trading rather than options trading, trading upward on volatility with calls and downward with puts.

A numerical analysis of the Black-Scholes equation will reveal these relationships in fact. For example, if we take as a reference point a call option whose strike price is \$160, the expiration date is 5 months, stock current price is \$140, volatility is 0.5 and the compounded risk-free interest rate is 0.15, then the price of this option will turn out to be \$81.82. In other words, given the current parameters, an investor will be willing to pay \$81.82 for the right to buy the stock at a

price of \$160 over the next 5 months. If we let the current price vary, we obviously see that as the stock price increases (at the time the call option is acquired), the price of the option increases and vice versa when the current stock prices decreases.

- Variation in the current stock price

Stock price	120	130	140	150	160
Option price	65.01	73.33	81.82	90.46	90.22

- Variation in the expiration date

Expiration $T$	3	4	5	6	7
Option price	60.53	72.14	81.82	89.98	96.92

- Variation in the strike price

Strike price	140	150	160	170	180
Option price	86.81	84.26	81.82	79.50	77.27

- Variation in volatility

Volatility	0.3	0.4	0.5	0.6	0.7
Option price	68.07	74.76	81.82	88.82	95.52

- Variation in the risk-free compounded interest rate

Interest rate	0.05	0.10	0.15	0.20	0.25
Option price	63.80	73.15	81.82	89.68	96.67

Variations in the strike price, the expiration date of the option, the stock volatility and the compounded risk-free interest rate are outlined below as well. Note that when the stock price, the expiration date, the volatility and the interest rate

increase, the option price increases; while when the strike price increases, the option price declines. However, in practice, we note that beyond some level of the strike, the option price increases — this is called the smile and will be discussed in Chapter 8.

Call and put options are broadly used by fund managers for the leverage they provide or to cover a position. For example, the fund manager may buy a call option out of the money (OTM) in the hope that he will be in the money (ITM) and thus make an appreciable profit. If the fund manager owns an important number of shares of a given stock, he may then buy put options at a given price (generally OTM). In the case of a loss, the stock price decline might be compensated by exercising the put options. Unlike the fund manager, a trader can use call options to ‘trade on volatility’. If the trader buys a call, the price paid will be associated to the implicit volatility (namely a price calculated based on Black–Scholes option value formula). If the volatility is in fact higher than the implicit volatility, the trader can probably realize a profit, and vice versa.

#### 6.4.1 Options, their sensitivity and hedging parameters

Consider a derived asset as a function of its spot price  $S$ , time  $t$ , the standard deviation (volatility  $\sigma$ ) and the riskless interest rate  $R_f$ . In other words, we set the derived asset price as a function  $f \equiv f(S, t, \sigma, R_f)$  whose solution is known. Consider next small deviations in these parameters, then by Taylor series approximation, we can write:

$$df = \frac{\partial f}{\partial S} dS + \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial \sigma} d\sigma + \frac{\partial f}{\partial R_f} dR_f + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} (dS)^2 + \frac{1}{2} \frac{\partial^2 f}{\partial t^2} (dt)^2 + \dots$$

Terms with coefficients of order greater than  $dt$  are deemed negligible (for example  $\partial^2 f / \partial t^2$ ). Each of the terms in the Taylor series expansion provides a measurement of local sensitivity with respect to the parameter defining the derivative price. In particular, in financial studies the following ‘Greeks’ are defined:

$$DELTA = \Delta = \frac{\partial f}{\partial S} : \text{sensitivity to the spot price}$$

$$THETA = \Theta = \frac{\partial f}{\partial t} : \text{sensitivity to time to expiration}$$

$$VEGA = \nu = \frac{\partial f}{\partial \sigma} : \text{sensitivity to volatility}$$

$$RHO = \rho = \frac{\partial f}{\partial R_f} : \text{sensitivity to the interest rate}$$

$$GAMMA = \Gamma = \frac{\partial^2 f}{\partial S^2}$$

Inserting into the derived asset differential equation, we have:

$$dF = \Delta dS + \Theta dt + \nu d\sigma + \rho dR_f + \frac{1}{2} \Gamma (dS)^2$$

For the sensitivity equations to make sense, however, the option price, a function of the spot price must assume certain mathematical relationships that imply convexity of the option price. Evidently, this equation will differ from one derived asset to another (for example, for a bond, a currency, a portfolio of securities etc. we will have an equation which expresses the parameters at hand and of course the underlying partial differential equation of the derived asset). The ‘Greeks’ can be calculated easily using widely available programs (such as MATLAB, MATHEMATICA etc.) that also provide graphical representations of ‘Greeks’ variations.

### 6.4.2 Option bounds and put–call parity

#### Bounds

An option is a right, not an obligation to buy or sell an asset at a predetermined (strike price) and at a given period in the future (maturity). A forward contract differs from the option in that it is an obligation and not a right to buy or sell. Hence, an option is inherently more valuable than a forward or futures contract for it can never lead to a loss at maturity. Explicitly, the value of the forward contract is the discounted payoff at maturity  $F_T - F_0$  for a long futures contract and  $F_0 - F_T$  for a short. The predetermined futures price  $F_0$  is the strike price  $K$  in option terminology. Hence, the option price must obey the following inequalities that provide *lower bounds* on the option’s call and put values (where we replaced the forward’s price by its value derived previously):

$$c_E \geq e^{-R_f(T-t)}(S e^{R_f(T-t)} - F_0) = S - K e^{-R_f(T-t)}$$

$$p_E \geq e^{-R_f(T-t)}(F_0 - S e^{R_f(T-t)}) = K e^{-R_f(T-t)} - S$$

where  $c_E$  and  $p_E$  are the call and the put of a European option. Further, at the limit:

$$\lim_{t \rightarrow T} c_E = S - K$$

$$\lim_{t \rightarrow T} p_E = K - S$$

Similarly, we can construct option bounds on American options. Since these options have the additional right to exercise the option in the course of its lifetime, option writers are likely to ask for an additional premium to cover the additional risk transfer from the option buyer. Thus, as long as time is valuable to the investor the following bounds must also hold. Explicitly, let the price of an American and a European call option be  $C$  and  $c$  respectively while for put options we have also  $P$  and  $p$ . Then, for a non-paying dividend option it can be verified that (based on the equivalence of cash flows of two portfolios using European and American put and call options):

$$C = c, P > p \quad \text{when} \quad R_f > 0$$

**Put-call parity**

The put-call parity relationship establishes a relationship between  $p$  and  $c$ . It can be derived by a simple arbitrage between two equivalent portfolios, yielding the same payoff regardless of the stock price. As a result, their value must be the same. To do this, we construct the following two portfolios at time  $t$ :

$$\left[ \begin{array}{ccc} & \text{Time } t & \text{Time } T \\ & S_T < K & S_T > K \\ c + K e^{-R_f(T-t)} & K & (S_T - K) + K = S_T \\ p + S_t & K = (K - S_T) + S_T & S_T \end{array} \right]$$

We see that at time  $T$ , the two portfolios yield the same payoff  $\text{Max}(S_T, X)$  which implies the same price at time  $t$ . Thus:

$$c + K e^{-R_f(T-t)} = p + S_t$$

If this is not the case, then there would be some arbitrage opportunity. In this sense, computing European options prices is simplified since, knowing one leads necessarily to knowing the other.

When we consider dividend-paying options, the put-call Parity relationships are slightly altered. Let  $D$  denote the present value of the dividend payments during the lifetime of the option (occurring at the time of its ex-dividend date), then:

$$\begin{aligned} c &> S - D - K e^{-R_f(T-t)} \\ p &> D + K e^{-R_f(T-t)} - S \end{aligned}$$

Similarly, for put-call parity in a dividend-paying option, we have the following bounds:

$$S - D - K < C - P < S - K e^{-R_f(T-t)}$$

**Upper bounds**

An option's upper bound can be derived intuitively by considering the payoff irrespective of the options being American or European. The largest payoff for a put option  $\text{Max}[K - S, 0]$  occurs when the stock price is null. The put option upper bound is thus,

$$p < P < K$$

For a European call option, a similar argument leads to the conclusion that the call price must be below the price of the stock at maturity. This is irrelevant to a trader who cannot predict the stock price. However, for an American option,  $\text{Max}[S - K, 0]$ , the largest payoff occurs when the strike price is set to zero and therefore, the American call upper bound is the stock price,

$$c < C < S$$

These relationships can be obtained also by using arbitrage arguments (see, for example, Merton 1973). Finally other bounds on options are considered in Chapter 8.

### 6.4.3 American put options

American options, unlike European options may be exercised prior to the expiration date. To value such options, we can proceed intuitively by noting that the valuation is defined in terms of exercise and continuation regions over the stock price. In a continuation region, the value of the option is larger than the value of its exercise and, therefore, it is optimal to wait. In the exercise region, it is optimal to exercise the option and cash in the profits. If the time to the option's expiration date is  $t$ , then the exercise of the option provides a profit  $K - S(t)$ . In this latter case, the exercise time is a 'stopping time', and the problem is terminated. Another way to express such a statement is:

$$f(S, t) = \text{Max}[K - S(t), e^{-R_f dt} E f(S + dS, t + dt)]$$

where  $f(S, t)$  is the option price at time  $t$  when the underlying stock price is  $S$  and one or the other of the two alternatives hold at equality. At the contracted strike time of the option, we have necessarily,  $f(S, 0) = K - S(0)$ . The solution of the option's exercise time is difficult, however, and has generated a large number of studies seeking to solve the problem analytically or numerically. Noting that the solution is of the barrier type, meaning that there is some barrier  $X^*(t)$  that separates the exercise and continuation regions, we have:

$$\begin{cases} \text{If } K - S(t) \geq X^*(t) & \text{exercise region: stopping time} \\ K - S(t) < X^*(t) & \text{continuation region} \end{cases}$$

The solution of the American put problem consists then in selecting the optimal exercise barrier (Bensoussan, 1982, 1985). A number of studies have attempted to do so, including Broadie and Detemple (1996), Carr *et al.* (1992) and Huang *et al.* (1996) as well as many other authors. Although the analytical solution of American put options is hard to achieve, we shall consider here some very simple and analytical problems. For most practical problems, numerical and simulation techniques are used.

#### **Example: An American put option and dynamic programming\***

American options, unlike European ones, provide the holder of the option with the option to exercise it whenever he may wish to do so within the relevant option's lifetime. For American call options the call price of the European equals the call price of the American. This is not the case for put options, however. Assume that an American put option derived from this stock is exercised at time  $\tau < T$  where  $T$  is the option exercise period while the option exercise price is  $K$ . Let the underlying stock price be:

$$\frac{dS(t)}{S(t)} = R_f dt + \sigma dW(t), S(0) = S_0$$

Under risk-neutral pricing, the value of the option equals the discounted value (at the risk-free rate) at the optimal exercise time  $\tau^* < T$ , namely:

$$J(S, T) = \text{Max}_{\tau \leq T} E_S e^{-R_f \tau} (K - S(\tau), 0)$$

Thus,

$$J(S, t) = \begin{cases} K - S(t) & \text{exercise region: stopping time} \\ e^{-R_f dt} E J(S + dS, t - dt) & \text{continuation region} \end{cases}$$

In the continuation region we have explicitly:

$$\begin{aligned} J(S, t) &= e^{-R_f dt} E J(S + dS, t + dt) \\ &= (1 - R_f dt) E \left( J(S, t) + \frac{\partial J}{\partial t} dt + \frac{\partial J}{\partial S} dS + \frac{1}{2} \frac{\partial^2 J}{\partial S^2} (dS)^2 \right) \end{aligned}$$

which is reduced to the following partial differential equation:

$$-\frac{\partial J}{\partial t} = -R_f J(S, t) + \frac{\partial J}{\partial S} R_f S + \frac{1}{2} \frac{\partial^2 J}{\partial S^2} \sigma^2 S^2$$

While in the exercise region:

$$J(S, t) = K - S(t)$$

For a perpetual option, note that the option price is not a function of time but of price only and therefore  $\partial J / \partial t = 0$  and the option price is:

$$0 = -R_f J(S) + \frac{dJ}{dS} R_f S + \frac{1}{2} \frac{d^2 J}{dS^2} \sigma^2 S^2$$

Here the partial differential equation is reduced to an ordinary differential equation of the second order. Assume that an interior solution exists, meaning that the option is exercised if its (optimal) exercise price is  $S^*$ . In this case, the option is exercised as soon as  $S(t) \leq S^*$ ,  $S^* \leq K$ . These specify the two boundary conditions required to solve our equation.

- In the exercise region:  $J(S^*) = K - S^*$
- For optimal exercise price

$$\frac{dJ(S)}{dS} \Big|_{S=S^*} = -1$$

Let the solution be of the type  $J(S) = qS^{-\lambda}$ . This reduces the differential equation to an equation we solve for  $\lambda$ :

$$\sigma^2 \frac{\lambda(\lambda + 1)}{2} - \lambda R_f - R_f = 0 \quad \text{and} \quad \lambda^* = \frac{2R_f}{\sigma^2}$$

At the exercise boundary  $S^*$ , however:

$$J(S^*) = qS^{*-\lambda^*} = K - S^*; \frac{dJ(S^*)}{dS^*} = -\lambda^* qS^{*-\lambda^*-1} = -1$$

These two equations are solved for  $q$  and  $S^*$  leading to:

$$S^* = \frac{\lambda^*}{1 + \lambda^*} K \quad \text{and} \quad q = \frac{(\lambda^*)^{\lambda^*} K^{1+\lambda^*}}{(1 + \lambda^*)^{1+\lambda^*}}$$

And the option price is:

$$J(S) = \left[ \frac{(\lambda^*)^{\lambda^*} K^{1+\lambda^*}}{(1 + \lambda^*)^{1+\lambda^*}} \right] S^{-\lambda^*}, \quad \lambda^* = \frac{2R_f}{\sigma^2}, \quad S^* = \frac{\lambda^*}{1 + \lambda^*} K$$

In other words, the solution of the American put will be:

$$\begin{cases} \text{sell if} & S \leq S^* \\ \text{hold if} & S > S^* \end{cases}$$

When the option time is finite, say  $T$ , the condition for optimality is reduced to one of the two equations equating zero:

$$0 = \begin{cases} J(S, t) - (K - S(t)) \\ -\frac{\partial J}{\partial t} + R_f J(S, t) + \frac{\partial J}{\partial S} R_f S + \frac{1}{2} \frac{\partial^2 J}{\partial S^2} \sigma^2 S^2 \end{cases}$$

This problem is much more difficult to solve, however. Below, we consider a paper that has in fact been solved analytically.

**Example\*: A solved case (Kim and Yu, 1993)**

Let the underlying price process be a lognormal process:

$$\frac{dS}{S} = \mu dt + \sigma dw, \quad S(0) = S_0$$

As long as the option is kept, its price evolves following the (Black–Scholes) partial differential equation:

$$\frac{\partial f}{\partial t} + \mu S \frac{\partial f}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2} - R_f f = 0$$

In addition, we have the following boundaries:

$$f(S_T, T) = \text{Max} [0, K - S_T]$$

$$\lim_{S_t \rightarrow \infty} f(S_t, t) = 0$$

$$\lim_{S_t \rightarrow S_t^*} f(S_t, t) = K - S_t^*$$

The first boundary condition assumes that the option is exercised at its expiration date  $T$ , the second assumes that the value of the option is null if the stock price is infinite (in which case, it will never pay to sell the option) and finally, the third boundary condition measures the option's payoff at its exercise at time  $t$ . Let  $S_t^*$  be the optimal exercise price at time  $t$ , when the option is exercised prior to maturity, in which case (assuming that  $f(S_t, t)$  admits first and second derivatives), we have:

$$\lim_{S_t \rightarrow S_t^*} \frac{\partial f(S_t, t)}{\partial S_t} = -1$$

Although the solution of such a problem is quite difficult, Carr *et al.* (1992) and Kim and Yu (1993) have shown that the solution can be written as the sum of the option price for the European part of the option plus another sum which accounts for the premium that the American option provides. This expression is explicitly given by:

$$P_0 = P(S_0, 0) = p_0 + \pi$$

$$\pi = R_f K \int_0^T e^{-R_f t} \int_0^{S_t^*} \Psi(S_t, S_0) dS_t dt - (R_f - \mu) \int_0^T e^{-R_f t} \int_0^{S_t^*} S_t \Psi(S_t, S_0) dS_t dt$$

where  $p_0$  is the option price of a European put, the ‘flexibility premium’ associated with the American option is  $\pi$ , while  $\Psi(S_t, S_0)$  is the transition probability density function to a price  $S_t$  at time  $t$  from a price  $S_0$  at  $t = 0$ .

The analytical, as well as the numerical, solution of these problems is of course cumbersome. In the next chapter we shall consider a similar class of problems that seek to resolve simple problems of the type ‘when to sell, when to buy, should we hold’ both assets and options.

## REFERENCES AND ADDITIONAL READING

- Back, K. (1993) Asymmetric information and options, *Review of Financial Studies*, **6**, 435–472.
- Beibel, M., and H.R. Lerche (1997) A new look at optimal stopping problems related to mathematical finance, *Statistica Sinica*, **7**, 93–108.
- Bensoussan, A. (1982) *Stochastic Control by Functional Analytic Methods*, North Holland, Amsterdam.
- Bensoussan, A. (1984) On the Theory of Option Pricing, *ACTA Applicandae Mathematicae*, **2**, 139–158.
- Bensoussan, A., and H. Julien (2000) On the pricing of contingent claims with friction, *Mathematical Finance*, **10**, 89–108.
- Bergman, Yaacov A. (1985) Time preference and capital asset pricing models, *Journal of Financial Economics*, **14**, 145–159.
- Black, F., and M. Scholes (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.
- Boyle, P. P. (1992) *Options and the Management of Financial Risk*, Society of Actuaries, New York.
- Brennan, M.J. (1979) The pricing of contingent claims in discrete time models, *The Journal of Finance*, **1**, 53–63.
- Brennan, M.J., and E.S. Schwartz (1979) A Continuous Time Approach to the Pricing of Corporate Bonds, *Journal of Banking and Finance*, **3**, 133–155.
- Brennan, M.J., and E.S. Schwartz (1989) Portfolio insurance and financial market equilibrium, *Journal of Business*, **62**(4), 455–472.
- Briys, E., M. Crouhy and H. Schlesinger (1990) Optimal hedging under intertemporally dependent preferences, *The Journal of Finance*, **45**(4), 1315–1324.
- Broadie, M., and J. Detemple (1996) American options valuation, new bounds, approximations and a comparison of existing methods, *Review of Financial Studies*, **9**, 1211–1250.
- Brown, R.H., and S.M. Schaefer (1994) The term structure of real interest rates and the Cox, Ingersoll and Ross model, *Journal of Financial Economics*, **35**(1), 3–42.
- Carr, P., R. Jarrow and R. Myneni (1992) Alternative characterizations of American Put options, *Mathematical Finance*, **2**, 87–106.

- Cox, J.C., J.E. Ingersoll Jr and S. A. Ross (1981) The relation between forward prices and futures prices, *Journal of Financial Economics*, **9**(4), 321–346.
- Cox, J.C., and S.A. Ross (1976) The valuation of options for alternative stochastic processes, *Journal of Financial Economics*, **3**, 145–166.
- Cox, J.C., and S.A. Ross (1978) A survey of some new results in financial option pricing theory, *Journal of Finance*, **31**, 383–402.
- Cox, J.C., S.A. Ross and M. Rubenstein (1979) Option pricing approach, *Journal of Financial Economics*, **7**, 229–263.
- Cox, J., and M. Rubinstein (1985) *Options Markets*, Prentice Hall, Englewood Cliffs, NJ.
- Davis, M.H.A., V.G. Panas and T. Zariphopoulou (1993) European option pricing with transaction costs, *SIAM Journal on Control and Optimization*, **31**, 470–493.
- Duffie, D. (1988) *Security Markets: Stochastic Models*, Academic Press, New York.
- Duffie, D. (1992) *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, N. J.
- Geman, H., and M. Yor (1993) Bessel processes, Asian options and perpetuities, *Mathematical Finance*, **3**(4), 349–375.
- Geske, R., and K. Shastri (1985) Valuation by approximation: A comparison of alternative option valuation techniques, *Journal of Financial and Quantitative Analysis*, **20**, 45–71.
- Grabbe, J. O. (1991) *International Financial Markets* (2nd edn), Elsevier, New York.
- Harrison, J.M., and D.M. Kreps (1979) Martingales and arbitrage in multiperiod security markets, *Journal of Economic Theory*, vol. 20, no. 3, 381–408.
- Harrison, J.M., and S.R. Pliska (1981) Martingales and stochastic integrals with theory of continuous trading, *Stochastic Processes and Applications*, **11**, 261–271.
- Haug, E.G. (1997) *The Complete Guide to Option Pricing Formulas*, McGraw-Hill, New York.
- Henry, C. (1974) Investment decisions under uncertainty: The irreversibility effect, *American Economic Review*, **64**, 1006–1012.
- Huang, C.F., and R. Litzenberger (1988) *Foundations for Financial Economics*, North Holland, Amsterdam.
- Huang, J., M.G. Subrahmanyam and G. G. Yu (1996) Pricing and hedging American options, *Review of Financial Studies*, **9**(3), pp. 277–300.
- Hull, J. (1993) *Options, Futures and Other Derivatives Securities* (2nd edn), Prentice Hall, Englewood Cliffs, NJ.
- Jacka, S.D. (1991) Optimal stopping and the American Put, *Journal of Mathematical Finance*, **1**, 1–14.
- Jarrow, R.A. (1988) *Finance Theory*, Prentice Hall, Englewood Cliffs, NJ.
- Karatzas, I., and S.E. Shreve (1998) *Methods of Mathematical Finance*, Springer, New York.
- Kim, I.J., and G. Yu (1990) A simplified approach to the valuation of American options and its application, New York University, Working paper.
- Kim, I.J. (1993) The analytic valuation of American options, *Review of Financial Studies*, **3**, 547–572.
- Leroy, Stephen F. (1982) Expectation models of asset prices: A survey of theory, *Journal of Finance*, **37**, 185–217.
- McKean, H.P. (1965) A free boundary problem for the heat equation arising from a problem in mathematical economics, *Industrial Management Review*, **6**, 32–39.
- Merton, R. (1969) Lifetime portfolio selection under uncertainty: The continuous time case, *Review of Economics and Statistics*, **50**, 247–257.
- Merton, R.C. (1973) Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4**, 141–183.
- Merton, R.C. (1977) Optimum consumption and portfolio rules in a continuous time model, *Journal of Economic Theory*, **3**, 373–413.
- Merton, R.C. (1992) *Continuous Time Finance*, Blackwell, Cambridge, MA.
- Pliska, S.R. (1975) Controlled jump processes, *Stochastic Processes and Applications*, **3**, 25, 282.
- Ross, S.A. (1976) Options and efficiency, *Quarterly Journal of Economics*, **90**.
- Ross, S.A. (1976) The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, December, **13**(3), 341–360.

- Smith, C.W. (1976) Option pricing: A review, *Journal of Financial Economics*, **3**, 3–51.
- Stoll, Hans, R. (1969) The relationship between put and call option prices, *Journal of Finance*, **24**, 802–824.
- Wilmott, P. (2000) *Paul Wilmott on Quantitative Finance*, John Wiley & Sons, Ltd, Chichester.
- Wilmott, P., J. Dewynne and S.D. Howison (1993) *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press, Oxford.

## CHAPTER 7

# Options and Practice

### 7.1 INTRODUCTION

Option writers, are entrepreneurs in search of profits. As in any fight, fairness is not rewarded. In this spirit, option writers and their financial engineers seek to avoid fair competition by differentiating their products and fitting them to their clients' specific needs or responding to demands of new or seasoned hedgers and speculators. 'The best fight is the one that we cannot lose'! Profits may thus be realized when option writers create a market niche where competition is conspicuously lacking and where there may be some arbitrage profits. Of course, fees have to be set as a function of the writer's power which will depend on the risk of losing important clients, competition from other writers of the same and other products as well as the sophistication of large institutions with their own trading centres. Option writers, as other marketers in other areas, attempt to innovate by creating new products (or variants to currently marketed products), which is in fact a service of intangible characteristics catering to the attitude of investors, firms and individuals to uncertainty. The majority of investors, in fact, abhor uncertainty, while only few seek it or are willing to take positions that the majority will refuse. These participants in financial markets are 'human entities' and market gladiators are prospecting by providing services and trades that are sensitive to their 'psychological and economic' needs and profiles. For risky contracts (in times of crashes for example or very high volatility) speculators will be needed to provide liquidity. Hence, it is not surprising that complete markets and risk-neutral pricing breaks down when this is the case. When the supply of risk is overbearing and there may not be enough 'speculators' to assume it, markets will, at least, become incomplete. Market gladiators are neither risk-seeking nor pure hedgers, however. Management of conservative investment funds such as retirement funds also involve risk. Bonds, assumed generally safe investments, are also risky for they may default or, at least, their value may fluctuate as a function of interest rates, inflation and other economic variables. By the same token, as we have seen in Chapter 5, some hedge fund managers may share information regarding disparities between economic policies and economic fundamentals to generate a herd effect, or a potential run on a currency or an economic entity – bringing them back to alignment with a 'natural economic equilibrium'. Fortunes

are made and lost on these ‘casino runs’ where money is made in an instant and lost in another.

Niche-seeking and product innovation responding to speculative and hedging needs are not the only tools available to market gladiators. A continuous concern for market participation, a concern to avoid regulatory interventions and the urge to avoid tax payments in a legally defensible manner, underpin another source of product innovation. An outstanding example is the creation of Eurodollars deposits of a domestic currency in a foreign country – just as currency swaps were started in the 1980s by the World Bank and have been used since then by CFOs of international firms and banks. Similarly, the concept of offshore funds was conceived to avoid tax payments. This fact underscores current government regulation seeking to limit the use of these funds.

For these reasons, option writers have produced as many tailored options as business imagination can construct. They can be used individually as well as in a combined manner. Financial engineers create and price the cost of products but it is only the market that prices these products. The more ‘tailored’ the products the less price-efficient the market is, compared to standard widely traded products.

Although most people believe that derivatives are a recent innovation, they date as far back as twelfth-century practices by Flemish traders. The first futures and options contracts resembling current option types were in fact implemented in the seventeenth century in Amsterdam, which was at that time the financial capital of the Western world, and in the rice market of Osaka. Practice in derivatives has truly expanded into global financial markets since mathematical finance and economic theory has made it possible to value such derivatives contracts. The result is an expansion of trades for both ‘present and futures trades’ are traded at the same time, providing broad flexibilities to financial managers and investors to select the time-risk profile substitutions they prefer.

Options products may be grouped in several categories, summarized by the following:

- Packaged options are usually expressed and valued in terms of plain vanilla options, combining them to generate desired risk properties and profiles. Options strategies such as covered call; protective put; bull and bear spread; calendar spread, butterfly spread, condors, laps and flex, warrants, and others, are such cases we shall consider in this chapter.
- Compound options are derived options based on exercise prices that may be uncertain (for example, warrants, stock options, options on corporate bonds etc.). In this case, it is a ‘derived asset twice’ – first on the underlying asset and then on some other variable on the basis of which the option is constructed.
- Forward starts are options with different states, awarding thereby the right to exercise the option at several times in the future.
- Path-dependent options depend on the price and the trajectory of other variables. Asian options, knock-out options and many other option types are of this kind, as we shall see in this chapter.

- Multiple assets options involve options on several and often correlated risky assets (such as quantos, exchange options etc.).

In addition, there are options in application areas such as currency options, commodity options, and options on futures, as well as climatic options that assume an increasingly important role in both insurance and energy-related contracts. Warrants are used by firms as call options on the firm's equity. When the warrant is exercised, firms usually issue new stock, thereby diluting current stockholders' equity holdings. There are in addition numerous contracts such as swaps, caps and floors, swaptions and captions etc. that we shall also elaborate on in this chapter. The number of options used in practice is therefore very large and this precludes a complete coverage. For this reason, we shall consider a few such options as examples, providing an opening for both the theoretically and applications minded investor and financial manager. Further study will be needed, however, to appreciate the mathematical intricacies and limitations of dealing with these problems and to augment the sensitivity to the economic rationale such options presume when they are used in practice and are valued by the available quantitative tools.

## 7.2 PACKAGED OPTIONS

Packaged options are varied. We consider first binary options. A payoff for binary options occurs if the value of the underlying asset  $S(T)$  at maturity  $T$  is greater than a given strike price  $K$ . The amount paid may be constant or a function of the difference  $S(T) - K$ . The price of these options can be calculated easily if risk-neutral pricing is applicable (since, it equals the discounted value of the terminal payoff). When computations are cumbersome, it is still possible to apply standard (Monte Carlo) simulation techniques and calculate the expected discounted payoff (assuming again risk-neutral pricing, for otherwise simulation would be misleading). The variety of options that pay nothing or 'something' is large and therefore we can briefly summarize a few:

- *Cash or nothing*: Pays  $A$  if  $S(T) > K$ .
- *Asset or nothing*: Pays  $S(T)$  if  $S(T) \geq K$ .
- *Gap*: Pays  $S(T) - K$  if  $S(T) \geq K$ .
- *Supershare*: Pays  $S(T)$  if  $K_L \leq S(T) \leq K_H$ .
- *Switch*: Pays a fixed amount for every day in  $[0, T]$  that the stock trades above a given level  $K$ .
- *Corridor (or range notes)*: Pays a fixed amount for every day in  $[0, T]$  that the stock trades above a level  $K$  and below a level  $L$ .
- *Lookback options*: Floating-strike lookback options that provide a payout based on a lookback period (say three months), equalling the difference between the largest value and the current price. There are Min and Max lookback options:

$$\text{Min: } V(T) = \text{Max}[0, S(T) - S_{\min}]; \text{ Max: } V(T) = \text{Max}[0, S_{\max} - S(T)]$$

- *Asian options:* Asian options are calculated by replacing the strike price by the average stock price in the period. Let the average price be:

$$\bar{S} = \frac{1}{T} \int_0^T S(t) dt; t \in [0, T]$$

Then the value of the call and put of an Asian option is simply:

$$\text{Put: } V(T) = \text{Max} [0, \bar{S} - S(T)]; \text{Call: } V(T) = \text{Max}[0, S(T) - \bar{S}]$$

- *Exchange:* A multi-asset option that provides the option for a juxtaposition of two assets ( $S_1, S_2$ ) and given by  $\text{Max} [S_2(T) - S_1(T), 0]$ . Such options can also be used to construct options on the maximum or minimum of two assets. For example, buying the option to exchange one currency ( $S_1$ ) with another ( $S_2$ ) leads to:

$$V(T) = \min [S_1(T), S_2(T)] = S_2 - \text{Max} [S_2(T) - S_1(T), 0]$$

$$V(T) = \max [S_1(T), S_2(T)] = S_1(T) + \text{Max} [S_2(T) - S_1(T), 0]$$

- *Chooser:* Provides the option to buy either a call or a put. Explicitly, say that ( $T_1, T_2$ ) are the maturity dates of call and put options with strikes ( $K_1, K_2$ ). Now assume that an option is bought on either of the options with strike  $T \leq (T_1, T_2)$ . The payoff at maturity  $T$  is then equal to the max of a call  $C[S(T), T_1 - T; K_1]$  and the put  $P[S(T), T_2 - T; K_2]$ :

$$\text{Max}\{C[S(T), T_1 - T; K_1], P[S(T), T_2 - T; K_2]\}$$

- *Barrier and other options:* Barrier options have a payoff contingent on the underlying assets reaching some specified level before expiry. These options have knock-in features (namely in barrier) as well as knock-out features (out-barrier). These options are solved in a manner similar to the Black–Scholes equation considered in the previous chapter, except for a specification of boundary conditions at the barriers. We can also consider barrier options with exotic and other features such as options on options, calls on puts, calls on calls, puts on calls etc., as well as calls on forwards and vice versa. These are compound options and are written using both the maturity dates and strike prices for both the assets involved. For example, consider a call option with maturity date and strike price given by ( $T_1, K_1$ ). In this case, the payoff of a call on a call with maturity date  $T$  and strike  $K$  is a compound option given by:

$$C_c(T_1, K_1) = \text{Max}\{0, C[S(T), T_1 - T, K_1] - K_1\}$$

where  $C[S(T), T_1 - T, K_1]$  is the value at time  $T$  of a European call option with maturity  $T_1 - T$  and strike price  $K_1$ . By the same token, a compound put option on a call pays at maturity:

$$P_c(T_1, K_1) = \text{Max} (0, K - C(S(T), T_1 - T, K_1))$$

Practically, the valuation of such options is straightforward under risk-neutral pricing since their value equals their present discounted terminal payoff (at the exercise time).

- *Passport options*: These are options that make it possible for the investor to engage in short/long (sell/buy) trading of his own choice while the option writer has the obligation to cover all net losses. For example, if the buyer of the option takes positions at times  $t_i, i = 1, \dots, n - 1, t_0 = 0, t_n = T$  by buying or selling European calls on the stock, then the passport option provides the following payoff at time  $T$  – the option exercise time:

$$\text{Max} \left( \sum_{i=0}^{n-1} u_i [S(t_{i+1}) - S(t_i)], 0 \right)$$

where  $u_i$  is the number of shares (if bought, it is positive; if sold, it is negative) at time  $t_i$  and resolved at period  $t_{i+1}$ . In this case, the period profit or loss would be:  $[S(t_{i+1}) - S(t_i)]$ . Particular characteristic can be added such as the choice of the asset to trade, the number of trades allowed etc.

- *As you like it options*: These options allow the investor to chose after a specified period of time  $T$ , whether the option is a call or a put. If the option is European and the call and the put have the same strike price  $K$ , then put-call parity can be used. The value at exercise is  $\text{Max}(c, p)$  and consists in selecting either the call or the put at the time the option exercise is made. Thus, put-call parity with continuous and compounded discounting and a dividend-paying stock at a rate of  $q$  implies (as we shall see later on):

$$c + e^{-q(T-t)} \text{Max}[0, K e^{-(R_f - q)(T-t)} - S(t)]$$

In other words, ‘as you like it options’ consist of a call option with strike  $K$  at  $T$  and  $e^{-q(T-t)}$  put options with a strike of  $[K e^{-(R_f - q)(T-t)}]$  at maturity  $T$ .

The finance trade and academic literature abounds with options that are tailored to clients’ needs and to the market potential for such options. Therefore, we shall consider a mere few while the motivated reader should consult the numerous references at the end of the previous and the current chapter for further study and references to specific option types.

### 7.3 COMPOUND OPTIONS AND STOCK OPTIONS

Stocks are assets that represent equity shares issued by individual firms. They have various forms, granting various powers to stockholders. In general, stockholders are entitled to dividend payments made by the firm and to the right to vote at the firm’s assembly. Stocks are also a claim to the value of the firm that they share with bondholders. For example, if the firm defaults on its interest payments, bondholders can force the firm into bankruptcy to recover the loans. A stockholder, a junior claimant in this case, has generally nothing left to claim. Hence, a bondholder has the right to sell the company at a given threshold or,

equivalently, the bondholder holds a put on the value of the firm that the stockholder must hold short. Hence, a stock can be viewed as a claim or option on the value of the firm that is shared with bondholders. In practice, managers are often given stock options on their firm so they may align their welfare with those of the shareholders. The rationale of such compensation is that a manager whose income is heavily dependent on an upward move of the firm's stock price will be more likely to pursue an aggressive policy leading to a stock price rise as his payoff is a convex increasing function of the stock price. The shareholders will, of course, benefit from such a rise while it assumes some risk due to the call (stock) option's limited liability granted to the manager. This case illustrates some of the economic limits of risk-neutral pricing, which presumes that risk can be eliminated by trading it away. Further, this supposes the existence of another party willing to take the risk for no extra compensation. This can happen only if markets are perfectly liquid or there exists another investor willing to take on the exact opposite risk. Risk-neutrality presupposes therefore that there is always such an exact opposite. In reality, as is the case for executives' options, the strategy is set up so that the risk is not shifted away. For most applications, risk-neutrality may be used comfortably. But, the more out of the money options are, the less risk can be transferred and, thus, the more speculators are needed to take this risk. This means that in crash times or other extreme events, risk-neutral pricing tends to break down.

With these limitations in mind, we can apply risk-neutral pricing to value options or compound options (options on a stock option or some other underlying asset). Define a stock option (a claim) on the value of the firm (its stock price). To do so, say that a firm has  $N$  shares whose price is  $S$  and let the firm's debt be expressed by a pure discount bond  $B$  with maturity  $T$ . Initially, the value of the firm  $V$  can be written as  $V = NS + B$ . Assuming risk-neutral pricing, the stockprice (using an annual risk-free discount rate) over one and two periods is:

$$S = \frac{1}{(1 + R_f)} E^* \tilde{S}(1) = \frac{1}{(1 + R_f)^2} E^* \tilde{S}(2)$$

For a binomial process, shown in Figure 7.1, we have:  $\tilde{S}(1) = (S_h, S_d)$  and  $\tilde{S}(2) = (S_{hh}, S_{hd}S_{dd})$ . By the same token, we compute recursively the value of the compound (stock) option by:

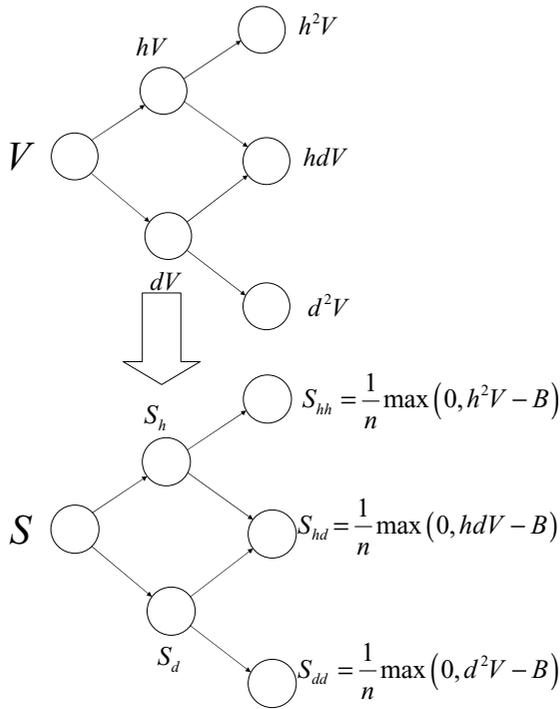
$$C^c = \frac{1}{(1 + R_f)} E^* \tilde{C}^c(1) = \frac{1}{(1 + R_f)^2} E^* \tilde{C}^c(2)$$

with  $\tilde{C}^c(1) = (\tilde{C}_h^c, \tilde{C}_d^c)$ ,  $\tilde{C}^c(2) = (\tilde{C}_{hh}^c, \tilde{C}_{hd}^c, \tilde{C}_{dd}^c)$

$$C_{hh}^c = \text{Max} [0, h^2 V - B]; C_{dd}^c = \text{Max} [0, d^2 V - B]; C_{hd}^c = \text{Max} [0, h d V - B]$$

and therefore:

$$C^c = \left( \frac{1}{1 + R_f} \right)^2 \left[ \begin{array}{l} p^{*2}(\text{Max}[0, h^2 V - B]) + \\ 2(1 - p^*)p^*(\text{Max} [0, h d V - B]) + \\ (1 - p^*)^2 \text{Max}(0, d^2 V - B) \end{array} \right]$$



**Figure 7.1** Compound option.

Here the risk neutral probability is:

$$p^* = \frac{1 + R_f - d}{h - d}$$

Note that this model differs from the simple plain vanilla model treated earlier, since in this case,  $S_h \neq hS$ ;  $S_d \neq dS$ . When the firm has no debt, the firm value is  $V = NP$ , a portion is invested in a risk-free asset and the other in a risky asset, similarly to the previous binomial case. For example, say that  $u = 1.3$  while  $d = 0.8$  and the risk free rate is  $R_f = 0.1$ . Thus the risk-neutral probability is:

$$p^* = \frac{1 + 0.1 - 0.8}{1.3 - 0.8} = \frac{0.3}{0.5} = 0.6,$$

$$C^c = \left(\frac{1}{1.1}\right)^2 \begin{bmatrix} 0.36 [\text{Max}(0, 1.69V - B)] + \\ 0.48 [\text{Max}(0, 0.8V - B)] + \\ 0.16 [\text{Max}(0, 0.64V - B)] \end{bmatrix}$$

Now, if bondholders have a claim on 40 % of the firm value, we have:

$$C^c = V \left(\frac{1}{1.1}\right)^2 [0.36(1.29) + 0.48(0.4) + 0.16(0.24)] = (0.57421)V$$

**Problem**

What are the effects of an increase of 5% on bondholders' share of the firm on the option's price?

**Problem**

High-tech firms (and in particular start-ups) often offer their employees stock options instead of salary increases. When is it better to 'take the money' over the options and vice versa. Construct a model to justify your case.

**7.3.1 Warrants**

Warrants are compound options, used by corporations that issue call options with their stock as an underlying asset. When the option is exercised, new stock is issued, diluting other stockholders' holdings but adding capital to the corporation. A warrant is valued as follows. Say that  $V$  is the firm's value and let there be  $n$  warrants, providing the right to buy one share of stock at a price of  $x$  and assume no other source of financing. If all warrants are exercised, then the new value of the firm is  $V + mx$  and thus, each warrant must at least be worth its price  $x$ , or:

$$\frac{V + mx}{N + m} > x$$

This means that a warrant is exercised only if:

$$V + mx > (N + m)x \quad \text{and} \quad V > Nx \quad \text{or} \quad x < V/N = S$$

If the value at time  $t$  is:  $W(V, \tau)$ ,  $\tau = T - t$  or at time  $t = 0$ ,

$$W(V, 0) = \begin{cases} \frac{V + mx}{N + m} - x & V > Nx \\ 0 & V \leq Nx \end{cases}$$

Then, if we set:  $\lambda = 1/[N + m]$  we have  $\lambda(V + mx) - x = \lambda V - x(1 - m\lambda)$  and thereby the price  $W(V, 0)$  can be written as follows:

$$W(V, 0) = \text{Max} [\lambda V - x(1 - m\lambda), 0] = \text{Max} \left[ V - \frac{x(1 - m\lambda)}{\lambda}, 0 \right]$$

This corresponds to an option whose price is  $V$ , the value of the firm, and whose strike (in a Black–Scholes model) is  $[x(1 - m\lambda)]/\lambda$ , thus, applying the Black–Scholes option pricing formula, we have at any one time:

$$W(V, \tau, \lambda, x) = \lambda W \left( V, \tau, \frac{x(1 - m\lambda)}{\lambda} \right) = W(\lambda V, \tau, x(1 - m\lambda))$$

Therefore, it is possible to value a warrant using the Black–Scholes option formula. For example, say that there are  $m = 500$  warrants with a strike price  $x = 100$ , a time to maturity of  $\tau = 0.25$  years, the yearly risk-free interest rate is  $R_f = 10\%$ , the stock price volatility is  $\sigma = 20\%$  a year and let there be  $N = 100\,000$  shares on the market while the current firm value is 150 000. Then,

the warrant's price is calculated by:

$$\begin{aligned} W[\lambda V, \tau, x(1 - m\lambda)] &= W\{1.5(10^5)\lambda, 0.25, 100[1 - (500)\lambda]\} \\ &= W(14.285, 0.25, 95.23) \end{aligned}$$

where  $\lambda = 1/(10\,500) = 0.095\,238$ .

In a similar manner, other compound options such as *options on a call* (call on call, put on call) and *options on put* (put on put, put on call) etc. may be valued.

### 7.3.2 Other options

We consider next and briefly a number of other options in a continuous-time framework. Throughout, we assume that the underlying process is a lognormal process.

*Options on dividend-paying stocks* are options on stocks that pay dividends at a rate of  $D$  proportional to the stock price. Note that the underlying price process with dividends is then:

$$\frac{dS}{S} = (\mu - D)dt + \sigma dW$$

Thus, applying risk-neutral pricing, the partial differential equation that values the option is given by:

$$-R_f V + \frac{\partial V}{\partial t} + (R_f - D)S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = 0$$

and the boundary condition for a European call option is  $V(S, T) = \text{Max}(S(T) - K, 0)$ . If we apply a no-arbitrage argument as we have in the previous chapter, we are left with  $-D dt$  which in essence deflates the price of the stock for the option holder (since the option holder, not owning the stock, does not benefit from dividend distribution). On this basis we obtain the option price deflated by dividends.

*Options on foreign currencies* are derived in the same manner. Instead of dividends, however, it is the foreign risk-free rate  $R_{for}$  that we use. In this case, the partial differential equation is:

$$-R_f V + \frac{\partial V}{\partial t} + (R_f - R_{for})S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = 0$$

Again, by specifying the appropriate boundaries, we can estimate the value of the corresponding option.

Unlike options on dividend paying stocks, *options on commodities* involve a carrying charge of, say  $qS dt$ , which is a fraction of the value of the commodity that goes toward paying the carrying charge. As a result, the corresponding differential equation is:

$$-R_f V + \frac{\partial V}{\partial t} + (R_f + q)S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = 0$$

with an appropriate boundary condition, specified according to the type of option we consider (call, put, etc.)

*Options on futures* are defined by noting that (see also Chapter 8):

$$F = S e^{R_f(T_F-t)}$$

Thus, the value of an option on a stock and an option on its futures are inherently connected by the above relationship. However, futures differ from options on stock in that the underlying security is a futures contract. Upon exercise, the option holder obtains a position in the futures contract. If we apply Ito's differential rule to determine the value of the option on the futures, we have:

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial S} dS + \frac{1}{2} \frac{\partial^2 F}{\partial S^2} (dS)^2 \quad \text{or} \quad dF + R_f S dt = e^{R_f(T_F-t)} dS$$

which is introduced in our partial differential equation to yield:

$$-R_f V_F + \frac{\partial V_F}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V_F}{\partial S^2} = 0$$

This is solved with the appropriate boundary constraint (determined by the contract we seek to value). Although options on futures have existed in Europe for some time, they have only recently become available in America. In 1982, the Commodity Futures Trading Commission allowed each commodity exchange to trade options on one of its futures contracts. In that year eight exchanges introduced options. These contracts included gold, heating oil, sugar, T-bonds and three market indices. Options on futures now trade on every major futures exchange. The underlying spot commodities include financial assets such as bonds, Eurodollars and stock indices, foreign currencies such as British pounds and euros, precious metals such as gold and silver, livestock commodities such as hogs and cattle and agricultural commodities such as corn and soybeans.

An *option on a futures price* for say, a commodity, can be related to the spot price by:

$$F = S e^{(R_f - q)(T_F - t)}$$

For a financial asset,  $q$  is the dividend yield on the asset, whereas for a commodity (which can be consumed),  $q$  must be modified to reflect the convenience yield less the carrying charge. Now in a risk-neutral economy the expected growth rate in the price of a stock which pays continuous dividends at a rate of  $q$  is  $R_f - q$ . In such an economy, the expected growth rate of a futures price should be zero, because trading a futures contract requires no initial investment. This means, that for pricing purposes, the value of  $q$  should be  $R_f$ . That is, for pricing an option on futures, the futures prices can be treated in the same way as a security paying a continuous dividend yield rate  $R_f$ . We substitute  $G(t) = F(t) e^{-R_f(T_F-t)}$  into Merton's model, leading to the model above and whose solution for a European call option is (as established by Fisher-Black):

$$V_F(0) = e^{-R_f T_F} [F(0)N(d_1^*) - XN(d_2^*)]; \quad d_1^* = \frac{1}{\sigma \sqrt{T_F}} \left[ \ln \left( \frac{F(0)}{X} \right) + \frac{1}{2} \sigma^2 T_F \right]$$

$$\text{and } d_2^* = d_1^* - \sigma \sqrt{T_F}.$$

## 7.4 OPTIONS AND PRACTICE

Financial options and engineering is about making money, or, inversely, not losing it. To do so, pricing (valuation), forecasting, speculation and risk reduction through trading (hedging and risk trading management) are an essential activity of traders. For investors, hedgers, speculators and arbitrageurs that consider the buying of options (call, put or of any other sort), it is important to understand the many statistics that abound and are provided by financial services and firms and how to apply such knowledge to questions such as:

- When to buy and sell (how long to hold on to an option or to a financial asset). In other words, what are the limits to buy/sell the stock on the asset.
- How to combine a portfolio of stocks, assets and options of various types and dates to obtain desirable (and feasible) investment risk profiles. In other words, how to structure an investment strategy.
- What are the risks and the profit potential that complex derivative products imply (and not only the price paid for them)?
- How to manage productively derivatives and trades.
- How to use derivatives to improve the firm positioning.
- How to integrate chart-trading strategies into a framework that takes into account financial theory, and how to value these strategies and so on.

The decision to buy (a long contract) and sell (a short contract, meaning that the contract is not necessarily owned by the investor) is not only based on current risk profiles, however. Prospective or expected changes in stock prices, in volatility, in interest rates and in related economic and financial markets (and statistics) are essential ingredients applied to solve the basic questions of ‘what to do, when and where’. In practice, these questions are approached from two perspectives: the individual investor and the market valuation. The former has his own set of preferences and knowledge, while the latter results from demand and supply market forces interacting in setting up the asset’s price. Trading and trading risks result from the diverging assessments of the market and the individual investor, and from external and environmental effects inducing market imperfections. These issues will be addressed here and in the subsequent chapters as well.

In practice, options and derivative products (forward, futures, their combinations etc.) are used for a broad set of purposes spanning hedging, credit and trading risk management, incentives for employees (serving often the dual purpose of an incentive to perform and a substitute for cash outlays in the form of salaries as we saw earlier) and as an essential tool for constructing financial packages (in mergers and acquisitions for example). Derivatives are also used to manage commodity trades, foreign exchange transactions, and interest risk (in bonds, in mortgage transactions etc.). The application of these financial products spans the simple ‘buy’/‘sell’ decisions and complex trading strategies over multiple products, multiple markets and multiple periods of time. While there are many strategies, we shall focus our attention on a selected few. These strategies can be organized as follows:

- Strategies based on plain vanilla options, including:
  - strategies using call options only,
  - strategies using put options only,
  - strategies combining call and put options.
- strategies based on exotic options.
- Other speculating – buy–sell–hold–buy – strategies

These strategies are determined by constructing a portfolio combining options, futures, forwards, stocks etc. in order to obtain cash flows with prescribed and desirable risk properties. The calculation and the design of such portfolios is necessarily computation-intensive, except for some simple cases we shall use here.

### 7.4.1 Plain vanilla strategies

#### Call and put options

Plain vanilla options can be used simply and in a complex manner. A *long call* option consists in buying the option with a given exercise price and strike time specified. The portfolio implied by such a financial transaction is summarized in the table below and consists in a premium payment of  $c$  for an option, a function of  $K$  and  $T$  and the underlying process, whose payoff at time  $T$  is  $\text{Max}(S_T - K, 0)$  (see also Figure 7.2):

Time $t = 0$	Final time $T$
$c_0$	$\text{Max}(S_T - K, 0)$

When the option is bought, the payoff is a random variable, a function of the future (at the expiration date  $T$ ) market stock price  $\tilde{S}_{T|0}$ , reflecting the current information the investor has regarding the price process at time  $t = 0$  and its strike  $K$ :

$$c_0 = E[e^{-R_i(T-t)} \text{Max}(\tilde{S}_{T|0} - K, 0)]$$

In this case, note that the discount rate  $R_i$  is the one applied by the investor. If markets are complete, we have risk-neutral pricing, then of course the option (market) price equals its expectation, and therefore:

$$c_0 = e^{-R_f(T-t)} E^*[\text{Max}(\tilde{S}_{T|0} - K, 0)]$$

where  $R_f$  is the risk-free rate and  $E^*$  is an expectation taken under the risk-neutral distribution. An individual investor may think otherwise, however, and his beliefs may of course be translated into (technical) decision rules where the individual attitudes to risk and beliefs as well as private and common knowledge combine to yield a decision to buy and sell (long or short) or hold on to the asset. In general, such call options are bought when the market (and/or the volatility) is bullish or when the investor expects the market to be bullish – in other words, when we

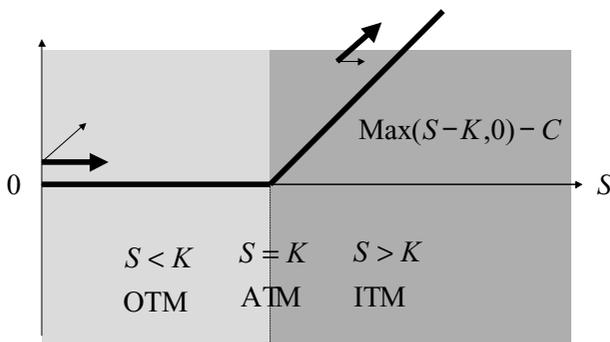
expect that the market price may be larger than the strike price at its exercise time. The advantage in buying an asset long, is that it combines a limited downside risk (limited to the option's call price) while maintaining a profit potential if the price of the underlying asset rises above the strike price. For example, by buying a European call on a share of stock whose current price is \$110 at a premium of \$5 with a strike price of \$120 in six months, we limit our risk exposure to the premium while benefiting from any upside movement of the stock above \$120. Generally, for an option contract defined at time  $t$  by  $(c_t, K, T)$  and traded over the time interval  $t \in [0, T]$ , the profit resulting from such a transaction at time  $t$  is given by:  $c_t - c_0$  where  $c_t$  is the option price traded at time  $t$ , reflecting the information available at this time. Using conditional estimates of this random variable, we have the following expectation:

$$c_t = e^{-R_f(T-t)} \int_K^\infty (\tilde{S} - K) dF_{T|t}(\tilde{S})$$

where  $F_{T|t}(\tilde{S})$  is the (risk neutral) distribution function of the stock (asset) price at time  $T$  based on the information at time  $t$ .

There may also be some private information, resulting from the individual investors' analysis and access to information (information albeit commonly available but not commonly used), etc. Of course, as time changes, information will change as well, altering thereby the value of such a transaction. In other words, a learning process (such as filtering and forecasting of the underlying process) might be applied to alter and improve the individual investor's estimates of future prices and their probabilities. For example, say that we move from time  $t$  to time  $t + \Delta t$ . Assuming that the option is tradable (it can be bought or sold at any time), the expected value would be:

$$c_{t+\Delta t} = e^{-R_f(T-t-\Delta t)} \int_K^\infty (\tilde{S} - K) dF_{T|t+\Delta t}(\tilde{S})$$



**Figure 7.2** A plain call option.

where  $F_{T|t+\Delta t}(\tilde{S})$  expresses the future stock price distribution reached at time  $t + \Delta t$ . Under risk-neutral pricing, the value of the option at any time equals its discounted expectation or  $c_{t+\Delta t}$  changing over time and with incoming new information. As a result, if the option was bought at time  $t$ , it might be sold at time  $t + \Delta t$  for a profit (or loss) of  $(c_{t+\Delta t} - c_t)$ , or it might be maintained with the accounting change in the value of the option registered if it is not sold. If the probabilities are 'objective' historical distributions, then of course, the profit – loss parameters are in fact random variables with moments we can calculate (theoretically or numerically). The decision to act one way or the other may be based on beliefs and the economic evaluation of the fundamentals or on technical analyses whose ultimate outcomes are: 'is the price of the stock rising or decreasing', 'will the volatility of the stock increase or decrease', 'are interest rates changing or not and in what direction' etc. The option's Greek sensitivity parameters (Delta, Vega etc.) provide an assessment of the effects of change in the respective parameters.

The same principle applies to other products and contracts that satisfy the same conditions, such as commodity trades, foreign exchange, industrial input factors, interest rates etc. In most cases, however, each contract type has its own specific characteristics that must be accounted for explicitly in our calculations. Further, for each contract bought there must be an investor (or speculator) supplying such a contract. In our case, in order to buy a long call, there must also be seller, who is buying the call short – in other words, collecting the premium  $c$  against which he will assume the loss of a profit in the case of the stock rising above the strike price at its exercise. Such transactions occur therefore because investors/speculators have varied preferences, allowing exchanges that lead to an equilibrium where demand and the supply for the specific contract are equal.

### Example: Short selling

A short sell consists in the promise to sell a security at a given price at some future date. To do so, the broker 'borrows' the security from another client and sells it in the market in the usual way. The short seller must then buy back the security at some specified time to replace it in the client's portfolio. The short seller assumes then all costs and dividends distributed in the relevant period of the financial sell contract. For example, if we short sell 100 GM shares sold in January at \$25 while in March the contract is exercised when the price of the stock is \$20, and in February 1, a dividend of \$1 was distributed to GM shareholders, then, the short seller profit is:  $100(25 - 20 - 1) = \$400$ .

In a similar manner we may consider *long put options*. They consist in the option to sell a certain asset at a certain date for a certain strike price  $K$  or at the market price  $S(T)$ , whichever is largest, or  $\text{Max}(S(T), K)$ . The cost of such an option is denoted by  $p$ , the strike is  $K$  and the exercise time is  $T$ . For a speculator, such an option is bought when the investor expects the market to be bearish and/or the asset volatility to be bullish. Unlike the long call, the long put combines a limited upside exposure with a high gearing in a falling market. The costs/payoffs of a portfolio based on a single long put option contract is therefore given by the following.

**A put option**

Time $t = 0$	Final time $T$
$p + S_0$	$\text{Max}(K, S_T)$

As a result, in risk-neutral pricing, the price of the put is:

$$p = e^{-R_f T} E^*[\text{Max}(K, \tilde{S}_{T|0})] - S_0$$

Equivalently, it is possible at the time the put is bought, that a forward contract  $F(0, T)$  for the security to be delivered at time  $T$  is taken (see also Chapter 8). In this case (assuming again risk-neutral pricing),  $F(0, T) = e^{R_f T} S_0$  and therefore the value of a put option can be written equivalently as follows:

$$p = e^{-R_f T} E^*\{\text{Max}[K - F(0, T), \tilde{S}_{T|0} - F(0, T)]\}$$

Put–call parity can be proved from the two equations derived here as well. Note that the value of the call can be written as follows:

$$c = e^{-R_f T} E^*(\text{Max}[\tilde{S}_{T|0} - K, 0]) = e^{-R_f T} E^*(\text{Max}[\tilde{S}_{T|0}, K]) - e^{-R_f T} K$$

and therefore:  $c + e^{-R_f T} K = e^{-R_f T} E^*(\text{Max}[\tilde{S}_{T|0}, K])$ . Thus, we have the call–put parity seen in the previous chapter:

$$p + S_0 = c + e^{-R_f T} K$$

Since  $p \geq 0$  and  $c \geq 0$  the put–call parity implies trivially the following bounds:  $p \geq e^{-R_f T} K - S_0$  and  $c \geq S_0 - e^{-R_f T} K$ . Thus, for a trader selling a put, the put writer, the maximum liability is the value of the underlying stock. These transactions are popular when they are combined with another (or several other) transactions. These strategies are used to both speculate and hedge. For example, say that a put option on CISCO is bought for a strike of \$42 per share whose premium is \$2.25 while the stock current price is \$46 per share. Then, an investor would be able to contain any loss due to stock decline, to the premium paid for the put (thereby hedging downside losses). Thus, if the stock falls below \$42, the maximum loss is:  $(46 - 42) + 2.25 = \$6.25$  per share. While, if the stock increases to \$52, the gain would be:  $-2.25 + \{52 - 46\} = 3.75$ .

Some firms use put options as a means to accumulate information. For example, some investment firms buy puts (as warrants) in order to generate a signal from the firm they intend to invest in (or not). If a firm responds positively to a request to a put (warrant) contracts, then this may be interpreted as a signal of ‘weakness’ – the firm willing to sell because it believes it is overpriced – and vice versa, if it does not want to sell it might mean that the firm estimates that it is underpriced. Such information eventually becomes common knowledge, but for some investment firms, the signals they receive are private information which remains private for at least a certain amount of time and provides such firms with a competitive advantage (usually, less than four months) which is worth paying for and to speculate with.

### 7.4.2 Covered call strategies: selling a call and a share

Say that a pension fund holds 1000 GM shares with a current price of \$130 per share. A decision is reached to sell these shares at \$140 as well as a call expiring in 90 days with an exercise price of \$140 at a premium of \$5 per share. As a result, the fund picks up an immediate income of \$5000 while the fund would lose its profit share for the stock when it reaches levels higher than \$140. However, since it intended to sell its holdings at \$140 anyway, such a profit would have not been made. This strategy is called a *covered call*. It is based on a portfolio consisting in the purchase of a share of stock with the simultaneous sale of a (short) call on that stock in order to pick up an extra income (the call option price), on a transaction that is to be performed in any case. The price (per share) of such a transaction is the expectation of the following random variable written as follows:

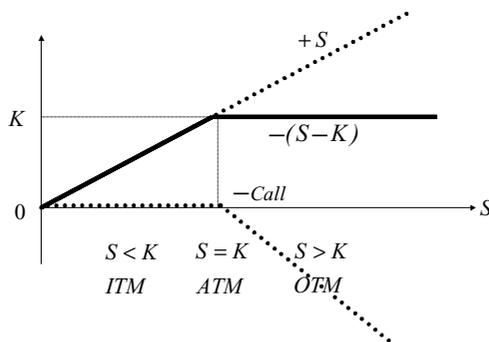
$$S - c = e^{-R_f T} E^*[\text{Min}(K, \tilde{S}_{T|0})]$$

where  $c$  is the call premium received,  $K$  is the strike price of the option with an exercise at time  $T$ , while  $S_T$  is the stock price at the option exercise time. Under risk-neutral pricing, there is a gain for the seller of the call since he picks up the premium on a transaction that he is likely to perform anyway.

#### A covered call

Time $t$	Final time $T$
$S - c$	$\text{Min}(K, S_T)$

The buyer of the (short) call, however, is willing to pay a premium because he needs the option to limit his potential losses. As a result, a market is created for buyers and sellers to benefit from such transaction. The terminal payoff of a covered call is given in the Figure 7.3.



**Figure 7.3** Terminal payoffs: covered call.

**Problem**

Write the payoff equation for a covered call when the seller of the stock is financed by a forward.

**7.4.3 Put and protective put strategies: buying a put and a stock**

The protective put is a portfolio that consists in buying a stock and a put on the stock. It is a strategy used when we seek protection from losses below the put option price. For this reason, it is often interpreted as an insurance against downside losses. Buying a put option on stock provides an investor with a limit on the downside risk while maintaining the potential for unlimited gains. Banks for example, use a protective put to protect their principal from interest rate increases. Similarly, say that a euro firm receives an income from the USA in dollars. If the dollar depreciates or the euro increases, it will of course be financially hurt. To protect the value of this income (in the local currency), the firm can buy a put option by selling dollars and obtain protection in case of a downward price movement. The protective put has therefore the following cash flow, summarized in the table below.

**A protective put strategy**

Time $t = 0$	Final time $T$
$p + S$	$\text{Max}(K, S_T)$

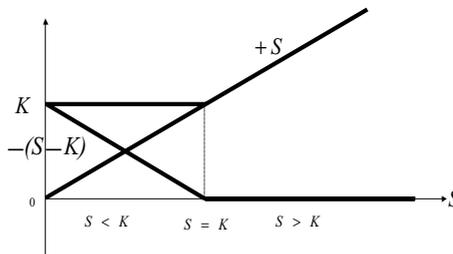
where  $p$  is the put premium and:

$$p + S = e^{-R_f T} E^*[\text{Max}(K, \tilde{S}_{T|0})]$$

or, equivalently:

$$p = e^{-R_f T} E^*[\text{Max}(K - e^{-R_f T} S, \tilde{S}_{T|0} - e^{-R_f T} S)]$$

This is shown graphically in Figure 7.4 and illustrated by exercise of the put.



**Figure 7.4** Terminal payoffs: Protective put.

### 7.4.4 Spread strategies

Spread strategy consists in constructing a portfolio by taking position in two or more options of the same type but with different strike prices. For example, a spread over two long call options can be written by  $W = c_1 + c_2$  Where  $c_i$ ,  $i = 1, 2$  are call option prices associated with the strike prices  $K_i$ ,  $i = 1, 2$ . The following table summarizes a spread strategy cash flow.

#### A call–call spread strategy

Time $t = 0$	Final time $T$
$c_1 + c_2$	$\text{Max}(K_1, S_T) + \text{Max}(K_2, S_T)$

There are also long and short put spread versus call and, vice versa, long and short call versus put. In a short put spread versus call, we sell a put with strike price  $B$  and sell a put at a lower strike  $A$  and buy a call at any strike. The long call will generally be at a higher strike price,  $C$ , than both puts. The return profile turns out to be similar to that of a short put spread, but the long call provides an unlimited profit potential should the underlying asset rise above  $C$ . Such a transaction is performed when the investor expects the market and the volatility to be bullish. In a rising market the potential profit would be unlimited while in a falling market, losses are limited. This is represented graphically below. For example, in the expectation of a stock price increasing, a speculator will buy a call at a low strike price  $K_1$  and sell another with a high strike price  $K_2 > K_1$  (this is also called a *bull spread*). This will have the effect of delimiting the profit/loss potential of such a trade, as shown in the Figure 7.5. In this case, the value of such a spread is as follows.

#### A bull spread

The premium collected initially at time  $t = 0$  is  $c_1 - c_2$ , while (under risk-neutral pricing):

$$c_1 - c_2 = e^{-R_f T} E^* \{ \text{Max}[(\tilde{S}_{T|0} - K_2), 0] \} - e^{-R_f T} E^* \{ \text{Max}[(\tilde{S}_{T|0} - K_1), 0] \}$$

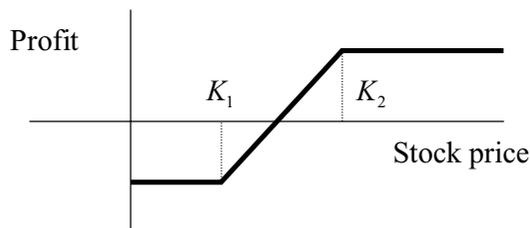
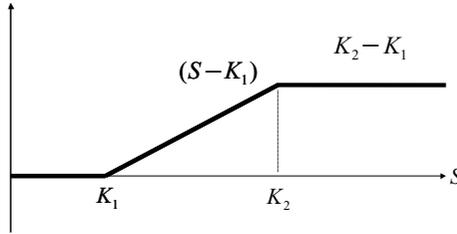


Figure 7.5 A bull spread.



**Figure 7.6** Terminal payoffs: bullish spread.

By contrast, in the expectation that stock prices will fall, the investor/speculator may buy a call option with a high strike price and sell a (short) call with a lower strike price. *This is also called a bear spread.* In this case, the initial cash inflow would be.

$$0 = c_1 - c_2 + e^{-R_f T} E^* \{ \text{Max}[(\tilde{S}_{T|0} - K_2), 0] \} - e^{-R_f T} E^* \{ \text{Max}[(\tilde{S}_{T|0} - K_1), 0] \}$$

For a bullish spread, however, we buy and sell:

$$\text{Long call } (K_1, T) + \text{Short call } (K_2, T) = C(K_1, T) - C(K_2, T)$$

while the payoff at maturity is:

$S_T < K_1$	$K_1 < S_T < K_2$	$S_T > K_2$	Calls
0	$S_T - K_1$	$S_T - K_1$	$C(K_1, T)$
0	0	$-(S_T - K_2)$	$-C(K_2, T)$
0	$S_T - K_1$	$K_2 - K_1$	

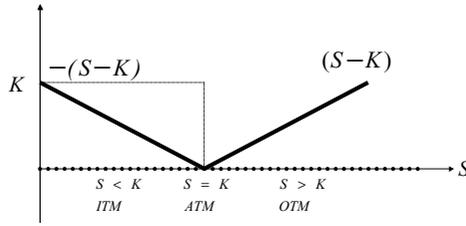
This is given graphically in Figure 7.6.

### 7.4.5 Straddle and strangle strategies

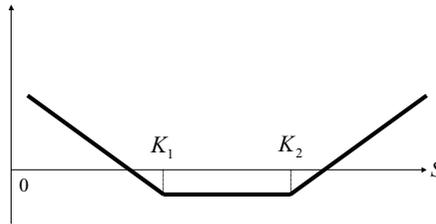
A *straddle* consists in buying both a call and a put on a stock, each with the same strike price,  $K$ , at the exercise date,  $T$ , and selling a call at any strike. A straddle is used by investors who believe that the stock will be volatile (moving strongly but in an unpredictable direction). A straddle can be long and short. A long straddle versus call consists of buying both a call and a put at the same strike but in addition, selling a call at any strike. Similarly, expectation of a takeover or an important announcement by the firm is also a good reason for a straddle. The cash flows associated with a straddle based on buying a put and a call with the same exercise price and the same expiration date is thus (see Figure 7.7):

$$p + c = e^{-R_f T} E^* \text{Max}[\tilde{S}_{T|0} - K, K - \tilde{S}_{T|0}]$$

In contrast to a straddle, a *strangle* consists of a portfolio but with different strikes for the put and the call. The graph for such a strangle is given in Figure 7.8.



**Figure 7.7** A Straddle strategy: terminal payoffs.



**Figure 7.8** A strangle strategy.

#### 7.4.6 Strip and strap strategies

If the investor believes that there is soon to be a strong stock price move, but with potentially a stronger probability of a downward move, then the investor can use a *strip*. This is similar to a straddle, but it is asymmetric. To implement a strip, the investor will take a long position in one call and two in puts. Inversely, if the investor believes that there is a stronger chance that the stock price will move upwards, then the investor will implement a *strap*, namely taking a long position on two calls and one on a put. In other words, the economic value of such strategies are given by the following payoff:

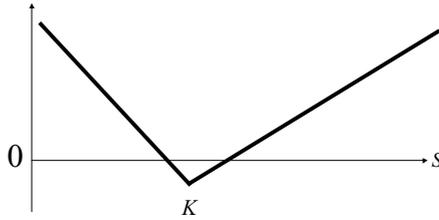
##### Strip strategies

	$S_T < K$	$S_T > K$
1 Long call	0	$S_T - K$
2 Long puts	$2(K - S_T)$	0
	$2(K - S_T)$	$S_T - K$

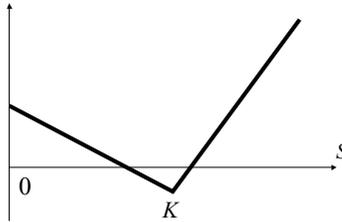
The graph of such a cash flow is given in Figure 7.9.

##### Strap strategies

In a strap (see Figure 7.10) we make an equal bet that the market will go up or down and thus a portfolio is constructed out of two calls and one put. The payoffs



**Figure 7.9** A strip strategy.



**Figure 7.10** A Strap strategy.

are given by the following table.

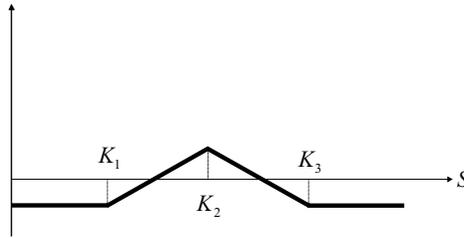
	$S_T < K$	$S_T > K$
2 Long calls	0	$2(S_T - K)$
1 Long put	$(K - S_T)$	0
	$(K - S_T)$	$2(S_T - K)$

### 7.4.7 Butterfly and condor spread strategies

When investors believe that prices will remain the same, they may use *butterfly spread strategies* (see Figure 7.11). This will ensure that if prices move upward or downward, then losses will be limited, while if prices remain at the same level the investor will make money. In this case, the investor will buy two call options, one with a high strike price and the other with a low strike price and at the same time, will sell two calls with a strike price roughly halfway in between (roughly equalling the spot price). As a result, butterfly spreads merely involve options with three different strikes. *Condor* spreads are similar to butterfly spreads but involve options with four different strikes.

### 7.4.8 Dynamic strategies and the Greeks

In the previous chapter we have drawn attention to the ‘Greeks’, expressing the option’s price sensitivity to the parameters used in calculating the option’s price. These measures are summarized below with an interpretation of their signs



**Figure 7.11** A butterfly spread strategy.

(Willmott, 2002). This sensitivity is used practically to take position on options and stocks:

### Delta

$$\Delta = \left\{ \frac{\text{Exposure to direction of price changes, Dollar change in position value}}{\text{Dollar change in underlying security price}} \right.$$

When Delta is negative, it implies a bearish situation with a position that benefits from a price decline. When Delta is null, then there is little change in the position as a function of the price change. Finally, when the sign is positive, it implies a bullish situation with a position that benefits from a price increase.

### Lambda

$$\Lambda = \left\{ \frac{\text{Leverage of position price elasticity, Percentage change in position value}}{\text{Percentage change in underlying security price}} \right.$$

The Lambda has the same implications as the Delta.

### Gamma

$$\Gamma = \left\{ \frac{\text{Exposure to price instability; 'non - directional price change', Change in position Delta}}{\text{Dollar change in underlying security price}} \right.$$

When Gamma is positive or negative, the position benefits from price instability, while when it is null, it is not affected by price instability.

### Theta

$$\Theta = \left\{ \frac{\text{Exposure to time decay, Dollar change in position value}}{\text{Decrease in time to expiration}} \right.$$

When Theta is negative the position value declines as a function of time and vice versa when Theta is positive. When Theta is null, the position is insensitive to time.

**Kappa**

$$\kappa = \left\{ \frac{\text{Exposure to changes in volatility of prices, Dollar change in position value}}{\text{One percent change in volatility of prices}} \right.$$

When Kappa is negative, the position benefits from a drop in volatility and vice versa. Of course, when Kappa is null, then the position is not affected by volatility.

The option strategies introduced above are often determined in practice by the ‘Greeks’. That is to say, based on their value (positive, negative, null) a strategy is implemented. A summary is given in Table 7.1 (Wilmott, 2002).

**Table 7.1** Common option strategies.

Delta	Gamma	Theta	Strategy	Implementation
Positive	Positive	Negative	Long call	Purchase long call option
Negative	Positive	Negative	Long put	Purchase long put option
Neutral	Positive	Negative	Straddle	Purchase call and put, both with same exercise and expiration date
Neutral	Positive	Negative	Strangle	Purchase call and put, each equally out of the money, and write a call and a put, each further out of the money, and each with the same expiration date
Neutral	Positive	Negative	Condor	Purchase call and put, each equally out of the money, and write a call and a put, each further out of the money than the call and put that were purchased. All options have the same expiration date
Neutral	Negative	Positive	Butterfly	Write two at-the-money calls, and buy two calls, one in the money and the other equally far out of the money
Positive	Neutral	Neutral	Vertical spread	Buy one call and write another call with a higher exercise price. Both options have the same time to expiration
Neutral	Negative	Positive	Time spread	Write one call and buy another call with a longer time to expiration. Both options have the same exercise price
Neutral	Positive	Negative	Back spread	Buy one call and write another call with a longer time to expiration. Both options have the same exercise price.
Neutral	Neutral	Neutral	Conversion	Buy the underlying security, write a call, and buy a put. The options have the same time to expiration and the same exercise price

**Problem**

Discuss the strategy buy a call and sell a put  $c - p$ . Discuss the strategy to buy a stock and borrow the present value  $S - K/(1 + R_f)^T$ . Verify that these two strategies are equivalent. In other words verify that:  $c - p = S - K/(1 + R_f)^T$ .

**Problem: A portfolio with return guarantees**

Consider an investor whose initial wealth is  $W_0$  seeking a guaranteed wealth level  $aS_T + b$  at time  $T$  where  $S_T$  is the stock price. This guarantee implies initially that  $W_0 \geq aS_0 + b e^{-R_f T}$  where  $R_f$  is the risk-free rate. Determine the optimal portfolio which consists of an investment in a zero coupon bond with nominal value,  $B_T$  at time  $T$  or  $B_T = B_0 e^{R_f T}$ , an investment in a risky asset whose price  $S_t$  is a given by a lognormal process and finally, an investment in a European put option whose current price is  $P_0$  with an exercise price  $K$ .

**7.5 STOPPING TIME STRATEGIES\*****7.5.1 Stopping time sell and buy strategies**

Buy low, sell high is a sure prescription for profits that has withstood the test of time and markets (Connolly, 1977; Goldman *et al.*, 1979). Waiting too long for the high may lead to a loss, while waiting too little may induce insignificant gains and perhaps losses as well. In this sense, trade strategies of the type ‘buy–hold–sell’ involve necessarily both gains and losses, appropriately balanced between what an investor is willing to gamble and how much these gambles are worth to him.

Risk-neutral pricing (when it can be applied) has resolved the dilemma of what utility function to use to price uncertain payoffs. Simply none are needed since the asset price is given by the (rational) expectation of its future values discounted at the risk-free discount rate. Market efficiency thus implies that it makes no sense for an investor to ‘learn’, to ‘seek an advantage’ or even believe that he can ‘beat the market’. In short, it denies the ability of an investor who believes that he can be cunning, perspicacious, intuition-prone or whatever else that may lead him to make profits by trading. In fact, in an efficient market, one makes money only if one is lucky, for in the short run, prices are utterly unpredictable. Yet, investors trade and invest trillions daily just because they think that they can make money. In other words, they may know something that we do not know, are plain gamblers or plain ‘stupid’. Or perhaps, they understand something that makes the market incomplete and find potential arbitrage profits. This presumption, that traders and investors trade because they believe that markets are incomplete, results essentially in ‘market’ forces that will correct market incompleteness and thus lead eventually to efficient markets. For as long as there are arbitrage profits they will take advantage of the market inefficiency till it is no longer possible to do so and the market becomes efficient. In this sense a market may provide an opportunity for profits which cannot be maintained forever since the market will be self-correcting. The cunning investor is thus one who understands that an

opportunity to profit from ‘inefficiency’ cannot last forever, knows when to detect it, identify and prospect on this opportunity and, importantly, knows as well when to get out. For at some time, this inefficiency will be obliterated by other investors who have realized that profits have been made and, wishing to share in it, will necessarily render the market efficient.

For these reasons, in practice, there may be problems in applying the risk-neutral framework. Numerous situations to be elaborated in Chapter 9, such as risk-sensitive individual investors expressing varied preferences, using discrete time and historical data, using private information etc. contribute to market inefficiencies. In such a framework, future price distributions possess risk properties that may be more or less desired by the individual investor/speculator who may either use a risk-adjusted discount factor or provide a risk qualification for the financial decisions he may wish to assume.

In this section we shall consider first a risk-neutral framework (essentially to maintain the risk-free discounting of risk-neutral pricing) and consider the decision to buy or sell an asset under the martingale probability measure. We shall show, using an example, that no profits can be made when the trade strategy is to sell as soon as a given price (greater than the current price) is reached. Of course, in practice we can use the actual probability measure (rather than the risk-neutral measure) but then it would be necessary to apply the individual investor’s discount rate. This procedure is followed subsequently and we demonstrate through examples how the risk premium associated with a trading strategy for a risk-sensitive investor/trader is determined. Finally, we consider a quantile risk-sensitive (Value at Risk – VaR) investor and assess a number of trading strategies for such an investor. Both discrete and continuous-time price processes are considered and therefore some of the problems considered may be in an incomplete market situation. For practical purposes, when the risk-neutral framework can be applied, simulation can be used to evaluate a trading strategy, however complex it may be (since simulation merely applies an experimental approach and assess performance of the trading strategy based on frequency and average concepts). When this is not the case, simulation must be applied carefully for the discount rate applied to simulated cash streams must necessarily account for the premium payments to be incurred for the cash flow’s random characteristics.

To demonstrate the technique used, we begin again with the lognormal stock price process:

$$\frac{dS}{S} = \alpha dt + \sigma dW, \quad S(0) = S_0$$

Its solution at time  $t$  is simply:

$$S(t) = S(0) \exp\left(\left(\alpha - \frac{\sigma^2}{2}\right)t + \sigma W(t)\right), \quad W(t) = \int_0^t dW(t)$$

where  $W(t)$  is the Brownian motion. It is possible to rewrite this expression as

follows (adding and subtracting in the exponential  $R_f t$ ):

$$S(t) = S(0) \exp \left\{ \left( R_f - \frac{\sigma^2}{2} \right) t + \sigma \left( W(t) + \frac{\alpha - R_f}{\sigma} t \right) \right\}$$

In order to apply the risk-neutral pricing framework, we define another probability measure or a numeraire with respect to which expectation is taken. Let

$$W^*(t) = \left( W(t) + \frac{\alpha - R_f}{\sigma} t \right),$$

then

$$S(t) = S(0) \exp \left\{ \left( R_f - \frac{\sigma^2}{2} \right) t + \sigma W^*(t) \right\}$$

which corresponds to the (transformed) price process:

$$\frac{dS}{S} = R_f dt + \sigma dW^*(t), \quad S(0) = S_0$$

And the current price equals the expected future price under risk neutral pricing since:

$$\begin{aligned} S(0) &= e^{-R_f t} E^* (S(t)) = e^{-R_f t} E^* \left( S(0) \exp \left\{ \left( R_f - \frac{\sigma^2}{2} \right) t + \sigma W^*(t) \right\} \right) \\ &= S(0) e^{-\sigma^2 t/2} E^* (e^{\sigma W^*(t)}) \end{aligned}$$

Since  $E^*$  is an expectation taken with respect to the risk neutral process, we have:

$$e^{-\sigma^2 t/2} E^* (e^{\sigma W^*(t)}) = e^{-\sigma^2 t/2} e^{\sigma^2 t/2} = 1$$

Thus, the current price equals an expectation of the future price. It is important to remember, however, that the proof of such a result is based on our ability to replicate such a process by a risk-free process (and thereby value it by the risk-free rate). In terms of the historical process, we have first:

$$dW^*(t) = dW(t) + \lambda dt, \quad \lambda = \frac{\alpha - R_f}{\sigma}$$

which we insert in the risk-neutral process and note that:

$$\frac{dS}{S} = R_f dt + \sigma [dW(t) + \lambda dt] = [R_f + \sigma \lambda] dt + \sigma dW(t)$$

where  $R_f$  is the return on a risk-free asset while  $\lambda = (\alpha - R_f)/\sigma$  is the premium (per unit volatility) for an asset whose mean rate of return is  $\alpha$  and its volatility is  $\sigma$ . In other words, risk-neutral pricing is reached by equating the stock price process whose return equals the risk-free rate plus a return of  $\alpha - \sigma \lambda$  that compensates for the stock risk, or:

$$\begin{aligned} \frac{dS}{S} &= (R_f + \alpha - \sigma \lambda) dt + \sigma [dW(t) + \lambda dt] \\ &= R_f dt + \sigma [dW(t) + \lambda dt] = R_f dt + \sigma dW^*(t) \end{aligned}$$

Under such a transformation, risk-neutral pricing is applicable and therefore financial assets may be valued by expectations using the transformed risk-neutral process or adjusting the price process by its underlying risk premium (if it can be assessed of course by, say, regressions that can estimate risky stock betas using the CAPM). Consider next the risk-neutral framework constructed and evaluate the decision to sell an asset we own (whose current price is  $S_0$ ) as soon as its price reaches a given level  $S^* > S_0$ . The profit of such a trade under risk-neutral pricing is:

$$\pi_0 = E^* e^{-R_f \tau} S^* - S_0$$

Here the stopping (sell) time is random, defined by the first time the target sell price is reached:

$$\tau = \text{Inf}\{t > 0, S(t) \geq S^*; S(0) = S_0\}$$

We shall prove that under the risk-neutral framework, there is an ‘equivalence’ to selling now or at a future date. Explicitly, we will show that  $\pi_0 = 0$ . Again, let the risk-neutral price process be:

$$\frac{dS}{S} = R_f dt + \sigma dW^*(t)$$

and consider the equivalent return process  $y = \ln S$

$$\begin{aligned} dy &= \left(R_f - \frac{\sigma^2}{2}\right) dt + \sigma dW^*(t), y(0) = \ln(S_0) \\ \tau &= \text{Inf}\{t > 0, y(t) \geq \ln(S^*); y(0) = \ln(S_0)\} \end{aligned}$$

As a result,  $E_S^*(e^{-R_f \tau}) = E_y^*(e^{-R_f \tau})$  which is the Laplace transform of the sell stopping time when the underlying process has a mean rate and volatility given respectively by  $\mu = R_f - \sigma^2/2$ ,  $\sigma$  (see also the mathematical Appendix to this chapter):

$$\begin{aligned} g_{R_f}^*(S^*, \ln S_0) &= \exp \left[ \frac{\ln S_0 - \ln S^*}{\sigma^2} \left( -\mu + \sqrt{\mu^2 + 2R_f \mu \sigma^2} \right) \right], \\ &\sigma > 0, -\infty < \ln S_0 \leq \ln S^* < \infty \end{aligned}$$

The expected profit arising from such a transaction is thus:

$$\begin{aligned} \pi_0 &= S^* E^*(e^{-R_f \tau}) - S_0 = S^* g_{R_f}^*(\ln S^*, \ln S_0) - S_0 \\ &= S^* \exp \left[ \frac{\ln S_0 - \ln S^*}{\sigma^2} \left( -\mu + \sqrt{\mu^2 + 2R_f \mu \sigma^2} \right) \right] - S_0 \end{aligned}$$

That is to say, such a strategy will in a risk-neutral world yield a positive return if  $\pi_0 > 0$ . Elementary manipulations show that this is to equivalent to:

$$\pi_0 > 0 \text{ if } \frac{\sigma^2}{2} > (R_f - 1)\mu \quad \text{or} \quad \frac{\sigma^2}{2} > (1 - R_f) \left( \frac{\sigma^2}{2} - R_f \right) \text{ if } R_f > \frac{\sigma^2}{2}$$

As a result,

$$\pi_0 = \begin{cases} > 0 & \text{if } R_f > \frac{\sigma^2}{2} \\ < 0 & \text{if } R_f < \frac{\sigma^2}{2} \end{cases}$$

The decision to sell or wait to sell at a future time is thus reduced to the simple condition stated above. An optimal selling price in these conditions can be found by optimizing the return of such a sell strategy which is found by noting that either it is optimal to have a selling price as large as possible (and thus never sell) or select the smallest price, implying selling now at the current (any) price. If the risk-free rate is 'small' compared to the volatility, then it is optimal to wait and, vice versa, a small volatility will induce the holder of the stock to sell. In other words,

$$\frac{d\pi_0}{dS^*} = \begin{cases} > 0 & R_f < \sigma^2/2 \\ < 0 & R_f > \sigma^2/2 \end{cases}$$

Combining this result with the profit condition of the trade, we note that:

$$\begin{cases} \frac{d\pi_0}{dS^*} > 0, \pi_0 < 0 & \text{if } R_f < \sigma^2/2 \\ \frac{d\pi_0}{dS^*} < 0, \pi_0 > 0 & \text{if } R_f > \sigma^2/2 \end{cases}$$

And therefore the only solution that can justify these conditions is  $\pi_0 = 0$ , implying that whether one keeps the asset or sells it is irrelevant, for under risk-neutral pricing, the profit realized from the trade or of maintaining the stock is equivalent. Say that  $R_f < \sigma^2/2$  then a 'wait to sell' transaction induces an expected trade loss and therefore it is best to obtain the current price. When  $R_f > \sigma^2/2$ , the expected profit from the trade is positive but it is optimal to select the lowest selling price which is, of course, the current price and then, again, the profit transaction,  $\pi_0 = 0$  will be null as our contention states.

Similar results are obtained when the time to exercise the sell strategy is finite. In this case,

$$\pi_{0,T} = S^* E^*(e^{-R_f \tau \Delta T}) - S_0 = 0$$

which turns out to have the same properties as above.

For a risk-sensitive investor (trader or speculator) whose utility for money is  $u(\cdot)$ , a decision to sell or wait will be based on the following (using in this case the actual probability measure and an individual discount rate  $R_i$ ):

$$\text{Max}_{S^* \geq S_0} Eu(\tilde{\pi}_0) = Eu(S^* e^{-R_i \bar{\tau}} - S_0)$$

If, at the end of the period, we sell the asset anyway, then the optimal trading/sell condition is:

$$\text{Max}_{S^* \geq S_0} \int_0^T Eu(S^* e^{-R_i \tau} - S_0) e^{-R_i \tau} g(\tau) d\tau + E_{S(T)} u(S(T) e^{-R_i T} - S_0) [1 - G(T)]$$

where  $g(\cdot)$  and  $G(\cdot)$  are the inverse Gaussian probability and cumulative distributions respectively.

$$g(S^*, t; S_0) = \frac{|S_0 - S^*|}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(S^* - S_0 - R_f t)}{2\sigma^2 t}\right],$$

$$1 - G(T) = \left[1 - 2\Phi\left(\frac{-(S^* - S_0)}{\sigma\sqrt{T}}\right)\right]$$

When we use historical data, the situation is different, as we discussed earlier. Say that the underlying asset has the following equation:

$$dS/S = b dt + \sigma dW, \quad S(0) = S_0$$

To determine the risk-sensitive rate associated with a trading strategy, we can proceed as follows. Say that a risk-free zero coupon bond pays  $S^*$  at time  $T$  whose current price is  $B^*$ . In other words,  $B^* = e^{-R_{f,T}T} S^*$ . Thus,  $S^* = B^* e^{R_{f,T}T}$  where  $R_{f,T}$  is a known discount rate applied to this bond. Since (see also the mathematical Appendix):

$$g_{R_f}^*(S^*, \ln S_0) = \exp\left[-\frac{\ln S^*/S_0}{\sigma^2} \left(-b + \sqrt{b^2 + 2R_f b\sigma^2}\right)\right], \text{ and}$$

$$S_0 = S^* \exp\left[-\frac{\ln S^*/S_0}{\sigma^2} \left(-b + \sqrt{b^2 + 2R_i b\sigma^2}\right)\right]$$

We obtain:

$$S_0 = B^* \exp\left[R_{f,T}T - \frac{\ln(B^* e^{R_{f,T}T}/S_0)}{\sigma^2} \left(-b + \sqrt{b^2 + 2R_i b\sigma^2}\right)\right]$$

This is solved for the risk-sensitive discount rate:

$$R_i = \left[1 + \sigma^2 \frac{[R_{f,T}T - \ln(S_0/B^*)]}{2b \ln(B^* e^{R_{f,T}T}/S_0)}\right] \frac{[R_{f,T}T - \ln(S_0/B^*)]}{\ln(B^* e^{R_{f,T}T}/S_0)}$$

### Example: Buying and selling on a random walk

The problem considered above can be similarly solved for a buy/sell strategy in a random walk. To do so, consider a binomial price process where price increase or decline by \$1 with probabilities  $p$  and  $q$  respectively. Set the initial price to  $S_0 = 0$ . When  $p = q = 1/2$ , the price process is a martingale. Assume that we own no stock initially but we construct the trading strategy: buy a stock as soon as it reaches the prices  $S_n = -a$  and sell it soon thereafter as soon as it reaches the price  $S_n = b$ . Our problem is to assess the average profit (or loss) of such a trade. First set the first time that a buy order is made to:

$$T_a = \text{Inf}\{n, S_n = -a\}, \quad a > 0$$

Once the price ‘ $-a$ ’ is reached and a stock is bought, it is held until it reaches the price  $b$ . This time is:

$$T_b = \text{Inf}\{n, S_n = b, S_0 = -a\}$$

At this time, the profit is  $(a + b)$ . In present value terms, the profit is a random variable given by:

$$\pi_{-a,b} = -aE(\rho^{-T_a}) + bE(\rho^{-(T_a+T_b)})$$

where  $\rho$  is the discount rate applied to the trade. Given the stopping time generating function (to be calculated below), we can calculate the expected profit. Of course, initially, we put down no money and therefore, in equilibrium, it is also worth no money. That is  $\pi_{-a,b} = 0$  and therefore  $a/b = E(\rho^{-(T_a+T_b)})/E(\rho^{-T_a})$ .

### Problem

Calculate  $\pi_{-a,b}$  and compare the results under a risk-free discount rate (when risk-neutral pricing can be applied and when it is not).

Now assume that we own a stock which we sell when the price decreases by  $a$  or when it increases by  $b$ , whichever comes first. If  $-a$  is reached first, a loss ' $-a$ ' is incurred, otherwise a profit  $b$  is realized. Let the probability of a loss be  $U_a$  and let the underlying price process be a symmetric random walk (and thereby a martingale), with  $E(S_n) = E(S_T) = 0$ ,  $T = \min(T_{-a}, T_b)$  while  $E(S_T) = U_a(-a) - (1 - U_a)b$  which leads to:

$$U_a = \frac{b}{a + b}$$

As a result, in present value terms, we have a profit given by:

$$\theta_{-a,b} = -aU_aE(\rho^{-T_a}) + b(1 - U_a)E(\rho^{-T_b})$$

Some of the simple trading problems based on the martingale process can be studied using Wald's identity. Namely say that a stock price jump is  $Y_i$ ,  $i = 1, 2, \dots$  assumed to be independent from other jumps and possessing a generating function:

$$\phi(v) = E(e^{vY_1}), \quad S_n = \sum_{i=1}^n Y_i, \quad S_0 = 0$$

Now set the first time  $T$  that the process is stopped by:

$$T = \text{Inf} \{n, S_n \leq -a, S_n \geq b\}$$

Wald's identity states that:

$$E[(\phi(v))^{-T} e^{vS_n}] = 1 \text{ for any } v \text{ satisfying } \phi(v) \geq 1.$$

Explicitly, set a  $\phi(v^*) = 1$  then by Wald's identity, we have:

$$E[(\phi(v^*))^{-T} e^{v^*S_n}] = E[e^{v^*S_n}] = 1$$

In expectation, we thus have:

$$1 = E[e^{v^*S_T} | S_T \leq -a] \Pr\{S_T \leq -a\} + E[e^{v^*S_T} | S_T \geq b] \Pr\{S_T \geq b\}$$

and therefore:

$$\Pr\{S_T \geq b\} = \frac{1 - E[e^{v^* S_T} | S_T \leq -a]}{E[e^{v^* S_T} | S_T \geq b] - E[e^{v^* S_T} | S_T \leq -a]}$$

As a result, the probability of ‘making money’ ( $b$ ) is  $\Pr\{S_T \geq b\}$  while the probability of losing money ( $-a$ ) is  $1 - \Pr\{S_T \geq b\}$ .

**Problem: Filter rule**

Assess the probabilities of making or losing money in a filter rule which consists in the following. Suppose that at time ‘0’ a sell decision has been generated. The trading rule generates the next buy signal (i.e. reaching price  $b$  first and then waiting for price  $-a$  to be reached).

**Problem: Trinomial models**

Consider for simplicity the risk-neutral process (in fact, we could consider equally any other diffusion process):

$$dS/S = R_f dt + \sigma dW, \quad S(0) = S_0$$

and apply Ito’s Lemma to the transformation  $y = \ln(S)$  and obtain:

$$dy = \left( R_f - \frac{1}{2} \sigma^2 \right) dt + \sigma dW, \quad y(0) = y_0$$

Given this normal (logarithmic) price process consider the trinomial random walk approximation:

$$Y_{t+1} = \begin{cases} Y_t + f_1 & \text{w.p. } p \\ Y_t + f_2 & \text{w.p. } 1 - p - q \\ Y_t + f_3 & \text{w.p. } q \end{cases}$$

where:

$$E(Y_{t+1} - Y_t) = pf_1 + (1 - p - q)f_2 + qf_3 \approx \left( R_f - \frac{1}{2} \sigma^2 \right)$$

$$E(Y_{t+1} - Y_t)^2 = pf_1^2 + (1 - p - q)f_2^2 + qf_3^2 \approx \sigma^2$$

First assume that  $p + q = 1$  and calculate the stopping sell time for an asset we own. Apply also risk neutral valuation to calculate the price of a European call option derived from this price if the strike is  $K$  and if the exercise time is 2. If  $p + q < 1$ , explain why it is not possible to apply risk-neutral pricing? In such a case calculate the expected stopping time of a strategy which consists in buying the stock at  $Y = -Y_a$  and selling at  $Y = Y_b$ .

Stopping times on random walks are often called the ‘gambler’s ruin’ problem, inspired by a gambler playing till he loses a certain amount of capital or taking his winnings as soon as they reach a given level. For an asymmetric birth–death random walk with:

$$P(Y_i = +1) = p, \quad P(Y_i = -1) = q, \quad \text{and} \quad P(Y_i = 0) = r$$

It is well known (for example, Cox and Miller, 1965, p. 75) that the probability of reaching one or the other boundaries is given by,

$$P(Y_t = -a) = \begin{cases} \frac{1 - (1/\lambda)^b}{1 - (1/\lambda)^{a+b}} & \lambda \neq 1 \\ b/(a+b) & \lambda = 1 \end{cases} \quad P(Y_t = b) = \begin{cases} \frac{(1/\lambda)^b - (1/\lambda)^{a+b}}{1 - (1/\lambda)^{a+b}} & \lambda \neq 1 \\ a/(a+b) & \lambda = 1 \end{cases}$$

where  $\lambda = q/p$ . Further,

$$E(T_{-a,b}) = \begin{cases} \left[ \frac{1}{1-r} \right] \left[ \frac{\lambda+1}{\lambda-1} \right] \left[ \frac{a(\lambda^b - 1) + b(\lambda^{-a} - 1)}{\lambda^b - \lambda^{-a}} \right] & \lambda = q/p \\ \frac{ab}{1-r} & \lambda = 1 \end{cases}$$

Note that  $(\lambda^{x_n}; n \geq 0)$  is a martingale which is used together with the stopping theorem to prove these results (see also Chapter 4). For example, as discussed earlier, at the first loss  $-a$  or at the profit  $b$  the probability of 'making money' is  $P(S_{T(-a,b)} = b)$  while the probability of losing it is  $P(S_{T(-a,b)} = -a)$ , as calculated above. The expected amount of time the trade will be active is also  $E(T_{-a,b})$ . For example, if a trader repeats such a process infinitely, the average profit of the trader strategy would be given by:

$$\bar{\pi}(-a, b) = \frac{bP(S_{T(-a,b)} = b) - aP(S_{T(-a,b)} = -a)}{E(T_{-a,b})}$$

Of course, the profit from such a trade is thus random and given by:

$$\tilde{\pi} = \begin{cases} -a & \text{w.p.} \begin{cases} \frac{1 - (1/\lambda)^b}{1 - (1/\lambda)^{a+b}} & \lambda \neq 1 \\ b/(a+b) & \lambda = 1 \end{cases} \\ b & \text{w.p.} \begin{cases} \frac{(1/\lambda)^b - (1/\lambda)^{a+b}}{1 - (1/\lambda)^{a+b}} & \lambda \neq 1 \\ a/(a+b) & \lambda = 1 \end{cases} \end{cases}$$

And therefore, the expected profit, its higher moments and the average profit can be calculated. When  $\lambda = 1$  in particular, the long run average profit is null and the variance equals  $2ab$ , or:

$$E(\tilde{\pi}) = \frac{(-ab + ba)}{a+b} = 0, \text{ var}(\tilde{\pi}) = 2ab$$

Thus, a risk-averse investor applying this rule will be better off doing nothing, since there is no expected gain. When  $\rho \neq 1$  while  $r = 0$  (the random walk) we have (Cox and Miller, 1965, p. 31):

$$P(S_{T(-a,b)} = b) = \frac{\lambda^a - 1}{\lambda^{a+b} - 1}; P(S_{T(-a,b)} = -a) = \frac{\lambda^{a+b} - \lambda^a}{\lambda^{a+b} - 1}$$

and

$$E(T_{-a,b}) = \left[ \frac{\lambda + 1}{\lambda - 1} \right] \left[ \frac{a(\lambda^b - 1) + b(\lambda^{-a} - 1)}{\lambda^b - \lambda^{-a}} \right]$$

while the long run average profit is:

$$\bar{\pi}(-a, b) = \frac{[b(\lambda^a - 1) - a(\lambda^{a+b} - \lambda^a)](\lambda - 1)(\lambda^b - \lambda^{-a})}{[b(\lambda^{-a} - 1) + a(\lambda^b - 1)](\lambda + 1)(\lambda^{a+b} - 1)}$$

An optimization of the average profit over the parameters  $(a, b)$  when the underlying process is a historical process provides then an approach for selling and buying.

### Problem

Consider the average profit above and optimize this profit with respect to  $a$  and  $b$  and as a function of  $\lambda > 1$  and  $\lambda < 1$ .

### Problem

Show that when  $p > q$ , then the mean time and its variance for a random walk to attain the value  $b$  is equal to:

$$E(T_b) = b/(p - q) \quad \text{and} \quad \text{var}(T_b) = ([1 - (p - q)^2]b)/(p - q)^3$$

Finally, show that when the boundary  $b$  becomes large that the standardized stopping time tends to a standard Normal distribution.

### Example: Pricing a buy/sell strategy on a random walk\*

Consider again an underlying random walk price (where the framework of risk-neutral pricing might not be applicable)  $(S_t)$ ,  $t = 1, 2, \dots$ . The probability that the price increases is  $s$  while the probability that the stock price decreases is  $q$ . Assume that the current price is  $i_0$  and let  $i$  be a target selling price  $i = i_0 + \Delta i > i_0$ , given by the binomial probability distribution:

$$\begin{aligned} P(S_n = i = i_0 + \Delta i) &= \left( \binom{n}{\left[ \frac{\Delta i + n}{2} \right]} \right) s^{(n+\Delta i)/2} q^{(n-\Delta i)/2} \\ &= \binom{\Delta i + 2\nu}{\Delta i + \nu} s^{\Delta i + \nu} q^\nu; \quad n - \Delta i = 2\nu, \quad \nu = 0, 1, 2, 3, \dots \end{aligned}$$

where  $[\ ]$  denotes the least integer. Since, prices  $i$  can be reached only at even values of  $\Delta i + n$ , it is convenient to rewrite the price process by:

$$\begin{aligned} P(S_{\Delta i + 2\nu} = i) &= \binom{\Delta i + 2\nu}{\Delta i + \nu} s^{\Delta i + \nu} q^\nu = \frac{s^{\Delta i + \nu} q^\nu}{\nu!} \prod_{k=1}^{\nu} (\Delta i + \nu + k); \\ n - \Delta i &= 2\nu, \quad \nu = 0, 1, 2, 3, \dots \end{aligned}$$

The amount of time  $\tau(i) = n = 2\nu + \Delta i$  for an underlying process ( $s > q$ ) that can reach this price, however, is given by Feller (1957):

$$P(\tau(i) = 2\nu + \Delta i) = \frac{\Delta i}{2\nu + \Delta i} P(S_{2\nu + \Delta i} = i) = \frac{i}{(2\nu + i)} \frac{s^{i+\nu} q^\nu}{\nu!} \prod_{k=1}^{\nu} (i + \nu + k);$$

$$\nu = 0, 1, 2, 3, \dots \quad s > q$$

Thus, if a sell order for a stock is to be exercised at price  $i$  and if we use a risk-sensitive adjusted discount rate  $\rho$ , then the current expected value of this transaction is:

$$i_0 = E(\tilde{i}_0) = iE(\rho^{\tilde{\tau}(i)}) = i^2(\rho s)^i \sum_{\nu=0}^{\infty} \frac{\prod_{k=1}^{\nu} (i + \nu + k)}{(2\nu + i)\nu!} (\rho^2 s q)^\nu$$

Under risk-neutral pricing of course, the discount rate equals the risk free rate  $\rho_f$ , that is  $\rho = \rho_f = 1/(1 + R_f)$  and for convenience:

$$i_0 = i^2(\rho s)^i \Lambda_i(\rho); \quad \Lambda_i(\rho) = \sum_{\nu=0}^{\infty} \frac{\prod_{k=1}^{\nu} (i + \nu + k)}{(2\nu + i)\nu!} (\rho^2 s q)^\nu$$

We can also set  $\rho = \rho_f + \pi$  where  $\pi$  is the risk premium associated with selling at a price  $i$ . A price  $i$  can also be obtained as well by buying a bond of nominal value  $i$  in  $m$  periods hence without risk. In this case, we will have:

$$B_i(m) = i(\rho_f)^m \quad \text{or} \quad i = B_i(m)(\rho_f)^{-m}$$

Replacing  $i$ , we have

$$i_0 = [B_i(m)(\rho_f)^{-m}]^2 (\rho s)^{[B_i(m)(\rho_f)^{-m}]} \Lambda_{B_i(m)(\rho_f)^{-m}}(\rho)$$

This provides a solution for the risk-sensitive discount rate and therefore its risk premium. Higher-order moments can be calculated as well. Set:

$$i_0 = i^2 \Lambda(\rho)$$

$$\text{var}(\tilde{i}_0) = i^3 \Lambda(\rho^2) - i^4 \Lambda^2(\rho)$$

The probability distribution  $P(\tilde{i}_0 | i_0, \text{var}(\tilde{i}_0))$  of the current trade which is a function of the discount rate provides a risk specification for such a trade. If we have a quantile risk given by,  $P(\tilde{i}_0 \leq i_0 - \Delta i | i_0, \text{var}(\tilde{i}_0)) \leq \xi$ , then by inserting the mean variance parameters given above, and expressed in terms the discount rate  $\rho$ , we obtain an expression of the relationship between this discount rate and the VaR parameters ( $\xi, i_0 - \Delta i$ ).

Say that the problem is to sell or wait and let  $i^* > i_0$  be the optimal future selling price (assumed to exist of course and calculated according to some criteria as we shall see below). If price  $i^*$  is reached for the first time at time  $n$ , then the price of such a trade is  $i^* \rho^n$  which can be greater or smaller than the current price. If we sell now, then the probability that this decision is ill-taken is  $P(i^* \rho^{\tau(i^*)} \geq i_0)$ . As a result, the probability of a loss due to a future price increase has a quantile risk

$P(i^* \rho^{\tau(i^*)} - i_0 \geq V_s) \leq 1 - \theta_s$  where  $V_s$  is the value at risk for such a decision while  $1 - \theta_s$  is the assigned probability associated with the risk of holding the stock. By the same token, if we wait and do not sell the stock, the probability of having made the wrong decision is now:

$$P(i_0 - i^* \rho^{\tau(i^*)} \geq V_h) = P(i^* \rho^{\tau(i^*)} - i_0 \leq V_h) \geq \theta_s$$

where  $V_h$  denotes the value at risk of holding the stock. If a transaction cost is associated with a trade and if we set  $c$  to be this cost (when there are no holding costs), then we have:

$$P[(i^* - c) \rho^{\tau(i^*)} - (i_0 - c) \geq V_s] \leq 1 - \theta_s$$

or

$$P\{i^* \rho^{\tau(i^*)} - [i_0 - c(1 - \rho^{\tau(i^*)})] \geq V_s\} \leq 1 - \theta_s$$

Therefore, a transaction cost has the net effect of depreciating the current price by  $c(1 - \rho^{\tau(i^*)}) > 0$  and thereby favouring selling later rather than now in order to delay the cost of the transaction. In this sense, a transaction cost has the effect of reducing the number of trades! By the same token, an investor may seek to buy an asset believing that its current value is underpriced. In such a case, the buyer will compare the future discounted price with the current price and reach a decision accordingly. For example, the optimal buy price, based on the expected discounted future prices would be:

$$i_0^{**} = \text{Max}_{i > i_0} \{iE(\rho^{\tau(i)} - i_0 | i^{**} > i_0 + d), \tau(i) = \text{Inf} \{n, i > i_0, \}$$

where  $d$  is the buy transaction cost. An appropriate buy–sell–hold strategy is then defined by:

$$\begin{cases} \text{Do nothing} & i_0^* \leq i_0 \leq i_0^{**} \\ \text{Sell if} & i_0^* \leq i_0 \\ \text{Buy if} & i_0^{**} \geq i_0 \end{cases}$$

In this framework, the quantile risk approach provides in a simple and a uniform manner an approach to stopping as a function of the risks the investor is willing to sustain. In Chapter 10, this measure of risk will be considered in greater detail, however.

## 7.6 SPECIFIC APPLICATION AREAS

Foreign exchange is a fertile ground for the application of financial products, their pricing and their analyses. Basic transactions (through the interbank or the wholesale market) on spot, futures, forwards and swap and other products are applied extensively. FX trading is assuming greater importance. The Philadelphia Exchange trades, for example, options on the British pound, German mark, Japanese yen, Swiss franc and the Canadian dollar. The most heavily traded contracts are the Deutschemark and Japanese yen American-style

options. The strike price for each foreign currency option is the US dollar price of a unit of foreign exchange. The expiration dates correspond to the delivery dates in futures. Specifically, the expiration dates correspond to the Saturday before the third Wednesday of the contract month. Contract months are March, June, September and December plus the two near-term contracts. The daily volume of contracts traded on the Philadelphia exchange has steadily increased to over 40 000 contracts per day.

Currency	Contract size	Strike price intervals	Premium quotations
Mark	62 500	1.0	Cents
Sterling	31 250	2.5	Cents
Swiss franc	62 500	1.0	Cents
Yen	6 250 000	0.01	Hundredth cent

Consider the British pound for example. Each contract is for 31 250 British pounds. Newspapers report the closing spot price, in cents per pound sterling. The strike prices are reported in cents per pounds, at 2.5-cent intervals. The call and put premiums are also in cents per pound. Consider the theoretical price of a European call option on the British pound that trades at the Philadelphia Exchange. The time to expiration is 6 months. The spot is \$1.60 per pound. An at-the-money call is to be valued where the exchange rate volatility is 10% per year. The domestic interest rate and the foreign interest rate are both equal to 8%. Using the theoretical price of the European call option given below, we find the option price to be \$0.0433. The equations for this problem are similar to the Black–Scholes model, as we saw earlier and in the previous chapter, and are summarized below.

### Price of a European call option on foreign exchange

$$C^E(0) = G(0)N(d_1^*) - K e^{-rT} N(d_2^*); d_1^* = \frac{1}{\sigma\sqrt{T}} \ln[G(0)/K] \\ + (r + \sigma^2/2)T; d_2^* = d_1^* - \sigma\sqrt{T}$$

with  $\sigma$  the volatility of foreign exchange and  $G(0) = S(0)e^{-r_F T}$  where  $r_F$  is the risk-free rate in the foreign currency.

FX swap contracts are made by drafting purchase–repurchase agreements by selling simultaneously one currency (say DM) in the spot and the forward market. Swap contracts are immensely popular contracts and will be treated in the next chapter in the context of interest-rate-related contracts.

In practice, financial products can be tailored to sources of risks and can respond to specific business, industrial or other needs. For example, financial products that meet firms' risks related to climatic risks and energy supplies. Climatic factors in

particular account for a substantial part of insurance firms' costs. The December 1999 storm that hit France may have cost 44.5 billion francs! – hitting hard both the French insurance companies and reinsurance firms throughout Europe. Climatic risks also have an important effect on the US economy, accounting for approximately 20 % of GNP according to the Department of Energy. There are in fact few sectors that are immune to weather effects and thereby the importance of all risk-management activities related to meteorological forecasting, robust construction, tourism, fashion etc. Energy needs in particular, are determined by the intensity of summer heat and winter cold, generating fluctuations in demand and supply for energy sources. An expanding climatic volatility has only added to the management of these risks. For this reason, firms such as ENRON (now defunct), Koch, Aquila and Southern Energy have focused attention on the use of financial energy-related products that can protect sources of supplies and meet demands. As a result, since 1997, there have been energy products on the CME and, since 2000, on London's LIFFE, providing financial services to energy investors, speculators and firms. The underlying sources of risk of energy firms (as well as many supply contract) span:

- (a) Price risk whose effects are given by  $V \Delta p$  where  $V$  is the quantity of the energy commodity and  $\Delta p$  is the price change of the commodity.
- (b) Quantity risk,  $(\Delta V)p$ .
- (c) Correlation risk  $\Delta V \Delta p$ .

To manage these risks, derivative products on both the price and the supply contract are used. These contracts, over multiple sources of risks, are difficult to assess and are currently the subject of extensive research and applications.

## 7.7 OPTION MISSES

In the mid-nineties, media and regulators' attention was focused on option misses because of huge derivative-based losses that have affected significantly both large corporate firms and institutions. The belief that options are primarily instruments for hedging was severely shaken and the complexity and risks implied in trading with derivatives revealed. Management and boards, certain that derivatives were used only to hedge and reduce price risk, were astounded by the consequences of positions taken in the futures and options markets – for the better and for the worse. According to the *Wall Street Journal* for 12 April, 1996, J.P. Morgan earnings in the first quarter jumped by 72 % from the previous year, helped by an unexpectedly strong derivatives business that more than doubled the bank's overall trading revenue! By the same token, firms were driven to bankruptcy due to derivative losses. Business managers also discovered that managing risk with derivatives can be tempting, often only understood by a few mathematically inclined financial academics. At the same time there is a profusion of derivatives contracts, having a broad set of characteristics and responding in different ways to the many factors that beset firms and individuals, which have invaded financial

**Table 7.2** Derivatives losses of industries and organizations.

Name	Losses (\$ million)	Main cause
AIG	90	Derivatives revaluation
Air Product	113	Leveraged and currency swaps
Arco (Pension Fund)	25	Structured notes
Askine Securities	605	MBS model
Bank of America	68	Fraud
Procter & Gamble	450	Leveraged and currency swaps
Barings PLC	1400	Futures trading
Barnet Banks	100	Leveraged swaps
Cargil	100	Mortgage derivatives
Codelco (Chile)	210	Futures trading—copper
Community Bankers	20	Leveraged swaps
Dell Computers	35	Leveraged swaps
Gestetner	10	Leveraged swaps
Glaxo	200	Derivatives and swaps
Harris Trust	52	Mortgage derivatives
Kashima Oil	1450	Currency derivatives
Kidder Peabody	350	Fraud trading
Mead	12	Leveraged swaps
Metalgesellschaft	1300	Futures trading
Granite Partners	600	Leveraged CMOs
Nippon Steel	30	Currency derivatives
Orange County	1700	Mortgage derivatives
Pacific Horizon	70	Structured notes
Piper Jaffray	700	Leveraged CMOs
Sandoz	80	Derivatives transactions
Showa/Shell Sekiyu	1400	Forward contracts
Salomon Brothers	1000	Fraud (cornering)
United Services	95	Leveraged swaps
Estimated losses	12 265	

markets. In many cases too, derivatives hype has ignited investors' imagination, for they provided a response to many practical and real problems hitherto not dealt with. An opportunity to manage risks, enhance yields, delay debt records to some future time, exploit arbitrage opportunities, provide corporate liquidity, leverage portfolios and do whatever might be needed are just a very few such instances.

A derivative mania has generated at the same time their misuse, leading to large losses, as many companies and individuals have experienced. Derivatives became the culprit for many losses, even if derivatives could not intelligently be blamed. For example, the continuous increase in interest rates during 1993–94 pushed down T-bond prices so that the market lost hundreds of billions in US dollars. Additionally, the sharp drop in the IBM stock price in 1992–94 from 175 to 50 created a market loss of approximately \$70 billion for one firm only! These losses could not surely be blamed on derivatives trading! Table 7.2

summarizes some of the losses, assembled from various sources and sustained by corporate America (Meir Amikam, 1996). These derivative losses were found to be due to a number of reasons including: (1) a failure to understand and identify firms' sensitivities to different types of risk and calculating risk exposure; (2) over-trusting – trusting traders with strong personality led to huge losses in Baring's, megalomania in Orange County, Gestetner etc.; (3) miscalculation of risks – overly large positions undertaken which turned out sour; (4) information asymmetry – the lack of internal control systems and audits of trading activities has led traders to assume unreasonable positions in hope; (5) poor technology – lack of computer-aided tools to follow up trading activity for example; (6) applying real-time trading techniques, responding to volatility rather than to fundamental economic analysis, that have also contributed to ignoring risks.

Of course, a number of risk management tools and models have been suggested to institutional and individual investors and traders to prevent these risks wherever possible. For example, value at risk, extremes loss distributions, mark to market and using the Delta of model-based risk, to be considered in Chapter 10, have been found useful. These models have their own limitations, however, and cannot replace the expectations of qualified professional judgement. By the same token, these expectations introduce a *systematic risk* that can lead to unexpected volatility and cause severe losses, not only to speculators, but also to hedgers. Some of the great failures in 1993–94, for example, were incurred because users were caught by a surprise interest rate hike. A similar scenario occurred when the price of oil dropped. Thus, the use of derivatives for speculation purposes can cause large losses if traders turn out to be on the wrong side of their market expectations.

There are some resounding losses, however, that have been the subject of intense scrutiny. Below we outline a few such losses.

*Bankers Trust/Procter & Gamble/Gibson Greeting*: 'Our policy calls for plain vanilla type swaps', Erick Nelson, CFO, Procter & Gamble.

*Procter & Gamble* incurred a loss of \$157 million loss from interest swaps in both US and German markets, swapping fixed for floating rates. In effect they had a put option given to Bankers Trust. P&G's strategic error was the belief that exchange rates would continue to fall both in the USA and in Germany. Swaps were thus made for the purpose of reducing interest costs. The actual state of interest rates turned out to be quite different, however. In the USA, the expectation of lower interest rates meant that the value of bonds would increase, rendering the put option worthless to Bankers Trust. In fact interest rates did not fall and therefore, Treasury bonds increased in value, forcing P&G to purchase the bonds from Bankers Trust at a higher price. This resulted in a first substantial loss. By the same token, the expectations of a decline in interest rates in Germany meant that the German Bund would decline rendering again the put option worthless to Bankers Trust. Therefore P&G would have lower interest costs as well. Rates increased instead, and therefore, the value of the German Bund decreased, forcing thereby a purchase of bonds from Banker's Trust at higher prices than the current market price – inducing again a loss. The combined losses reached \$157 million.

In other words, P&G took an interest-rate gamble instead of protecting itself in case the 'bet' turned out to be wrong.

In April 1994, Procter & Gamble and Gibson Greetings claimed that Bankers Trust, had sold them high-risk, leveraged derivatives. The companies claim that those instruments, on which profits and losses can multiply sharply in certain circumstances, had been bought without giving the companies adequate warning of the potential risks. Bankers Trust countered that the firms were trying to escape loss-making contracts. P&G sued Bankers Trust in October 1994, and again in February 1995, for additional damages on the leveraged interest-rate swap tied to the yields on Treasury bonds. P&G has also claimed damages on a leveraged swap tied to DM interest rates for \$195 million, insisting that the status of the first swap was not fully and accurately disclosed. The case was settled out of court in January 1995. It seemed in the course of the court deliberations that Bankers Trust may have not accurately disclosed losses and Bankers Trust may have hoped that market movements would turn to match their positions, so they did not notify Gibson immediately of the true magnitude of the intrinsic risk of the derivatives bought. Had Gibson been aware of the risk, the losses could have been minimized. In response the bank has fired one manager, reassigned others, and shacked up the leveraged derivatives unit. It also agreed to pay a \$10 million fine to regulators and entered into a written agreement with the New York Federal Reserve Bank that allows regulators unprecedented oversight of the bank's leveraged derivatives business. The agreement is open-ended and highly embarrassed Bankers Trust. In addition, the lawsuits have tarnished the reputation of Bankers Trust. Outsiders have wondered if it was more the deal-making culture that was to blame, rather than the official tale of a few rogue employees. Other banks, such as Merrill Lynch, First Boston and J.P. Morgan have also run into trouble selling leveraged derivatives and other high-risk financial instruments.

The *Orange County (California)* case was a cause célèbre amplified by the media and regulation agencies warning of the dangers of financial markets speculations by public authorities. The Orange County strategy is called 'On the street, a kind of leveraged reverse repo strategy, also coined the *death spiral* because one significant market move can blow down the strategy in one puff'. This can occur if managers fail to understand properly the firm's sensitivities to different sorts of risks or do not regard financial risks as an integral part of the institution or corporate strategy. Speculation by the Orange County treasurer, who initially generated huge profits, ultimately bankrupted it. As the treasurer supervisor stated: 'This is a person who has gotten us all millions of dollars. I don't care how the hell he does it, but it makes us all look good.'

The county lost \$2.5 billion. Out of a \$7.8 billion portfolio! The loss was faulted on borrowing short to invest long in risk-structured notes. In other words, the county treasurer leveraged the (public) portfolio by borrowing \$2 for each dollar in the portfolio, equivalent to investing on margin. The county then used the repo (reverse purchase agreement) market to borrow short in order to purchase

long (term) government bonds. In the repo agreement, the county pledged the long-term bonds it was purchasing as collateral for secured loans. These loans were then rolled over every three to six months. As interest rates began to rise, the cost of borrowing increased while the value of long-term bonds decreased. This situation resulted in a substantial loss. In effect, Orange County was betting that interest rates would remain low or even decrease some and that spreads between long- and short-term rates would remain high. Something else happened. When interest rates rose, the cost of short-term borrowing increased, the value of the long-term bonds purchased decreased, the rates on the inverse floaters (consisting for an initial period of a fixed rate and then of a variable rate) fell and the rates on the spread bonds (consisting of a fixed percentage plus a long-term interest minus a short-term rate) fell as the yield curve flattened. This generated a huge loss for Orange County.

*Metalgesellschaft*: 'Pride in integrity takes a blow' (Cooke and Cramb, *Financial Times*). Metalgesellschaft, a \$15 billion sales commodity and engineering conglomerate, blamed its near collapse on reckless speculation in energy derivatives by its New York subsidiary. To save the firm that employed 46 000 people, banks and shareholders provided a \$2.1 billion bail-out. The subsidiary MGRM (MG Refining & Marketing) negotiated long-term, fixed-price contracts to sell fuel to gas stations and other small businesses in 1992. The fixed price was slightly higher than the prevailing spot price. To lock in this profit and hedge against rising fuel prices, the company hedged itself by buying futures on the New York Mercantile Exchange (NYMEX). Maintaining such a 'stacked rollover hedge' when prices are falling could require large amounts of liquidity. To hedge the long-term contracts, MGRM was obligated to buy short-term futures contracts to cover its delivery commitment since matching its supply obligations with contracts of the same maturity was impossible. That strategy was based on rolling over the short-term futures just before they expired. The hedging depended on the assumption that oil markets, which were in backwardation over two-thirds of the time over the past decade, would remain in that state for most of the time. By entering into futures contracts, MGRM would be able to hedge their short positions in the forward sales contracts. This assumption was predicated on the fact that, as expiration approaches, the future price MGRM paid for the contracts would be less than the spot price. That trading and hedging strategy had some inherent risks that happened to be crucial:

- *Market fluctuations risk*: oil prices, futures and spot price, that did not meet expectations.
- *Proper hedge risk*: Resulting from mismatched timing of contracts and entering into a speculative hedge.
- *Funding risk*: futures contracts require marking to market; this margin call caused by futures losses is not offset by forward contract gains, which are unrealized until delivery.

By September 1993, MGRM's obligation was equivalent to 160 million barrels. In November 1993, oil prices dropped by \$5 to \$14.5 as a reaction to OPEC's decision to cut production. That drop wiped out 20 % of MGRM short-term futures contracts and led to a cumulative trading loss of \$660 million. German GAAP does not allow the offset of gains or losses on hedging positions using futures against corresponding gains or losses on the underlying hedge asset, however. Further, Deutsche Bank, advising MG and apprehensive about the short-term losses as they mounted, convinced MG to close out its positions, thus causing the loss. The risk MG was taking was using a short-term instrument hedging strategy for a long-term exposure, creating thereby a mismatch that could be considered a bet that turned out to be wrong. The paper loss converted into a real heavy loss as the futures positions were closed. In 1994 the group announced that the potential losses of unwinding its positions could bring the total losses up to \$1.9 billion over the next three years. It is argued, though, among academics that the strategy could have worked had MGRM not unwound their futures position. Thus, liquidity was essential for this strategy to work. Alternatively MGRM could have reversed its futures position when oil prices dropped. It is important to look at current market conditions in reassessing the merits of one's strategy. MGRM traders did not (and, in fact, could not) properly hedge. They were speculating on the correlation between the underlying and the cash market. They ignored the risks of a speculative hedge, trusting that they could predict the relationship and changes in prices from month to month.

*Barings*: 'Ultimately, if you want to cover something up, it's not that difficult . . . Derivative positions change all the time and balance sheets don't give a proper picture of what's going on. For anyone on the outside to keep track is virtually impossible' (SIMEX trader, quoted by the *Financial Times*, February 1995).

The much-publicized *Barings* loss of \$1.3 billion was incurred by its branch in Singapore. It was incurred in three weeks by trading on the Nikkei Index. Leeson the Singapore Office Head of Trading was speculating that the Nikkei Index would rally after the Kobe earthquake, so he amassed a \$27 billion long position in Nikkei Index futures. The Nikkei Index fell, however, and Leeson was forced to sell put and call options to cover the margin calls. In an effort to recoup losses, Leeson increased the size of his exposure and held 61 039 long contracts on the Nikkei 225 and 26 000 short contracts on Japanese bonds. When he decided to flee, the Nikkei dropped to 17 885. Leeson was betting that the Index would trade in a range and he would therefore earn the premium from the contracts (to pay the margins). No one was aware of such trades and the risk exposure it created for *Barings* (as a result, it generated a much-needed and heated discussion regarding the needs for controls. The main office 'seemed' to focus far more on the potential gains rather than on the potential losses! This loss induced the demise of *Barings*, a venerable and longstanding English institution, which was sold to ABN AMRO.

Initially Leeson was responsible for settlement. In a short time he turned out to be a successful trader whose main job was to arbitrage variations among the prices of futures and options on the Nikkei 225, having the unique advantage

that Barings had seats both on SIMEX and on OSE (Osaka Stock Exchange). Contracts on the Nikkei 225 and Nikkei 300 were OSE's only futures and options and accounted for 30 % of SIMEX business. As a member he enjoyed the privilege of seeing the orders ahead of non-members and of taking suitable positions with low risk. His strategy was mainly based on small spreads in which he invested large amount of money. Later on he was promoted to be responsible for trading and for settlement.

*Granite partners:* Granite Partners lost \$600 million in mortgage derivatives in the mid-nineties. Fund managers promised their investors little risk in their investment policy since they used derivatives mainly for hedging purposes. By using CMO derivatives they expected to take advantage of market movements. But the disclosure emphasized that Granite had the option to wait, if need be, until market conditions suited the funds' position. Leveraging with CMO derivatives was much more than what was promised. The portfolio was leveraged based on the assumption of some in-house models. To their detriment, the bond market took a direction that went against the funds' positions. Since the portfolio was highly leveraged, the losses grew tremendously and Granite was shut down.

*Freddie Mac:* In January 2003, PriceWaterhouseCoopers, Freddie Mac's auditor for less than a year, revealed that the company might have misreported some of its derivatives trades. As a result, Freddie Mac later said that some earnings that should have been reported in 2001 and 2002 were improperly shifted into the future. It is not clear that the top executives were not attempting to distort the company's books. But recent corporate crises suggest that if someone wants to hide something, derivatives can help (*New York Times*, 12 June 2003).

Lessons from these loss cases, as well as many others, are summarized in Table 7.3. Generally, the most common cause was speculation – the market moving in directions other than presumed, the trade strategy collapsed – causing unexpected losses. There is little information relating to internal and external audit and control, however, implying perhaps that in most cases management does not realize the risk exposures they take on and thus controls end up being very poor. There were no reports of written policies that were supposed to limit positions and losses. Had there been such policies, traders could have easily ignored them, trusting their strategic 'cunning and assessments'. Although these figures were assembled from various sources, including the media, and the actual reasons for such losses were varied, a distribution of the main causes for the losses were: management, 18; poor audit or no controls, 20; wrong methods and trade strategies applied, 21; market fluctuation (poor forecasts), 17; and, finally, frauds and traders' megalomania, 5. Problems associated with audit and controls are resurging in various contexts today. For example, in the wake of multibillion-dollar accounting scandals, companies are under intense pressure to make sure that their financial results do not paint a misleading, rosy picture. Insurance firms for example, are swamped with billions of dollars in corporate bonds that they bought years ago and that are still maintained at their original value in their

**Table 7.3** Main reasons which led to losses.

Firm	Management	Audit/ control	Methods/ strategy	Market fluctuations	Frauds/ megalomania
AIG		+	+	+	
Air Product	+	+	+		
Arco (Pension Fund)	+	+		+	
Askine Securities			+	+	
Bank of America		+	+		
Bankers Trust/ Procter and Gamble			+	+	+
Barings PLC	+	+	+		+
Barnnet Banks		+	+		
Cargil	+	+	+	+	
Codelco (Chile)	+	+	+	+	
Community Bankers	+	+	+	+	
Dell Computers	+	+	+		
Gestetner	+	+	+		+
Glaxo	+	+	+	+	
Harris Trust	+		+		
Kashima Oil		+	+	+	
Kidder Peabody	+	+	+	+	
Mead		+	+		
Metalgesellschaft	+	+		+	
Granite Partners			+	+	
Nippon Steel	+	+			
Orange County	+	+		+	+
Pacific Horizon	+	+		+	
Piper Jaffray	+		+	+	
Sandoz			+	+	
Showa/Shell Sekiyu	+	+			
Salomon Brothers				+	+
United Services		+	+		

books! Now, financial regulators are suggesting that they should be accounted at their true value, which could lead many insurance firms to the brink, or at the least to reporting huge losses and to borrowing large amounts of money to meet their capital requirements (International Herald Tribune, 17 June 2003, Business Section).

## REFERENCES AND ADDITIONAL READING

- Albizzati, M.O., and H. Geman (1994) Interest rate risk management and valuation of surrender option in life insurance policies, *Journal of Risk and Insurance*, **61**, 616–637.
- Amikam, H., (1996), Private Communication.
- Arndt, K. (1980) Asymptotic properties of the distribution of the supremum of a random walk on a Markov chain, *Probability Theory and Applications*, **46**, 139–159.

- Barone-Adesi, G., and R.E. Whaley (1987) Efficient analytical approximation of American option values, *Journal of Finance*, **42**, 301–320.
- Barone-Adesi, G., W. Allegretto and R. Elliott (1995) Numerical evaluation of the critical price and American options, *The European Journal of Finance*, **1**, 69–78.
- Basak, S., and A. Shapiro (2001) Value at risk based risk management: Optimal policies and asset prices, 2001, *Review of Financial Studies*, **14**, 371–405.
- Beibel, M., and H.R. Lerche (1997) A new look at warrant pricing and related optimal stopping problems. Empirical Bayes, sequential analysis and related topics in statistics and probability, *Statistica Sinica*, **7**, 93–108.
- Benninga, S. (1989) *Numerical Methods in Finance*, MIT Press, Cambridge, MA.
- Boyle, P. (1977) Options: A Monte Carlo approach, *Journal of Financial Economics*, **4**, 323–338.
- Boyle, P., and Y. Tse (1990) An algorithm for computing values of options on the maximum or minimum of several assets, *Journal of Financial and Quantitative Analysis*, **25**, 215–227.
- Brennan, M., and E. Schwartz (1977) The valuation of American put options, *Journal of Finance*, **32**, 449–462.
- Capocelli, R.M., and L.M. Ricciardi (1972) On the inverse of the first passage time probability problem, *Journal of Applied Probability*, **9**, 270–287.
- Caraux, G., and O. Gascuel (1992) Bounds on distribution functions of order statistics for dependent variates, *Statistical Letters*, **14**, 103–105.
- Carr, P., R. Jarrow and R. Myer (1992) Alternative characterizations of American put options, *Journal of Mathematical Finance*, **2**, 87–106.
- Cho, H., and K. Lee (1995) An extension of the three jump process models for contingent claim valuation, *Journal of Derivatives*, **3**, 102–108.
- Chow, Y.S., H. Robbins and D. Siegmund (1971) *The Theory of Optimal Stopping*, Dover Publications, New York.
- Chow, Y.S., H. Robbins and D. Siegmund (1971) *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston, MA.
- Coffman, E.G., P. Flajolet, L. Flatto and M. Hofri (1997) The max of a random walk and its application to rectangle packing, Research Report, INRIA (France), July.
- Connoly, K.B. (1977) *Buying and Selling Volatility*, John Wiley & Sons, Inc., New York.
- Cox, D.R., and H.D. Miller (1965) *The Theory of Stochastic Processes*, Chapman & Hall, London.
- Darling, D.A., and A.J.F. Siebert (1953) The first passage time for a continuous Markov process, *Annals of Math. Stat.*, **24**, 624–639.
- Duffie, D., and H.R. Richardson (1991) Mean-variance hedging in continuous time, *Annals of Applied Probability*, **1**, 1–15.
- Durbin, J. (1992) The first passage time of the Brownian motion process to a curved boundary, *Journal of Applied Probability*, **29**, 291–304.
- Embrechts, P., C. Kluppelberg and T. Mikosch (1997) *Modelling Extremal Events*, Springer Verlag, Berlin & New York.
- Feller, W. (1957) *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, Inc., New York.
- Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, Inc., New York.
- Garman, M.B., and S.W. Kohlhagen (1983) Foreign currencies option values, *Journal of International Money and Finance*, **2**, 231–237.
- Gerber, H.U., and E.S.W. Shiu (1994a) Martingale approach to pricing perpetual American options, *ASTIN Bulletin*, **24**, 195–220.
- Gerber, H.U., and E.S.W. Shiu (1994b) Pricing financial contracts with indexed homogeneous payoff, *Bulletin of the Swiss Association of Actuaries*, **94**, 143–166.
- Gerber, H.U., and E.S.W. Shiu (1996) Martingale approach to pricing perpetual American options on two stocks, *Mathematical Finance*, **6**, 303–322.
- Gerber, H.U., and Shiu, E.S.W. (1996). Actuarial bridges to dynamic hedging and option pricing, *Insurance: Mathematics and Economics*, **18**, 183–218.

- Geske, R., (1979) The valuation of compound options, *Journal of Financial Economics*, **7**, 63–81.
- Geske, R., and H.E. Johnson (1984) The American put option valued analytically, *Journal of Finance*, **39**, 1511–1524.
- Geske, R., and K. Shastri (1985) Valuation by approximation: A comparison of alternative option valuation techniques, *Journal of Financial and Quantitative Analysis*, **20**, 45–71.
- Goldman, M.B., H. Sosin and M. Gatto (1979) Path-dependent options: buy at the low, sell at the high, *Journal of Finance*, **34**, 1111–1128.
- Graversen, S.E., G. Peskir and A.N. Shiryaev (2001) Stopping Brownian motion without anticipation as close as possible to its ultimate maximum, *Theory Probability and Applications*, **45**(1), 41–50.
- He, H. (1990) Convergence from discrete to continuous time contingent claim prices, *The Review of Financial Studies*, **3**, 523–546.
- Hull, J., and A. White (1990) Valuing derivative securities using the explicit finite difference method, *Journal of Financial and Quantitative Analysis*, **25**, 87–100.
- Jacka, S.D. (1991) Optimal stopping and the American put, *Journal of Mathematical Finance*, **1**, 1–14.
- Johnson, N.L., and S. Kotz (1969) *Discrete Distributions*, Houghton Mifflin, New York.
- Johnson, N.L., and S. Kotz (1970a) *Continuous Univariate Distributions – 1*, Houghton Mifflin, New York.
- Johnson, N.L., and S. Kotz (1970b) *Continuous Univariate Distributions – 2*, Houghton Mifflin, New York.
- Karatzas, I. (1989) Optimization problems in the theory of continuous trading, *SIAM Journal on Control and Optimization*, **27**, 1221–1259.
- Kijima, M. and M. Ohnishi (1999) Stochastic orders and their applications to financial optimization, *Mathematical Methods of Operation Research*, **50**, 351–372.
- Kim, I.J. and G. Yu (1996) An alternative approach to the valuation of American options and applications, *Review of Derivative Research*, **1**, 61–85.
- Korczak, J., and P. Roger (2002) Stock timing and genetic algorithms, *Applied Stochastic Models in Business and Industry*, **18**, 121–134.
- Korshunov, D.A. (1997) On distribution tail of the maximum of a random walk, *Stochastic Processes and Applications*, **72**(1), 97–103.
- Korshunov, D.A. (2001) Large-deviation probabilities for maxima of sums of independent random variables with negative mean and subexponential distribution. *Theory Probab. Appl.*, **46**(2) 387–397. (In Russian.)
- Lamberton, D. (2002) Brownian optimal stopping and random walks, *Applied Mathematics and Optimization*, **45**, 283–324.
- Leadbetter, M.R., G. Lindgren and H. Rootzen (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York.
- Murphy, J. (1998) *Technical Analysis of the Financial Markets*, New York Institute of Finance, New York.
- Pekšir, G. (1998) Optimal stopping of the maximum process: The maximality principle, *Annal. Probab.*, **26**, 1614–1640.
- Révész, Pal (1994) *Random Walk in Random and Non-Random Environments*, World Scientific, Singapore.
- Ritchken, P., and R. Trevor (1999) Pricing options under generalized GARCH and stochastic volatility process, *Journal of Finance*, **54**, 377–402.
- Rychlik, T. (1992) Stochastically extremal distribution order statistics for dependent samples, *Statistical Probability Letters*, **13**, 337–341.
- Rychlik, T. (2001) Mean-variance bounds for order statistics from dependent DFR, IFR, DFRA and IFRA samples, *Journal of Statistical Planning and Inference*, **92**, 21–38.
- Schweizer, M. (1995) Variance-optimal hedging in discrete time, *Mathematics of Operations Research*, February (1), 1–32.
- Shaked, M., and J.G. Shantikumar (1994) *Stochastic Orders and their Applications*, Academic Press, San Diego, CA.

- Shepp, L.A., and A.N. Shiryaev (1993) The Russian option: Reduced regret, *Annals of Applied Probability*, **3**, 631–640.
- Shepp, L. A., and A.N. Shiryaev (1994) A new look at the Russian option, *Theory Prob. Appl.*, **39**, 103–119.
- Shirayayev, A.N. (1978) *Optimal Stopping Rules*, Springer-Verlag, New York.
- Shiryaev, A.N. (1999) *Essentials of Stochastic Finance*, World Scientific, Singapore.
- Tapiero, C.S. (1977) *Managerial Planning: An Optimum and Stochastic Control Approach*, Gordon & Breach, New York.
- Tapiero, C.S. (1988) *Applied Stochastic Models and Control in Management*, North Holland, New York.
- Tapiero, C.S. (1996) *The Management of Quality and its Control*, Chapman & Hall, London.
- Wilmott P., (2000) Paul Wilmott on Quantitative Finance, John Wiley & Sons Ltd., Chichester.
- Zhang, Q. (2001) Stock trading and optimal selling rule, *SIAM Journal on Control*, **40**(1), 64–87.

### APPENDIX: FIRST PASSAGE TIME\*

A first time to some state, say  $S$  (a given stock price, an exercise option price, a given interest rate level and so on), may be defined by:

$$T(x_0) = \text{Inf} \{t > 0; x_0 = x_0, x(t) \geq S\}$$

where  $x_0$  is the initial state (at time  $t = 0$ ). The ‘target state’ can be thought of as an *absorbing state*. Let  $f(x, t)$  be the probability of state  $x$  at time  $t$  of a Markov process. Thus, the probability that the passage time exceeds the current time is:

$$\text{Pr} \{T(x_0) > t\} = \int_{-\infty}^S f(x, t/x_0) dx$$

As a result, the passage time probability can be written by deriving: the  $\text{Pr}\{T(x_0) \leq t\} = 1 - \text{Pr}\{T(x_0) > t\}$ , leading to the distribution function  $g(S, t/x_0)$ ,  $0 \leq t < \infty$ :

$$g(t) = -\frac{\partial}{\partial t} \int_{-\infty}^S f(x, t/x_0) dx$$

with the additional (existence) conditions:

$$g(S, t/x_0) \geq 0, \forall S, t, x_0; 0 < \int_0^{\infty} g(S, t/x_0) dt \leq 1, \forall S, t, x_0;$$

$$\times \lim_{x_0 \rightarrow S} g(S, t/x_0) = \delta(t)$$

Of course, if the probability distribution  $f(\cdot, \cdot)$  can be found analytically, then the stopping time distribution can be calculated explicitly in some cases. An example to this effect is considered below, which clearly points out to some mathematical

difficulties when analytical solutions are sought. Consider a forward Kolmogorov (Fokker–Plank) equation (corresponding to the stochastic differential equation with drift  $b(x)$  and diffusion  $a(x)$ ):

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}[b(x)f] + \frac{\partial^2}{\partial x^2}[a(x)f]$$

which we write for convenience by the operator:

$$\frac{\partial f}{\partial t} = \mathbf{L}f, \quad \mathbf{L} = -\frac{\partial}{\partial x}[b(x)f] + \frac{\partial^2}{\partial x^2}[a(x)f]$$

Using the fact that state  $S$  is absorbing, an expectation of the passage time can be obtained by defining a simpler differential operator (expressed as a function of the initial condition  $x_0$  and not of time and as we can see by the application of Ito's differential rule). That is to say, the Laplace transform of the passage time distribution, defined in the terms of the initial state and the target (absorbing) state, is defined by:

$$g_\lambda^*(S, x_0) = \int_0^\infty e^{-\lambda t} g(S, t; x_0) dt, \quad 0 < g_0^*(S, x_0) \leq 1, \quad \lim_{x_0 \rightarrow S} g_\lambda^*(S, x_0) = 1$$

An application of Ito's differential rule yields the second-order differential equation:

$$a(x_0) \frac{d^2 g_\lambda^*}{dx_0^2} + b(x_0) \frac{dg_\lambda^*}{dx_0} - \lambda g_\lambda^* = 0$$

which we write in terms of an adjoint operator  $\mathbf{L}^+$  by:

$$\mathbf{L}^+ g_\lambda = \lambda g_\lambda, \quad \mathbf{L}^+ = \frac{\partial}{\partial x}[b(x_0)f] + \frac{\partial^2}{\partial x^2}[a(x_0)f]$$

If  $\lambda > 0$ , then the solution for  $g_\lambda$  is necessarily bounded and is the Laplace transform of a passage time distribution for an Ito stochastic differential equation which is given by:

$$dx = b(x) dt + a(x) dw, \quad x(0) = x_0$$

For  $(a, b)$  constants, we have as a special case:

$$g_\lambda^*(S, x_0) = \exp \left[ \frac{x_0 - S}{a^2} \left( -b + \sqrt{b^2 + 2\lambda b a^2} \right) \right], \quad a > 0, \quad -\infty < x_0 \leq S < \infty$$

whose inverse transform yields the inverse Gaussian distribution:

$$g(S, t; x_0) = \frac{(x_0 - S)}{\sqrt{2\pi a^2 t^3}} \exp \left[ -\frac{(S - x_0 - bt)^2}{2a^2 t} \right]$$

In other words, if the decision is to sell a stock at a price  $S$ , then the probability distribution of the time at which the stock is sold is given by  $g(S, t; x_0)$ . The current

discounted value of such a policy, however, is given by:  $V(S) = E(S e^{-R\tau})$  where  $E(e^{-R\tau})$  is the stopping time Laplace transform with the risk-free rate replacing the transform's variable. As a result, we have:

$$V(S) = S \exp \left[ \frac{x_0 - S}{a^2} \left( -b + \sqrt{b^2 + 2R_f b a^2} \right) \right]$$

For a study of first passage time problems the reader should refer to Darling and Siegert (1953) as well as Capocelli and Ricciardi (1972) who provides the first passage time distribution for a lognormal process as well.

## CHAPTER 8

# Fixed Income, Bonds and Interest Rates

### 8.1 BONDS AND YIELD CURVE MATHEMATICS

*Bonds* are binding obligations by a bond issuer to pay the holder of the bond pre-agreed amounts of money at future and given dates. Thus, unlike stocks, bonds have payouts of known quantities and at known dates. Bonds are important instruments that make it possible for governments and firms to raise funds now against future payments. They are considered mostly safe investments, although they can be subject to default and their dependence on interest rates affect their price. As a result, although the nominal values of bonds are known, their price is derived from underlying interest rates. There are as well many types of bonds, designed to meet investors' needs, firms and governments' needs and payment potential when raising capital and funds. For example, there are zero-coupon bonds, coupon-bearing bonds paid at discrete irregular and regular time intervals, there are floating rate bonds, fixed rate bonds, repos (involving a repurchase agreement at some future date and at an agreed-on price). There are also strips bonds (meaning Separate Trading of Registered Interest and Principal of Securities) in which the coupon and the principal of normal bonds are split up, creating an artificial zero-coupon bond of longer maturity. There are options on bonds, bonds with call provisions allowing their recall prior to redemption etc.

Bond values express investors' 'impatience' measured by the rate of interest (discount) used in determining their value. When a bond is totally risk-free, the risk-free rate (usually the Treasury Bills rate of the US Government) is used. When a bond is also subject to various sources of uncertainties (due to interest-rate processes, due to defaults, inflation etc.) then a risk-sensitive discount rate will be applied reflecting an attitude toward these uncertainties and interactions between 'impatience and risk'. We shall see later on that these 'risk-sensitive discount rates' can also be determined in terms of the ongoing risk-free rates and the rates term (of payments and time) structure.

Bond market sizes and trades dwarf all other financial markets and provide therefore a most important source and fundamental information for the valuation

of financial assets in general and the economic health of nations and firms. Rated bonds made by financial agencies, such as Standard and Poors, Moody's and their like, are closely watched indicators that have a most important impact on both firm's equity value and governments' liquidity.

In this chapter, essential elements of bond valuation and bond-derived contracts will be elaborated. Further we shall also provide an introduction to interest-rate modelling which is equivalent to bond modelling for one reflects the other and vice versa. It is also a topic of immense economic practical and research interest. Irving Fisher in his work on interest (1906, 1907, 1930) gave the first modern insight into the market interest rate as a balance between agents' impatience (and attitude towards time) and the productivity (returns) of capital (investments). These studies were performed in the spirit of a general equilibrium theory whose foundations were posed by Walras in his *Elements d'Economie Politique Pure* in 1874. Subsequent economic studies (Arrow, 1953) have introduced uncertainty in equilibrium theory based on this approach. A concise review can be found for example in Magill and Quinzzi (1996). Subsequent studies have formalized both the theory of interest rates and its relation to time (the term structure of interest rates) as well as model the exogenous and endogenous sources of uncertainty in interest rates evolution. These studies are of course of paramount importance and interest for bond pricing, whether they are risk-free or default-prone (as it is the case for some corporate bonds). When a bond is risk-free then of course we use the risk-free rate associated with the time of payment. However, since interest rates may vary over time, the bond 'productivity or yield' may shift in various ways (according to the uncertain evolution of interest rates as well as the demand by borrowers and the supply by lenders), which renders the risk-free rate time-varying. When bonds are subject to default of various types, risks are compounded, affecting thereby the discount rate applied to the payment of bonds (and thus the bond price). In this sense, the study and the valuation of bonds is imbued with uncertainty and the risks it generates.

In this chapter we introduce some basic notions for the valuation of bonds. We consider rated bonds, with and without default, with reliable and unreliable rating, for which a number of results and examples will be treated. These results are kept simple except in some cases where over-simplification can hide some important aspects in bond valuation. In such cases we 'star' the appropriate section. In addition a number of results regarding options on bonds, the use of bonds to value the cash flows of corporate rated firms (such as computing 'net present values' of investment projects by such firms etc.) are derived.

Bond markets are, as stated above, both extremely large and active. By far, the most-traded bonds are Treasury bills. These are zero-coupon bonds with a maturity of less than one year. Treasury bills are issued in increments of \$5 000 above a minimum amount of \$10 000. In economic journals, T-bills are quoted by their maturity, followed by a price expressed by the bank discount yield. Below some simple examples are treated to appreciate both the simplicity and the complexity of bond valuation. At the same time, we shall elaborate on a broad number of transactions that can be valued using the bond terminology.

**8.1.1 The zero-coupon, default-free bond**

A zero coupon bond consists in an obligation to pay at a given future date  $T$  (the maturity date), a certain amount of money (the bond nominal value). For simplicity, let this amount be \$1. The price of such a bond at a given time  $t$ ,  $B(t, T)$ , denotes the current price of a dollar payment at time  $T$ . The value of this bond is essentially a function of:

- (i) the time to payment or  $\tau = T - t$  and
- (ii) the discount factor used at  $t$  for a payment at  $T$  or, equivalently, it is expressed in terms of the bond yield, denoted by  $y(t, T)$ .

In other words, the value of a bond can be written in terms of these variables by:

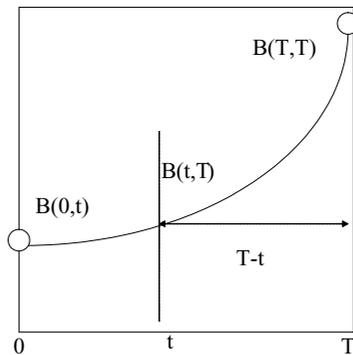
$$B(t, T) = V(y(t, T), \tau); \tau = T - t; \frac{\partial V(y(t, T), \tau)}{\partial y} < 0, \frac{\partial V(y(t, T), \tau)}{\partial \tau} < 0$$

Note that the larger the amount of time left to payment (the bond redemption time) the smaller the bond price (explaining its negative derivative). Further, the larger the discount factor-yield at time  $t$ , the smaller the value of the bond (see Figure 8.1).

The definition of a bond's price and its estimation is essential for financial management and mathematics. The behaviour of such a value and its properties underlies the process of interest-rate formation and vice versa, interest-rate processes define the value of bonds. Some obvious properties for bond values are:

$$B(t, t) = 1; \lim_{(T-t) \rightarrow \infty} B(t, T) = 0; B(s', s) > B(s'', s) \text{ if } s'' > s'$$

In other words, a bond paid instantly equals its nominal value, while a bond redeemed at infinity is null. Finally, two similar bond payouts, with one bond due before the other, imply that the one is worth more than the other. Thus, to value a bond, we need to express the time preference for money by the yield, representing



**Figure 8.1**

the effective bond discount rate  $y(t, T)$  at time  $t$  associated with a payment  $T - t$  periods later. The yield is one of the important functions investors and speculators alike seek to define. It is used by economists to capture the overall movement of interest rates (which are known as ‘yields’ in Wall Street parlance). There are various interest rates moving up and down, not necessarily in unison. Bonds of various maturities may move independently with short-term rates and long-term rates often moving in opposite directions simultaneously. The overall pattern of interest-rate movement – and what it means about the future of the economy and Wall Street – are the important issues to reckon with. They are thus like tea leaves, only much more reliable if one knows how to read them. Ordinarily, short-term bonds carry lower yields to reflect the fact that an investor’s money has less risk. The longer our cash is tied up, the theory goes, the more we should be rewarded for the risk taken. A normal yield curve, therefore, slopes gently upward as maturities lengthen and yields rise. From time to time, however, the curve twists itself into a few recognizable shapes, each of which signals a crucial, but different, turning point in the economy. When those shapes appear, it is often time to alter one’s assumptions about economic growth.

In discrete time, the value of a bond is given by discounting the future payout over the remaining period of time using the yield associated to the payment to be received. Of course, we can also calculate the yield as a function of the bond price and its time to maturity. This is done when data regarding bond values are more readily available than yields. For a discrete time bond, it would be written as follows:

$$B(t, T) = [1 + y(t, T)]^{-(T-t)} \quad \text{or} \quad y(t, T) = 1 - [B(t, T)]^{-1/T-t}$$

while in continuous time, it is written as follows:

$$B(t, T) = e^{-y(t, T)(T-t)} \quad \text{or} \quad y(t, T) = -\frac{\ln(B(t, T))}{(T-t)}$$

In other words, the yield and the bond are priced uniquely – one reflecting the other and vice versa. If this were not the case, then markets would be incomplete and ‘bond arbitrageurs’, for example, would identify such situations and profit from the ‘mis-pricing’ of bonds. For example, say that a bond paying \$1 in 2 years has a current market value of 0.85. Thus, the yield is found by solving the following equation:

$$\begin{aligned} 0.85 &= [1 + y(0, 2)]^{-2} \quad \text{and therefore the yield is } y(0, 2) \\ &= \sqrt{\left[ \frac{1}{0.85} \right]} - 1 = 0.08455 \end{aligned}$$

In this case, the bond has a return of 8.455%. If risk-free interest rates for the same period are 9%, then clearly it is economically appealing to use the difference in interest rates to make money.

### 8.1.2 Coupon-bearing bonds

Pure discount bonds such as the above are one of the ‘building blocks’ of finance and can be used to evaluate a variety of financial instruments. For example, if a default-free bond pays a periodic payment of \$ $c$  (the coupon payment) as well as a terminal nominal payment  $F$  at time  $T$  (the bond face value at maturity), then its price can be expressed in terms of zero-coupon bonds. Its value would be:

$$B_c(t, T) = c \sum_{k=t+1}^{T-1} B(t, k) + FB(t, T)$$

The same value expressed in terms of the yield will be, of course:

$$B_c(t, T) = c \sum_{k=t+1}^T \left[ \frac{1}{1 + y(t, k)} \right]^{k-t} + F \frac{1}{[1 + y(t, T)]^{T-t}}$$

For example, a bond whose face value is \$1000, with a coupon payout of \$50 yearly with a 9% interest rate has a current value of \$713.57. By the same token, another bond whose current price is \$800 and has the same properties (payout and face value) has a yield which is necessarily smaller than putting money in the bank and collecting money at the risk-free rate of 9%.

Valuation in continuous time yields the following equation:

$$B_c(t, T) = c \int_t^T B(t, \tau) d\tau + FB(t, T) \text{ or as yields}$$

$$B_c(t, T) = c \int_t^T e^{-y(t, \tau-t)(\tau-t)} d\tau + F e^{-y(t, T)(T-t)}$$

Mortgage payments, debt payment of various forms and sorts, investments yielding a fixed income etc. can be written in terms of bonds (assuming all payments to be default-free). In some cases, such as reverse mortgage, one may have to be careful in using bonds for the valuation of a financial contract. For example, in reverse mortgage, the bank would assume the responsibility of paying a fixed amount, say  $c$ , every month to a homeowner, as long as he lives. At death (which is a random time) the bank would ‘at last’ take ownership of the home. The value of the bond, thus, equals a coupon payout made for a random amount of time while receiving at the final random time (when the homeowner passes away) an amount equalling the home (random) value. These situations, of course, render the valuation of such contracts more difficult. What may seem at first a profitable contract may turn out to be disastrous subsequently. For this reason, considerable attention is devoted to these situations so that an appropriate pricing procedure and protection (hedging) may be structured. If sources of uncertainty can be determined in a fairly reliable manner, we can at least write the value of the bond equation in terms of these uncertain ingredients and proceed to numerical or simulation techniques to obtain a solution, providing that we can equate these

**Table 8.1** Term structure interest rates (source: ECB, 2000).

	1 year	2 years	3 years	4 years	5 years
$y(0, t)$	0.0527	0.053	0.0537	0.0543	0.0551
$y(1, t)$	—	0.0533	0.0542	0.0548	0.0557
$y(2, t)$	—	—	0.0551	0.0556	0.0565
$y(3, t)$	—	—	—	0.0561	0.0572
$y(4, t)$	—	—	—	—	0.0583

values to some replicating risk-free portfolio that would allow calculation of the appropriate discount rate. Additional problems are met when we introduce rated bonds, default bonds, junk bonds etc., as we shall see subsequently.

### Example

Consider a coupon-paying bond with a payout of \$100 a year for 4 years at the end of which the principal of \$1000 is redeemed. The current yield is found and given in Table 8.1. On the basis of this information, we are able to calculate the current bond price. Namely, assuming that this is a default-free bond, the bond price is:

$$B_{100}(0, 5) = 100 \sum_{k=1}^4 \left[ \frac{1}{1 + y(0, k)} \right]^k + 1000 \frac{1}{[1 + y(0, 5)]^5}$$

Table 8.1 provides the yields at and for various periods. Yields are calculated by noting that if there is no arbitrage then a dollar invested at time '0' for  $t$  periods should have the same value as a dollar invested for  $s$  periods and then reinvested for the remaining  $t - s$  periods. In other words, in complete markets, when there can be no arbitrage profit, we have:

$$[1 + y(0, t)]^t = [1 + y(0, s)]^s [1 + y(s, t)]^{t-s} \quad \text{and}$$

$$y(s, t) = \left( \frac{[1 + y(0, t)]^t}{[1 + y(0, s)]^s} \right)^{1/(t-s)} - 1$$

Thus, yields  $y(0, t)$  provide all the information needed to calculate the bond current price. Say that we are currently in the year 2000. This means that we have to insert the term structure rates of year 2000 in our equation in order to calculate the current bond price, or:

$$B_{100}(0, 5) = 100 \left[ \frac{1}{(1 + 0.0527)} + \frac{1}{(1 + 0.053)^2} + \frac{1}{(1 + 0.0537)^3} + \frac{1}{(1 + 0.0543)^4} \right] + 1100 \frac{1}{(1 + 0.0551)^5} = 1192.84$$

To determine the price a period hence, the appropriate table for the rates term structure will have to be used. If we assume no changes in rates, then the bond value is calculated by:

$$B_{100}(1, 5) = 100 \sum_{k=1}^4 \left[ \frac{1}{1 + y(1, k)} \right]^k + 1000 \frac{1}{[1 + y(1, 5)]^5} = 1155.72$$

which is a decline in the bond value of  $1192.84 - 1155.72 = 37.12$  dollars.

### 8.1.3 Net present values (NPV)

The NPV of an investment providing a stream of known-for-sure payments over a given time span can be also written in terms of zero-coupon bonds. The traditional NPV of a payment stream  $C_0, C_1, C_2, C_3, \dots, C_n$  with a fixed risk-free discount rate  $R_f$  is:

$$NPV = C_0 + \frac{C_1}{1 + R_f} + \frac{C_2}{(1 + R_f)^2} + \frac{C_3}{(1 + R_f)^3} + \dots + \frac{C_n}{(1 + R_f)^n}$$

There are some problems with this formula, however, for it is not market-sensitive, ignoring the rates term structure and the uncertainty associated with future payouts. If the payout is risk-free, it is possible to write each payment  $C_i$ ,  $i = 0, 1, 2, \dots, n$  in terms of zero-coupon (risk-free) bonds. At time  $t = 0$ ,

$$NPV_0 = C_0 + C_1 B(0, 1) + C_2 B(0, 2) + C_3 B(0, 3) + \dots + C_n B(0, n)$$

While a period later, we have:

$$NPV_1 = C_1 + C_2 B(1, 2) + C_3 B(1, 3) + \dots + C_n B(1, n)$$

with each bond valued according to its maturity. When a zero-coupon bond is rated or subject to default (which has not been considered so far), applying a constant discount rate to evaluate the NPV can be misleading since it might not account for changes in interest rates over time, their uncertainty as well as the risks associated with the bond payouts and the ability of the bond issuer to redeem it as planned. If a bond yield is time-varying, deterministically or in a random manner, then the value of the bond will change commensurately, altering over time the NPV. Corporate bonds (rated by financial agencies such as Standard and Poors, Moody's, Fitch), the value of corporations' cash flows must similarly reflect the corporate rating and their associated risks. In section 8.3, we consider these bonds and thereby provide an approach to valuing cash flows of rated corporations as well. The net present value at time  $t$  of a corporate cash flow is thus a random variable reflecting interest-rate uncertainty and the corporate rate and its reliability. An appropriate and equivalent way to write the NPV (assuming that cash payments are made for sure) using the yield  $y(0, i)$  associated with each

zero-coupon bond with maturity  $i$ ), is:

$$NPV(0|\mathfrak{R}) = C_0 + \frac{C_1}{1 + y(0, 1)} + \frac{C_2}{(1 + y(0, 2))^2} + \frac{C_3}{(1 + y(0, 3))^3} + \dots \\ + \frac{C_n}{(1 + y(0, n))^n}$$

Thus, generally, we can write a net present value at time  $t$  by:

$$NPV(t|\mathfrak{R}) = \sum_{i=t}^n C_i B(t, i) = \sum_{i=t}^n \frac{C_i}{[1 + y(t, i)]^{i-t}}$$

In a similar manner, a wide variety of cash flows and expenses may be valued. The implication of this discussion is that all cash flows, their timing and the uncertainty associated with these flows may also be valued using ‘bond mathematics’. When coupon payments are subject to default, we can represent the NPV as a sum of default-prone bonds, as will be discussed later on. Similarly, if the NPV we calculate is associated to a corporation whose debt (bond) is rated, then such rating also affects the value of the bond and thereby the corporation’s cash flow. Generally, bonds are used in many ways to measure asset values, to measure risks and to provide an estimate of many contracts that can be decomposed into bonds that can be, or are, traded.

#### 8.1.4 Duration and convexity

‘Duration’ is a measure for exposure to risk. It expresses the sensitivity of the bond price to (small) variations in interest rates. In other words, the duration at time  $t$  of a bond maturing at time  $T$ , written by  $D(t, T)$ , measures the return per unit for a move  $\Delta y$  in the yield, or

$$D(t, T) = -\frac{1}{B(t, T)} \frac{[\Delta B(t, T)]}{\Delta y(t, T)}$$

For small intervals of time, we can rewrite this expression as follows:

$$D(t, T) = -\frac{[\Delta \log B(t, T)]}{\Delta y(t, T)} \approx -\frac{d[\log B(t, T)]}{dy(t, T)}$$

Since the bond rate of return is  $R(t, T) = \Delta B(t, T)/B(t, T) \approx \Delta \log B(t, T)$  and  $\Delta y(t, T)$  is a rate move at time  $t$ , we can write:

$$R(t, T) = -D(t, T) [\Delta y(t, T)]$$

In words:

$$\text{Rate of returns on bonds} = -(\text{Duration}) * (\text{Yield rate move})$$

At time  $t$ , a zero-coupon bond maturing at time  $T$  has, of course, a duration of  $T - t$ . For a coupon bond with payments of  $C_i$  at times  $t_i$ ,  $i = 1, \dots, n$  and a bond

price yield denoted by  $B(0, n)$  and  $y(0, n)$ , then (in continuous-time discounting):

$$B(0, n) = \sum_{i=1}^n C_i e^{-y(0,n)t_i}$$

and the duration is measured by time-weighted average of the bond prices:

$$D(0, n) = \frac{\sum_{i=1}^n C_i t_i e^{-y(0,n)t_i}}{\sum_{i=1}^n C_i e^{-y(0,n)t_i}}$$

This result can be proved by simple mathematical manipulations since:

$$-\frac{d(\log B)}{dy} = D \text{ implies } -\frac{d\left(\log \sum_{i=1}^n C_i e^{-yt_i}\right)}{dy} = \frac{\sum_{i=1}^n C_i t_i e^{-yt_i}}{\sum_{i=1}^n C_i e^{-yt_i}} = D$$

While duration reflects a first-order change of the bond return with respect to its yield, convexity captures second-order effects in yield variations. Explicitly, let us take a second-order approximation to a bond whose value is a function of the yield. Informally, let us write the first three terms of a Taylor series expansion of the bond value:

$$B(t, y + \Delta y) = B(t, y) + \frac{\partial B(t, y)}{\partial y} \Delta y + \frac{1}{2} \frac{\partial^2 B(t, y)}{\partial y^2} (\Delta y)^2$$

Dividing by the bond value, we have:

$$\frac{\Delta B(t, y)}{B} = \frac{1}{B} \frac{\partial B(t, y)}{\partial y} \Delta y + \frac{1}{2} \frac{1}{B} \frac{\partial^2 B(t, y)}{\partial y^2} (\Delta y)^2$$

And, approximately, for a small variation in the yield of  $\Delta y$ , we have (replacing partial differentiation by differences):

$$\frac{\Delta B}{B} = \frac{1}{B} \frac{\Delta B}{\Delta y} \Delta y + \frac{1}{2} \frac{1}{B} \frac{\Delta^2 B}{\Delta y^2} (\Delta y)^2$$

If we define convexity by:

$$\Upsilon(t, T) = \frac{1}{B} \frac{\Delta^2 B}{\Delta y^2}$$

then, an expression for the bond rate of return in terms of the duration and the convexity is:

$$\frac{\Delta B}{B} = -D(t, T) \Delta y + \frac{1}{2} \Upsilon(t, T) (\Delta y)^2,$$

or

$$\left( \begin{array}{c} \text{Rate of} \\ \text{returns on bonds} \end{array} \right) = -(\text{Duration}) * \left( \begin{array}{c} \text{Yield} \\ \text{rate move} \end{array} \right) \\ + \frac{1}{2}(\text{Convexity}) * \left( \begin{array}{c} \text{Yield} \\ \text{rate move} \end{array} \right)^2$$

Thus, a fixed-income bond will lose value as the interest rate (i.e.  $\Delta y > 0$ ) increases and, conversely, it loses value when the interest decreases (i.e.  $\Delta y < 0$ ). For example, say that a coupon-bearing bond at time  $t_i$ ,  $i = 1, 2, 3, \dots$  with yield  $y$  is given at time  $t$  by:

$$B(t, T) = K e^{-y(T-t)} + \sum_{i=1}^{C_i} e^{-y(t_i-t)}$$

Note that:

$$\frac{dB}{dy} = -K(T-t)e^{-y(T-t)} - \sum_{i=1}^{C_i} C_i(t_i-t)e^{-y(t_i-t)}$$

$$\frac{d^2B}{dy^2} = -K(T-t)^2 e^{-y(T-t)} - \sum_{i=1}^{C_i} C_i(t_i-t)^2 e^{-y(t_i-t)}$$

And therefore the duration and the convexity express explicitly first- and second-order effects of yield variation, or:

$$D(t, T) = \frac{dB}{Bdy}; \quad \Upsilon(t, T) = \frac{d^2B}{Bdy^2}$$

### Example

We consider the following bond and calculate its duration:

Actual price:	100
Nominal interest rate:	10 % (p.a.)
Buy back value:	100
Years remaining:	4
The actual market interest rate:	10 %

The duration is defined by:

$$\text{Macaulay duration} = -\frac{PV'(Y)}{PV(Y)} = \frac{\sum_{i=1}^n t_i * c_i * (1+Y)^{-t_i}}{\sum_{i=1}^{n-1} c_i * (1+Y)^{-t_i} + (100 + c_n) * (1+Y)^{-t_n}}$$

where  $PV'$  is the derivative of  $PV(Y)$ . This means that the duration of the bond equals:

$$D_M^C = \frac{[1 * 10\,000 * (1.1)^{-1} + 2 * 10\,000 * (1.1)^{-2} + 3 * 10\,000 * (1.1)^{-3} + 4 * 110\,000 * (1.1)^{-4}]}{100\,000} = 3.5$$

If one invests in a bond at a given time and for a given period, the yield does not represent the rate of return of such an investment. This is due to the fact that coupon payments are reinvested at the same yield, which is not precise since yields are changing over time and coupon payments are reinvested at the prevailing yields when coupons are distributed. As a result, changing yield has two opposite effects on the investor rate of return. On the one hand, an increase in the yield decreases the bond value, as we saw earlier, while it increases the rate of return on the coupon. These two effects cancel out exactly when the investor holds the bond for a time period equal to its duration. Thus, by doing so, the rate of return will be exactly the yield at the time he acquired the bond and thus his investment is immune to changing yields. This strategy is called immunization. This strategy is in fact true only for small changes in the yield.

Explicitly, let  $B(t, y) = B(t, y : T)$  be the bond price at time  $t$  when the yield is  $y$  and the maturity  $T$ . Consider another instant of time  $t + \Delta t$  and let the yield at this time be equal  $y + \Delta y$ . In the (continuous) time interval  $[t, t + \Delta t]$  the coupon payment  $c$  is reinvested continuously at the new yield and therefore the bond values at time  $t$  and  $t + \Delta t$  are given by:

Time  $t$ :  $B(t, y)$

$$\text{Time } t + \Delta t: B(t + \Delta t, y + \Delta y) + \int_t^{t+\Delta t} c e^{-(y+\Delta y)(t+\Delta t-z)} dz$$

Thus, for immunization we require that the bond rate of return equals its current yield, or:

$$y = \frac{1}{\Delta t} \left[ \frac{B(t + \Delta t, y + \Delta y) + \int_t^{t+\Delta t} c e^{-(y+\Delta y)(t+\Delta t-z)} dz - B(t, y)}{B(t, y)} \right]$$

Since

$$\int_t^{t+\Delta t} c e^{-(y+\Delta y)(t+\Delta t-z)} dz = \frac{c}{(y + \Delta y)} [1 - e^{-(y+\Delta y)\Delta t}] \approx c \Delta t$$

and for small  $\Delta t$ ,

$$B(t + \Delta t, y + \Delta y) \approx B(t, y + \Delta y) + \frac{\partial B(t, y + \Delta y)}{\partial t} \Delta t$$

Inserting in our equation, we have:

$$1 + y \Delta t = \left[ \frac{B(t, y + \Delta y) + \frac{\partial B(t, y + \Delta y)}{\partial t} \Delta t + c \Delta t}{B(t, y)} \right]$$

and

$$\begin{aligned} \frac{\partial B(t + \Delta y)}{\partial t} &= \frac{\partial}{\partial t} \left( \int_t^T c e^{-(y+\Delta y)(z-t)} dz + e^{-(y+\Delta y)(T-t)} \right) \\ &= c(y + \Delta y) \int_t^T e^{-(y+\Delta y)(z-t)} dz - c + (y + \Delta y) e^{-(y+\Delta y)(T-t)} \\ &= (y + \Delta y) B - c \end{aligned}$$

Replacing these terms in the previous equation, we have:

$$1 + y \Delta t = \left[ \frac{B(t, y + \Delta y) + (y + \Delta y) B \Delta t - c \Delta t + c \Delta t}{B(t, y)} \right]$$

This is reduced to:

$$\begin{aligned} 1 + y \Delta t &= \left[ \frac{B(t, y + \Delta y)}{B(t, y)} \right] [1 + (y + \Delta y) \Delta t] \\ &= \left[ 1 + \frac{B(t, y + \Delta y) - B(t, y)}{B(t, y)} \right] [1 + (y + \Delta y) \Delta t] \end{aligned}$$

or

$$\begin{aligned} 1 + y \Delta t &= \left[ 1 + \frac{\Delta B}{B \Delta y} \Delta y \right] [1 + (y + \Delta y) \Delta t] \\ &= [1 - D(t, y) \Delta y] [1 + (y + \Delta y) \Delta t] \end{aligned}$$

Additional manipulations lead to the condition for immunization, namely that  $\Delta t$  equals the duration or  $\Delta t = D(t, y) - D(t, y)(y + \Delta y) \Delta t \approx D(t, y)$  and finally for very small  $\Delta t$

$$\Delta t = D(t, y)$$

## 8.2 BONDS AND FORWARD RATES

A forward rate is denoted by  $F(t, t_1, t_2)$  and is agreed on at time  $t$ , but for payments starting to take effect at a future time  $t_1$  and for a certain amount of time  $t_2 - t_1$ . In Figure 8.2, these times are specified.

A relationship between forward rates and spot rates hinges on an arbitrage argument. Roughly, this argument states (as we saw earlier), that two equivalent

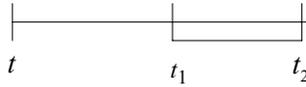


Figure 8.2

investments (from all points of view) have necessarily the same returns. Say that at time  $t$  we invest \$1 for a given amount of time  $t_2 - t$  at the available spot rate (its yield). The price of such an investment using a bond is then:  $B(t, t_2)$ . Alternatively, we could invest \$1 for a certain amount of time, say  $t_1 - t$ ,  $t_1 \leq t_2$  at which time the moneys available will be reinvested at a forward rate for the remaining time interval:  $t_2 - t_1$ . The price of such an investment will then be  $B(t, t_1)B_f(t_1, t_2)$  where  $B_f(t_1, t_2) = [1 + F(t, t_1, t_2)]^{-(t_2-t_1)}$  is the value of the bond at time  $t_1$  paying \$1 at time  $t_2$  using the agreed-on (at time  $t$ ) forward rate  $F(t, t_1, t_2)$ . Since both payments result in \$1 both received at time  $t_2$  they have the same value, for otherwise there will be an opportunity for arbitrage. For this reason, assuming no arbitrage, the following relationship must hold (and see Figure 8.3):

$$B(t, t_2) = B(t, t_1)B_f(t_1, t_2) \text{ implying } B_f(t_1, t_2) = \frac{B(t, t_2)}{B(t, t_1)}$$

In discrete and continuous time, assuming no arbitrage, this leads to the following forward rates:

$$[1 + F(t, t_1, t_2)]^{t_2-t_1} = \frac{[1 + y(t, t_2)]^{t_2-t}}{[1 + y(t, t_1)]^{t_1-t}} \text{ (discrete time)}$$

$$F(t, t_1, t_2) = \frac{y(t, t_2)(t_2 - t) - y(t, t_1)(t_1 - t)}{(t_2 - t_1)} \text{ (continuous time)}$$

In practice, arbitrageurs can make money by using inconsistent valuations by bond and forward rate prices. For complete markets (where no arbitrage is possible), the spot rate (yield) contains all the information regarding the forward market rate and, vice versa, the forward market contains all the information regarding the spot market rate, and thus it will not be possible to derive arbitrage profits. In practice, however, some pricing differences may be observed, as stated above, opening up arbitrage opportunities.

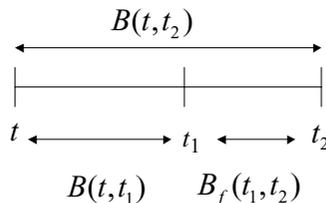


Figure 8.3

**Problem**

An annuity pays the holder a scheduled payment over a given amount of time (finite or infinite). Determine the value of such an annuity using bond values at the current time. What would this value be in two years using the current observed rates?

**Problem**

What will be the value of an annuity that starts in  $T$  years and will be paid for  $M$  years afterwards? How would you write this annuity if it is terminated at the time the annuity holder passes away (assuming that all payments are then stopped)?

**Problem**

Say that we have an obligation whose nominal value is \$1000 at the fixed rate of 10 % with a maturity of 3 years, reimbursed in fine. In other words, the firm obtains a capital of \$1000 whose cost is 10 %. What is the financial value of the obligation? Now, assume that just after the obligation is issued the interest rate falls from 10 to 8 %. The firm's cost of finance could have been smaller. What is the value of the obligation (after the change in interest rates) and what is the 'loss' to the firm.

### 8.3 DEFAULT BONDS AND RISKY DEBT

Bonds are rated to qualify their standard risks. Standard and Poors, Moody's and other rating agencies use for example, AAA, AA, A, BB, etc. to rate bonds as more or less risky. We shall see in section 8.4 that these rating agencies also provide Markov chains, expressing the probabilities that rated firms switch from one rating to another, periodically adapted to reflect market environment and the conditions particularly affecting the rated firm (for example, the rise and fall of the technology sector, war and peace, and their likes).

Consider a portfolio of B-rated bonds yielding 14 %; typically, these are bonds which currently are paying their coupons, but have a high likelihood of defaulting or have done so in the recent past. A Treasury bond of similar duration yields 5.5 %. Thus, in this example, the Junk–Treasury Spread (JTS) is 8.5 %. Now, let us take a look at the spread's history over the past 13 years (Jay Diamond, *Grant's Interest Rate Observer* data).

The spread depicted in Figure 8.4 corresponds roughly to a B-rated debt. Note the very wide range of spreads, from just below 3 % to almost 10 %. What does a JTS of 3 % mean? Very bad news for the junk buyer, because he or she will have been better off in Treasuries if the loss rate exceeds 3 %. And even if the loss rate is only half of that, a 1.5 % return premium does not seem adequate to compensate for this risk. There is a wealth of data on the bankruptcy/default rate, allowing us to evaluate whether the prevailing risk premium amounts to adequate compensation.



**Figure 8.4** Junk-Treasury spread 1988–2000 (Jay Diamond, *Grant's Interest Rate Observer* data).

Rating agencies often use terms such as default rate and loss rate which are important to understand. The former defines the proportion of companies defaulting per year. But not all companies that default go bankrupt. The recovery rate is the proportion of defaulting companies that do not eventually go bankrupt. So a portfolio's reduction in return is calculated as the default rate times one minus the recovery rate: if the default rate is 4 % and the recovery rate is 40 %, then the portfolio's total return has been reduced by 2.4 %. The loss rate, how much of the portfolio actually disappears, is simply the default rate minus the absolute percentage of companies which recover. According to Moody's, the annual long-term default rate of bonds rated BBB/Baa (the lowest 'investment grade') is about 0.3 %; for BB/Ba, about 1.5 %; and for B, about 7 %. But in any given year, the default rate varies widely. Further, because of the changes in the high-yield market that occurred 15 years ago, the pre-1985 experience may not be of great relevance to high-yield investing today.

Prior to the use of junk bonds the overwhelming majority of speculative issues were 'fallen angels', former investment-grade debt which had fallen on hard times. But, after 1985, most high-yield securities were speculative right from their initial offering. Once relegated to bank loans, poorly rated companies were for the first time able to issue debt themselves. This was not a change for the better. Similar to speculative stock IPOs, these new high-yield bond issues tended to have less secure 'coverage' (based on an accounting term defined as the ratio of earnings-before-taxes-and-interest to total interest charges) than the fallen angels of yore, and their default rates were correspondingly higher.

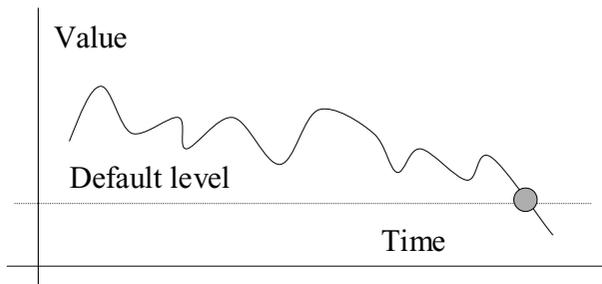
Many financial institutions hold large amounts of default-prone risky bonds and securities of various degrees of complexity in their portfolios that require a reliable estimate of the credit exposure associated with these holdings. Models of default-prone bonds fall into one of two categories: *structural models* and *reduced-form models*. Structural models specify that default occurs when the firm value falls below some explicit threshold (for example, when the debt to equity ratio crosses a given threshold). In this sense, default is a 'stopping time' defined by the evolution of a representative stochastic process. Merton (1974) first considered such a problem; it was studied further by many researchers including Black and Cox (1976), Leland (1994), and Longstaff and Schwartz (1995). These models determine both equity and debt prices in a self-consistent manner via arbitrage, or contingent-claims pricing. Equity is assumed to possess characteristics similar to a call option, while debt claims have features analogous to claims on the firm's value. This interpretation is useful for predicting the determinants of credit-spread changes, for example.

Some models assume as well that debt-holders get back a fraction of the debt, called the recovery ratio. This ratio is mostly specified a priori, however. While this is quite unrealistic, such an assumption removes problems associated to the debt seniority structure, which is a drawback of Merton's (1974) model. Some authors, for example, Longstaff and Schwartz (1995), argue that, by looking at the history of defaults and recovery ratios for various classes of debt of comparable firms, one can find a reliable estimate of the recovery ratio. Structural models are, however, difficult to use in valuing default-prone debt, due to difficulties associated with determining the parameters of the firm's value process needed to value bonds. But one may argue that parameters could always be retrieved from market prices of the firm's traded bonds. Further, they cannot incorporate credit-rating changes that occur frequently for default-prone (risky) corporate debts.

Many corporate bonds undergo credit downgrades by credit-rating agencies before they actually default, and bond prices react to these changes (often brutally) either in anticipation or when they occur. Thus, any valuation model should take into account the uncertainty associated with credit-rating changes as well as the uncertainty surrounding default and the market's reactions to such changes. These shortcomings make it necessary to look at other models for the valuation of defaultable bonds and securities that are not predicated on the value of the firm and that take into account credit-rating changes. For example, a meltdown of financial markets, wars, political events of economic importance are such cases, where the risk is exogenous (rather than endogenous). This leads to reduced-form models.

The problem of rating the credit of bonds and credit markets is in fact more difficult than presumed by analytical models. Information asymmetries compound these difficulties. Akerlof in his 2001 Nobel allocation pointed to these effects further.

A bank granting a credit has less information than the borrower, on his actual default risk. . . . On the same token, banks expanding into new, unknown markets are at a particular risk. On the one hand, due to their imperfect market knowledge, they must rely on the equilibrium between supply and demand to a large extent. On the other hand, under asymmetric information, it is very



**Figure 8.5** Structural models of default.

easy for clients to hide risks and to give too optimistic profit estimates, possibly approaching fraud in extreme cases. Adverse selection then implies a markedly increased default risk for such banks. Banks can use interest rates and additional security as instruments for screening the creditworthiness of clients when they estimate that their information is insufficient. Credit risk and pricing models, of course, are complementary tools. Based on information provided by the client, they produce risk-adjusted credit spreads and thus may set limits to the principle of supply and demand. On the other hand, borrowers with a credit rating may use this rating to signal the otherwise private information on their solvency, to the bank. In exchange, they expect to receive better credit conditions than they would if the bank could only use information on sample averages.

Technically, the value process is defined in terms of a stochastic process  $\{x, t \geq 0\}$  while default is defined by the first time  $\tau$  (the stopping time) the process reaches a predefined threshold-default level. In other words, let the threshold space be  $\mathfrak{R}$ , then:

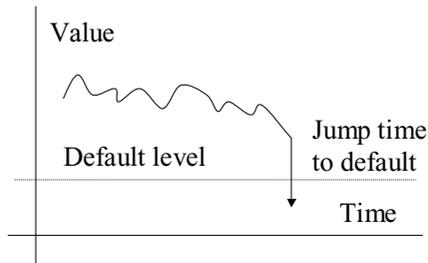
$$\tau = \text{Inf} \{t > 0, x(t) \notin \mathfrak{R}\}$$

where  $\mathfrak{R}$  is used to specify the set of feasible states for an operating firm. As soon as the firm's value is out of these states, default occurs.

Reduced-form models specify the default process explicitly, interpreting it as an exogenously motivated jump process, usually expressed as a function of the firm value. This class of models has been investigated, for example by Jarrow and Turnbull (1995), Jarrow *et al.* (1997) and others. Although these models are useful for fitting default to observed credit spreads, they mostly neglect the underlying value process of the firm and thus they can be less useful when it is necessary to determine credit spread variations. Jarrow *et al.* (1997) in particular have adopted the rating matrix used by financial institutions such as Moody's, Standard and Poors and others as a model of credit rating (as we too shall do in the next section).

Technically, default is defined exogenously by a random variable  $\tilde{T}$  where  $t < \tilde{T} < T$ , with  $T$ , the bond expiry date. The conditional probability of default is assumed given by:

$$P(\tilde{T} \in (t + dt) | t < \tilde{T} < T) = q(x) dt + 0(dt)$$



**Figure 8.6** Reduced-form models default.

This means that the conditional probability of default  $q$  in a small time interval  $(t + dt)$ , given that no default has occurred previously, is a function of an underlying stochastic process  $\{x, t \geq 0\}$ . If the probability  $q$  is independent of the process  $\{x, t \geq 0\}$ , this implies that the probability of default is of the exponential type. That is to say, it implies that at each instant of time, the probability of default is time-independent and independent of the underlying economic fundamentals. These are very strong assumptions and therefore, in practice, one should be very careful in using these models.

A comparison between structural and reduced-form models (see Figure 8.6) is outlined in Table 8.2. Selecting one model or the other is limited by the underlying risk considered and the mathematical and statistical tractability in applying such a model. These problems are extensively studied, as the references at the end of the chapter indicate.

A general technical formulation, combining both structural and reduced-form models leads to a time to default we can write by  $\text{Min}(\tau, \tilde{T}, T)$  where  $T$  is the maturity reached if no default occurs, while exogenous and endogenous default are given by the random variables  $(\tau, \tilde{T})$ . If the yield of such bonds at time  $t$  for a payout at  $s$  is given by,  $Y(t, s) \equiv y(\tau, \tilde{T}, s)$ , the value of a pure default-prone bond paying \$1 at redemption is then  $E \exp(-Y(t, T) \text{Min}(\tau, \tilde{T}, T))$ . Of course if there was no default, the yield would be  $y(t, s)$  and therefore  $Y(t, s) > y(t, s)$  in order to compensate for the default risk. The essential difficulty of these problems is to determine the appropriate yield which accounts for such risks, however. For example, consider the current value of a bond retired at  $\text{Min}(\tau, \tilde{T}, T)$  and paying an indexed coupon payout indexed to some economic variable or economic index (inflation, interest rate etc.). Uncertainty regarding the coupon payment, its nominal value and the bond default must then be appropriately valued through the bond yield.

When a bond is freely traded, the coupon payment can also be interpreted as a ‘bribe’ paid to maintain bond holding. For example, when a firm has coupon payments that are too large, it might redeem the bond (provided it incurs the costs associated with such redemption). By the same token, given an investor with other opportunities, deemed better than holding bonds, it might lead the investor to forgo future payouts and principal redemption, and sell the bond at its current

**Table 8.2** A comparison of selected models.

Model	Advantages	Drawbacks
Merton (1974)	Simple to implement.	(a) Requires inputs about the firm value. (b) Default occurs only at debt maturity. (c) Information about default and credit-rating changes cannot be used.
Longstaff and Schwartz (1995)	(a) Simple to implement. (b) Allows for stochastic term structure and correlation between defaults and interest rates.	(a) Requires inputs related to the firm value. (b) Information in the history of defaults and credit-rating changes cannot be used.
Jarrow, Lando, and Turnbull (1997)	(a) Simple to implement. (b) Can match exactly existing prices of default-risky bonds and thus infer risk-neutral probabilities for default and credit-rating changes. (c) Uses the history of default and credit-rating change.	(a) Correlation not allowed between default probabilities and the level of interest rates. (b) Credit spreads change only when credit ratings change.
Lando (1998)	(a) Allows correlation between default probabilities and interest rates. (b) Allows many existing term-structure models to be easily embedded in the valuation framework.	Historical probabilities of defaults and credit-rating changes are used assuming that the risk premiums due to defaults and rating changes, is null.
Duffie and Singleton (1997)	(a) Allows correlation between default probabilities and the level of interest rates. (b) Recovery ratio can be random and depend on the pre-default value of the security. (c) Any default-free term-structure model can be accommodated, and existing valuation results for default-free term-structure models can be readily used.	Information regarding credit-rating history and defaults cannot be used.
Duffie and Huang (1996) (swaps)	(a) Has all the advantages of Duffie and Singleton. (b) Asymmetry in credit qualities is easily accommodated. (c) ISDA guidelines for settlement upon swap default can be incorporated.	(a) Information regarding credit-rating history and defaults cannot be used. (b) Computationally difficult to implement for some swaps, such as cross-currency swaps, if domestic and foreign interest rates are assumed to be random.

market value. The number of cases we might consider is very large indeed, but only a few such cases will be considered explicitly here.

Structural and reduced-form models for valuing default-prone debt do not incorporate financial restructuring (and potential recovery) that often follows default. Actions such as renegotiating the terms of a debt by extending the maturity or lowering/postponing promised payments, exchanging debt for other forms of security, or some combination of the above (often being the case after default), are not considered. Similarly, institutional and reorganization features (such as bankruptcy) cannot be incorporated in any of these models simply. Further, anticipated debt restructurings by the market is priced in the value of a defaultable bond in ways that none of these models captures. In fact, many default-prone securities are also thinly traded. Thus, a liquidity premium is usually incorporated into these bond prices, hiding their risk of default. Finally, empirical evidence for these models is rather thin. Duffie and Singleton (1997, 1999) find that reduced-form models have problems explaining the observed term structure of credit spreads across firms of different credit qualities. Such problems could arise from incorrect statistical specifications of default probabilities and interest rates or from models' inability to incorporate some of the features of default/bankruptcy mentioned above. Bond research, just like finance in general, remains therefore a domain of study with many avenues to explore and questions that are still far from resolved.

## 8.4 RATED BONDS AND DEFAULT

The potential default of bonds and changes in rating are common and outstanding issues to price and reckon with in bonds trading and investment. They can occur for a number of reasons, including some of the following:

- (1) Default of the payout or default on the redemption of the principal.
- (2) Purchasing power risk arising because inflationary forces can alter the value of the bond. For example, a bond which is not indexed to a cost of living index may in fact generate a loss to the borrower in favour of the lender should inflation be lower than anticipated thereby increasing the real or inflation-deflated payments.
- (3) Interest-rate risk, resulting from predictable and unpredictable variations in market interest rates and therefore the bond yield.
- (4) Delayed payment risk, and many other situations associated with the financial health of the bond issuer and its credit reliability.

These situations are difficult to analyse but rating agencies specializing in the analysis and the valuation of financial assets provide ratings to nations and corporate entities that are used to price bonds. These agencies provide explicit matrices that associate to various bond classes (AAA, AA, B etc.) probabilities (a Markov chain) of remaining in a given class or switching to another (higher or lower) risk (rating) class. Table 8.3 shows a scale of ratings assigned to bonds by financial

**Table 8.3** Ratings.

Moody's	S&P	Definition
Aaa	AAA	Highest rating available
Aa	AA	Very high quality
A	A	High quality
Baa	BBB	Minimum investment grade
Ba	BB	Low grade
B	B	Very speculative
Caa	CCC	Substantial risk
Ca	CC	Very poor quality
C	D	Imminent default or in default

firms services (Moody's, Standard and Poors etc.) that start from the best quality to the lowest. In addition to these ratings, Moody's adds a '1' to indicate a slightly higher credit quality; for instance, a rating of 'A1' is slightly higher than a rating of 'A' whereas 'A3' is slightly lower. S&P ratings may be modified by the addition of a '+' or '-' (plus or minus). 'A+' is a slightly higher grade than 'A' and 'A-' is slightly lower. Occasionally one may see some bonds with an 'NR' in either Moody's or S&P. This means 'not rated'; it does not necessarily mean that the bonds are of low quality. It basically means that the issuer did not apply to either Moody's or S&P for a rating. Government agencies are a good example of very high quality bonds that are not rated by S&P. Other things being equal, the lower the rating, the higher the yield one can expect. Insured bonds have the highest degree of safety of all non-government bonds. Bond insurance agencies guarantee the payment of principal and interest on the bonds they have insured (since insurance reduces the bond's risk). When bonds are insured by one of the major insurance agencies, they automatically attain 'AAA' rating, identifying the bond as one of the highest quality one can buy. Some of the major bond insurers are AMBAC, MBIA, FGIC and FSA. In such circumstances, bonds have almost no default risk.

The Moody's rating matrix shown in Table 8.4 is an example. For AAA bonds, the probability that it maintains such a rating is .9193 while there is a .0746 probability that the bond rating is downgraded to a AA bond and so on for remaining values. These matrices are updated and changed from time to time as business conditions change. Given these matrices we observe that even a triple AAA bond is 'risky' since there is a probability that it be downgraded and its price reduced to reflect such added risk.

In some cases, the price of downgrading the credit rating of a firm can be much larger than presumed. For example, buried in Dynegy Inc.'s annual report for 2001 is a '\$301 million paragraph'. The provision is listed on page 28 of a 114-page document is the only published disclosure showing that Dynegy will have to post that much collateral if the ratings of its Dynegy Holdings Inc. unit are cut to junk status, or below investment grade. Debtors like Dynegy, WorldCom Inc. and Vivendi Universal SA are obligated to pay back billions of dollars if their

**Table 8.4** A typical Moody's rating matrix.

	AAA	AA	A	BBB	BB	B	CCC	D	NR
AAA	91.93 %	7.46 %	0.48 %	0.08 %	0.04 %	0.00 %	0.00 %	0.00 %	—
AA	0.64 %	91.81 %	6.76 %	0.60 %	0.06 %	0.12 %	0.03 %	0.00 %	—
A	0.07 %	2.27 %	91.69 %	5.12 %	0.56 %	0.25 %	0.01 %	0.04 %	—
BBB	0.04 %	0.27 %	5.56 %	87.88 %	4.83 %	1.02 %	0.17 %	0.24 %	—
BB	0.04 %	0.10 %	0.61 %	7.75 %	81.48 %	7.90 %	1.11 %	1.01 %	—
B	0.00 %	0.10 %	0.28 %	0.46 %	6.95 %	82.80 %	3.96 %	5.45 %	—
CCC	0.19 %	0.00 %	0.37 %	0.75 %	2.43 %	12.13 %	60.45 %	23.69 %	—
D	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	100.00 %	—

credit ratings fall, their stock drops or they fail to meet financial targets. Half of these so-called triggers have not been disclosed publicly, according to Moody's Investors Service Inc., which has been investigating the presence of such clauses since the collapse of the Enron Corp. (International Herald Tribune, May 9 May 2002, p. 15). In other words, in addition to a corporation's rating, there are other sources of information, some revealed and some hidden, that differentiate the value of debt for such corporations, even if they are equally rated. In other words, their yield may not be the same even if they are equally rated. The rating of bonds is thus problematic, although there is an extensive insurance market for bonds that index premiums to the bond rating. A dealer's quotes in Moody's provides, for example, an estimate of insurance costs for certain bonds (determined by the swap market), some of which are reproduced in Table 8.5. The premium paid varies widely, however, based on both the rating and the perceived viability of the company whose bond is insured.

**Table 8.5** Date-Moody's, 20 May 2002.

Company	Premium cost per \$M for 5 years	Moody's Senior Debt Rating
Merrill Lynch	10 000	Aa3
Lehman Brothers	9 500	A2
American Express	8 000	A1
Bear Stearns	7 500	A2
Goldman Sachs	6 500	A1
GE Capital	6 500	Aaa
Morgan Stanley	6 500	Aa3
JP Morgan Chase	6 500	Aa3*
AIG	5 300	Aaa
Citigroup	4 000	Aa1*
Bank of America	4 000	Aa2*
Bank One	3 500	Aaa

The valuation of rated bonds is treated next by making some simplifying assumptions to maintain an analytical and computational tractability, and by solving some problems that highlight approaches to valuing rated bonds. Rating can only serve as a first indicator to future default risk. Good accounting, information (statistical and otherwise) and economic analyses are still necessary.

### 8.4.1 A Markov chain and rating

Consider first a universe of artificial (and in fact non-existing) coupon-bearing rated bonds with a payment of a dollar fraction  $\ell_i$  at maturity  $T$ , depending on the rating of the bond at maturity. Risk is thus induced only by the fraction  $1 - \ell_i$  lost at maturity. Further, define the bond  $m$ -ratings matrix by a Markov chain  $[p_{ij}]$  where

$$0 \leq p_{ij} \leq 1, \quad \sum_{j=1}^m p_{ij} = 1$$

denotes the objective transition probability that a bond rated  $i$  in a given period will be rated  $j$  in the following one. Discount factors are a function of the rating states, thus a bond rated  $i$  has a spot (one period) yield  $R_{it}$ ,  $R_{it} \leq R_{jt}$  for  $i < j$  at time  $t$ . As a result, a bond rated  $i$  at time  $t$  and paying a coupon  $c_{it}$  at this time has, under the usual conditions, a value given by:

$$B_{i,t} = c_{it} + \sum_{j=1}^m \frac{p_{ij}}{1 + R_{jt}} B_{j,t+1}; \quad B_{i,T} = \ell_i, \quad i = 1, 2, 3, \dots, m$$

Note that the discount rate  $R_{jt}$  is applied to a bond rated  $j$  in the next period. For example, for an ‘imaginary’ rated bond  $A$  and  $D$  only, each with (short-term) yields  $(R_{At}, R_{Dt})$  at time  $t$  and a rating matrix specified by the transition probabilities,  $[p_{ij}]$ ;  $i, j = A, D$ , we have:

$$B_{A,t} = c_{A,t} + \left[ \frac{p_{AA}}{1 + R_{At}} B_{A,t+1} + \frac{p_{AD}}{1 + R_{Dt}} B_{D,t+1} \right]; \quad B_{A,T} = \ell_A = 1$$

$$B_{D,t} = c_{D,t} + \left[ \frac{p_{DA}}{1 + R_{At}} B_{A,t+1} + \frac{p_{DD}}{1 + R_{Dt}} B_{D,t+1} \right]; \quad B_{D,T} = \ell_D$$

where  $(c_{A,t}, c_{D,t})$  are the payouts associated with the bond rating and  $(\ell_A, \ell_D)$  are the bond redemption values at maturity  $T$ , both a function of the bond rating. In other words, the current value of a rated bond equals current payout plus the expected discount value of the bond rated at all classes, using the corresponding yield for each class at time  $t + 1$ . Of course, at the terminal time, when the bond is due (since there is not yet any default), the bond value equals its nominal value. If at maturity the bond is rated A, it will pay the nominal value of one dollar ( $\ell_A = 1$ ) while if it is rated D, it will imply a loss of  $1 - \ell_D$  for a bond rated initially A. If we define D as a default state (i.e. where the bondholder cannot recuperate the bond nominal value), the  $\ell_D$  can be interpreted as the recuperation ratio. Of course, we can assume as well  $\ell_D = 0$  as will be the case in a number

of examples below. In vector notation we have:

$$\begin{bmatrix} B_{A,t} \\ B_{D,t} \end{bmatrix} = \begin{bmatrix} c_{A,t} \\ c_{D,t} \end{bmatrix} + \begin{bmatrix} \frac{p_{AA}}{1+R_{At}} & \frac{p_{AD}}{1+R_{Dt}} \\ \frac{p_{DA}}{1+R_{At}} & \frac{p_{DD}}{1+R_{Dt}} \end{bmatrix} \begin{bmatrix} B_{A,t+1} \\ B_{D,t+1} \end{bmatrix}; \quad \begin{bmatrix} B_{A,T} \\ B_{D,T} \end{bmatrix} = \begin{bmatrix} \ell_A \\ \ell_D \end{bmatrix};$$

where  $\ell_i$  is the nominal value of a bond rated  $i$  at maturity. And generally, for an  $m$ -rated bond,

$$\mathbf{B}_t = \mathbf{c}_t + \mathbf{F}_t \mathbf{B}_{t+1}, \quad \mathbf{B}_T = \mathbf{L}$$

Note that the matrix  $\mathbf{F}_t$  has entries  $[p_{ij}/(1+R_{jt})]$  and  $\mathbf{L}$  is a diagonal matrix of entries  $\ell_i$ ,  $i = 1, 2, \dots, m$ . For a zero-coupon bond, we have:

$$\mathbf{B}_t = \prod_{k=t}^T \mathbf{F}_k \mathbf{L}.$$

By the same token, rated bonds discounts  $q_{it} = 1/(1+R_{it})$  are found by solving the matrix equation:

$$\begin{bmatrix} q_{1t} \\ q_{2t} \\ \dots \\ \dots \\ q_{mt} \end{bmatrix} = \begin{bmatrix} p_{11}B_{1,t+1} & p_{12}B_{2,t+1} & \dots & \dots & p_{1m}B_{m,t+1} \\ p_{21}B_{1,t+1} & p_{22}B_{2,t+1} & & & p_{2m}B_{m,t+1} \\ \dots & \dots & & & \dots \\ \dots & \dots & & & \dots \\ p_{m1}B_{1,t+1} & p_{m2}B_{2,t+1} & & & p_{mm}B_{m,t+1} \end{bmatrix}^{-1} \begin{bmatrix} B_{1,t} - c_{1t} \\ B_{2,t} - c_{2t} \\ \dots \\ \dots \\ B_{m,t} - c_{mt} \end{bmatrix}$$

where at maturity  $T$ ,  $B_{i,T} = \ell_i$ . Thus, in matrix notation, we have:

$$\bar{\mathbf{q}}_t = \Gamma_{t+1}^{-1} (\mathbf{B}_t - \mathbf{c}_t)$$

Note that one period prior to maturity, we have:  $\bar{\mathbf{q}}_{T-1} = \Gamma_T^{-1} (\mathbf{B}_{T-1} - \mathbf{c}_{T-1})$  where  $\Gamma_T$  is a matrix with entries  $p_{ij}B_{j,T} = p_{ij}\ell_j$ . For example, for the two-ratings bond, we have:

$$\begin{aligned} B_{1,t} &= c_{1t} + q_{1t}p_{11}B_{1,t+1} + q_{2t}p_{12}B_{2,t+1}; & B_{1,T} &= \ell_1 \\ B_{2,t} &= c_{2t} + q_{1t}p_{21}B_{1,t+1} + q_{2t}p_{22}B_{2,t+1}; & B_{2,T} &= \ell_2 \end{aligned}$$

Equivalently, in matrix notation, this is given by:

$$\begin{bmatrix} q_{1t} \\ q_{2t} \end{bmatrix} = \begin{bmatrix} p_{11}B_{1,t+1} & p_{12}B_{2,t+1} \\ p_{21}B_{1,t+1} & p_{22}B_{2,t+1} \end{bmatrix}^{-1} \begin{bmatrix} B_{1,t} - c_{1t} \\ B_{2,t} - c_{2t} \end{bmatrix}$$

In this sense the forward bond price can be calculated by the rated bond discount rate and vice versa.

### Example

We consider the matrix representing a rated bond supplied by Moody's (Table 8.6). The discount rates  $R_i$ ,  $i = AAA, \dots, D$  for each class and the corresponding coupon payments are given in Table 8.7. For example, the discount rate of an AAA bond is 0.06 yearly while that of a BBB bond is 0.1. In addition, the AAA-rated bond has a coupon paying \$1, while if it were rated BB its coupon payment

**Table 8.6**

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	0.9193	0.0746	0.0048	0.0008	0.0004	0	0	0
AA	0.0064	0.9181	0.0676	0.006	0.0006	0.0012	0.0003	0
A	0.0007	0.0227	0.9169	0.0512	0.0056	0.0025	0.0001	0.0004
BBB	0.0004	0.0027	0.0556	0.8788	0.0483	0.0102	0.0017	0.0024
BB	0.0004	0.001	0.0061	0.0775	0.8148	0.079	0.0111	0.0101
B	0	0.001	0.0028	0.0046	0.0695	0.828	0.0396	0.0545
CCC	0.0019	0	0.0037	0.0075	0.0243	0.1213	0.6045	0.2369
D	0	0	0	0	0	0	0	1

would have to be \$1.4. In this sense, both the size of the coupon and the discount applied to the rated bond are used to pay for the risk associated with the bond. The bond nominal value is \$100 with a lifetime of ten years. An elementary program will yield then the following bond value, shown above for each class at each year till the bond’s redemption. For example, initially, the premium paid for a AAA bond compared to a AA one is  $(63.10 - 58.81) = \$4.29$ . At the end of the fifth year, however, the AAA-AA bond price differential is only  $82.59 - 80.11 = \$2.48$ . In fact, we note that the smaller the amount of time left to bond redemption, the smaller the premium.

**8.4.2 Bond sensitivity to rates – Duration**

For the artificial rated bond considered above, we can calculate the duration of a rated bond through the rated bond sensitivity to the yields of each rating. For simplicity, assume that short yields are constants and calculate the partial derivatives for a bond rated A or D only. In this case, we seek to calculate the partials:

$$\left[ \frac{\partial B_{A,t}}{\partial R_A}, \frac{\partial B_{A,t}}{\partial R_D}, \frac{\partial B_{D,t}}{\partial R_A}, \frac{\partial B_{D,t}}{\partial R_D} \right],$$

$$\frac{\partial B_{A,t}}{\partial R_A} = -\frac{p_{AA}}{(1 + R_A)^2} B_{A,t+1} + \frac{p_{AA}}{1 + R_A} \frac{\partial B_{A,t+1}}{\partial R_A} + \frac{p_{AD}}{1 + R_D} \frac{\partial B_{D,t+1}}{\partial R_A}; \frac{\partial B_{A,T}}{\partial R_A} = 0$$

$$\frac{\partial B_{A,t}}{\partial R_D} = \left[ \frac{p_{AA}}{1 + R_A} \frac{\partial B_{A,t+1}}{\partial R_D} - \frac{p_{AD}}{(1 + R_D)^2} B_{D,t+1} + \frac{p_{AD}}{1 + R_D} \frac{\partial B_{D,t+1}}{\partial R_D} \right]; \frac{\partial B_{A,T}}{\partial R_D} = 0$$

$$\frac{\partial B_{D,t}}{\partial R_A} = \left[ -\frac{p_{DA}}{(1 + R_A)^2} B_{A,t+1} + \frac{p_{DA}}{1 + R_A} \frac{\partial B_{A,t+1}}{\partial R_A} + \frac{p_{DD}}{1 + R_D} \frac{\partial B_{D,t+1}}{\partial R_A} \right]; \frac{\partial B_{D,T}}{\partial R_A} = 0$$

$$\frac{\partial B_{D,t}}{\partial R_D} = \left[ \frac{p_{DA}}{1 + R_A} \frac{\partial B_{A,t+1}}{\partial R_D} - \frac{p_{DD}}{(1 + R_D)^2} B_{D,t+1} + \frac{p_{DD}}{1 + R_D} \frac{\partial B_{D,t+1}}{\partial R_D} \right]; \frac{\partial B_{D,T}}{\partial R_D} = 0$$

**Table 8.7**

Results	AAA	AA	A	BBB	BB	B	CCC	D
T=0	63.10	58.81	54.85	51.30	47.89	44.84	42.32	39.88
T=1	65.89	61.78	57.96	54.50	51.15	48.13	45.61	43.15
T=2	68.84	64.96	61.32	58.00	54.74	51.78	49.29	46.83
T=3	71.98	68.37	64.95	61.80	58.70	55.84	53.41	51.00
T=4	75.31	72.02	68.88	65.95	63.04	60.34	58.03	55.71
T=5	78.84	75.92	73.12	70.48	67.83	65.35	63.20	61.03
T=6	82.59	80.11	77.70	75.41	73.10	70.91	68.99	67.05
T=7	86.56	84.58	82.64	80.79	78.89	77.08	75.48	73.84
T=8	90.78	89.38	87.99	86.65	85.27	83.93	82.75	81.52
T=9	95.25	94.51	93.76	93.04	92.28	91.55	90.88	90.20
T=10	100	100	100	100	100	100	100	100

AAA	0.06
AA	0.07
A	0.08
BBB	0.09
BB	0.1
B	0.11
CCC	0.12
D	0.13

AAA	1
AA	1.1
A	1.2
BBB	1.3
BB	1.4
B	1.5
CCC	1.6
D	1.7

We can write in vector notation a system of six simultaneous equations given by:

$$\Gamma_t = \left[ B_{A,t}, B_{D,t}, \frac{\partial B_{A,t}}{\partial R_A}, \frac{\partial B_{A,t}}{\partial R_D}, \frac{\partial B_{D,t}}{\partial R_A}, \frac{\partial B_{D,t}}{\partial R_D} \right];$$

where  $\Gamma_t = \mathbf{C}_t + \Phi \Gamma_{t+1}$ ,  $\Gamma_T = [\ell_A, \ell_D, 0, 0, 0, 0]$ ;  $\mathbf{C}_t = [c_A, c_D, 0, 0, 0, 0]$

$$\Phi = \begin{bmatrix} \frac{P_{AA}}{1 + R_A} & \frac{P_{AD}}{1 + R_D} & 0 & 0 & 0 & 0 \\ \frac{P_{DA}}{1 + R_A} & \frac{P_{DD}}{1 + R_D} & 0 & 0 & 0 & 0 \\ -\frac{P_{AA}}{(1 + R_A)^2} & 0 & \frac{P_{AA}}{1 + R_A} & 0 & \frac{P_{AD}}{1 + R_D} & 0 \\ 0 & -\frac{P_{AD}}{(1 + R_D)^2} & 0 & \frac{P_{AA}}{1 + R_A} & 0 & \frac{P_{AD}}{1 + R_D} \\ -\frac{P_{DA}}{(1 + R_A)^2} & 0 & \frac{P_{DA}}{1 + R_A} & 0 & \frac{P_{DD}}{1 + R_D} & 0 \\ 0 & -\frac{P_{DD}}{(1 + R_D)^2} & 0 & \frac{P_{DA}}{1 + R_A} & 0 & \frac{P_{DD}}{1 + R_D} \end{bmatrix}$$

A solution to this system of equations is found similarly by backward recursion. Namely, for a time-invariant coupon payout, we have:

$$\Gamma_{T-n} = \sum_{j=1}^n [\Phi]^j \mathbf{C} + [\Phi]^n \Gamma_T \mathbf{L}, \quad n = 0, 1, 2, \dots$$

while for a nonpaying coupon bond, we have:

$$\Gamma_{T-n} = [\Phi]^n \Gamma_T, \quad n = 0, 1, 2, \dots$$

These equations can be solved numerically providing thereby a combined estimate of rated bond prices and their yield sensitivity.

Generally, we can also calculate rated bonds' duration and their 'cross-duration'. Bond duration is now defined in terms of partial durations, expressing the effects of all yields rates. Explicitly, say that a bond is rated  $i$  at time  $t$  and for simplicity let the yields be time-invariant. The duration of a bond rated  $i$  with respect to its yield is denoted by  $D_{ii}(t, T)$ ,  $i = 1, 2, \dots, m$ , with,

$$D_{ii}(t, T) = -\frac{1}{B_{i,t}} \frac{\partial B_{i,t}}{\partial R_i}$$

while the partial duration, of the bond rated  $i$  with respect to any other yield,  $R_j$ ,  $i \neq j$  is:

$$D_{ij}(t, T) = -\frac{1}{B_{i,t}} \frac{\partial B_{i,t}}{\partial R_j} = -\frac{\partial (\log B_{i,t})}{\partial R_j}; \quad i \neq j$$

By the same token, for convexity we have:

$$\Upsilon_{ii}(t, T) = \frac{1}{B_{i,t}} \frac{\partial^2 B_{i,t}}{\partial R_i^2} \quad \text{and} \quad \Upsilon_{ij}(t, T) = \frac{1}{B_{i,t}} \frac{\partial^2 B_{i,t}}{\partial R_i \partial R_j}$$

The partial durations and convexities express the sensitivity of a bond rated  $i$  is thus:

$$\frac{dB_{i,t}}{B_{i,t}} = \frac{1}{B_{i,t}} \frac{\partial B_{i,t}}{\partial t} dt + \frac{1}{B_{i,t}} \sum_{j=1}^m \frac{\partial B_{i,t}}{\partial R_j} dR_j + \frac{1}{2} \frac{1}{B_{i,t}} \sum_{j=1}^m \frac{\partial^2 B_{i,t}}{\partial R_j \partial R_i} dR_i dR_j$$

Or:

$$\frac{dB_{i,t}}{B_{i,t}} - \frac{\partial \log B_{i,t}}{\partial t} dt = - \sum_{j=1}^m D_{ij} dR_j + \frac{1}{2} \sum_{j=1}^m \Upsilon_{ij} dR_i dR_j$$

### Example

Consider a bond with three ratings (1, 2 and 3) and assume constant yields for each given by  $(R_1, R_2, R_3) = (0.05; 0.07; 0.10)$ . Let the ratings transition matrix be:

$$\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.05 & 0.8 & 0.15 \\ 0.00 & 0.05 & 0.95 \end{bmatrix}$$

Then the bond recursive equation for a pure bond paying \$1 in two periods is:

$$B_{i,t} = \sum_{j=1}^m \frac{P_{ij}}{1 + R_j} B_{j,t+1}; \quad B_{i,2} = \ell_i, \quad i = 1, 2, 3, \dots, m, \quad t = 0, 1, 2$$

For  $m = 3$ , this is reduced to:

$$B_{1,0} = (0.81q_{11} + 0.005q_{12})\ell_1 + (0.09q_{12} + q_{22}0.08)\ell_2 + 0.015q_{23}\ell_3$$

$$B_{2,0} = (0.045q_{11} + 0.04q_{12})\ell_1 + (0.005q_{12} + 0.64q_{22} + 0.0075q_{23})\ell_2 \\ + (0.12q_{23} + 0.1425q_{33})\ell_3$$

$$B_{3,0} = (0.0025q_{12})\ell_1 + (0.04q_{22} + 0.0475q_{23})\ell_2 + (0.0075q_{23} + 0.9025q_{33})\ell_3$$

with the notation:

$$q_{ij} = \frac{1}{1 + R_i} \frac{1}{1 + R_j}$$

and,

$$q_{11} = .9068, \quad q_{12} = 0.8899, \quad q_{13} = 0.8656, \\ q_{22} = 0, \quad q_{23} = 0.8494, \quad q_{33} = 0.8262$$

Thus,

$$\begin{aligned} B_{1,0} &= 0.7389\ell_1 + 0.09788\ell_2 + 0.0127\ell_3 \\ B_{2,0} &= 0.0763\ell_1 + 0.5695\ell_2 + 0.2196\ell_3 \\ B_{3,0} &= 0.00222\ell_1 + 0.0752\ell_2 + 0.7456\ell_3 \end{aligned}$$

Therefore, a bond rated '1' is worth more than a bond rated '2' and a '2' is worth more than a '3' if  $B_{1,0} > B_{2,0} > B_{3,0}$ . Their difference accounts for the yield differential associated with each bond rating. Of course, if the rated bond is secured throughout the two periods (i.e. it does not switch from class to class), we have:

$$\begin{aligned} B_{1,0} &= \frac{1}{(1 + R_1)^2} = 0.907; \quad B_{2,0} = \frac{1}{(1 + R_2)^2} = 0.8734; \\ B_{3,0} &= \frac{1}{(1 + R_3)^2} = 0.8264 \end{aligned}$$

The difference between these numbers accounts for a premium implied by the ratings switching matrix. For the bond rated '3', we note that the secured 'rate 3' bond is worth less than the rated bond, accounting for the potential gain in yield if the bond credit quality is improved. Alternatively, we can use the risk-free discount rate  $R_f = 0.04$  assumed for simplicity to equal 4% yearly (since the bond has no default risk). In this case,

$$B_{f,0} = \frac{1}{(1 + R_f)^2} = \frac{1}{(1 + 0.04)^2} = 0.9245$$

The premium for such a risk-free bond compared to a secured bond rated '1', is  $0.9245 - 0.9071$ .

### 8.4.3 Pricing rated bonds and the term structure risk-free rates\*

When the risk-free term structure is available, and assuming no arbitrage, we can construct a portfolio replicating the bond, thereby valuing the rated bond yields for each bond class. Explicitly, consider a portfolio of rated bonds consisting of  $N_i$ ,  $i = 1, 2, 3, \dots, m$  bonds rated  $i$ , each providing  $\ell_i$  dollars at maturity. Let the portfolio value at maturity be equal to \$1. That is to say

$$\sum_{i=1}^m N_i B_{i,T} = \sum_{i=1}^m N_i \ell_i = 1$$

One period (year) prior to maturity, such a portfolio would be worth

$$\sum_{i=1}^m N_i B_{i,T-1}$$

dollars. By the same token, if we denote by  $R_{f,T-1}$  the risk-free discount rate for one year, then assuming no arbitrage, one period prior to maturity, we have:

$$\sum_{i=1}^m N_i B_{i,T-1} = \frac{1}{1 + R_{f,T-1}}; B_{i,T-1} = c_{it} + \sum_{j=1}^m q_{j,T-1} p_{ij} B_{j,T};$$

$$B_{i,T} = \ell_i, \quad i = 1, 2, \dots, m$$

with  $q_{jt} = 1/(1 + R_{j,t})$  and  $R_{j,t}$  is the one-period discount rate applied to a  $j$ -rated bond. This system of equations provides  $2m$  unknown rates and the portfolio composition with only one equation is therefore under-determined. For two periods we have an additional equation:

$$\sum_{i=1}^m N_i B_{i,T-2} = \frac{1}{(1 + R_{f,T-1})^2}$$

While the bond price is given by:

$$B_{i,T-2} = c_{i,T-2} + \sum_{j=1}^m q_{j,T-2} p_{ij} B_{j,T-1}; B_{i,T} = \ell_i, \quad i = 1, 2, \dots, m$$

as well as:

$$B_{i,T-2} = c_{i,T-2} + \sum_{j=1}^m q_{j,T-2,2} p_{ij}^{(2)} B_{j,T}; B_{i,T} = \ell_i, \quad i = 1, 2, \dots, m$$

where  $p_{ij}^{(2)}$  is the probability that the bond is rated  $j$  two periods hence while  $q_{j,T-2,2}$  is the discount rate for a  $j$ -rated bond for two periods forward (that might differ from the rate  $q_{j,T-2,1}$  applied for one period only). Here again, we see that there are  $2m$  rates while there are only two equations. For three periods we will have three equations per rating and so on. Generally,  $k$  periods prior to maturity, assuming no arbitrage, we have the following conditions for no arbitrage:

$$\sum_{i=1}^m N_i B_{i,T-k} = \frac{1}{(1 + R_{f,T-k})^k} \quad k = 0, 1, 2, 3, \dots, T$$

$$B_{i,T-k} = c_{i,T-k} + \sum_{j=1}^m q_{j,T-k,h} p_{ij}^{(h)} B_{j,T-(k-h)}; B_{i,T} = \ell_i,$$

$$i = 1, 2, \dots, m; \quad k = 1, 2, 3, \dots, T; \quad h = 1, 2, 3, \dots, k$$

where  $R_{f,T-k}$ ,  $k = 1, 2, 3, \dots$ , is the risk-free rate term structure and  $p_{ij}^{(h)}$  is the  $ij$  entry of the  $h$ -power of the rating matrix. These provide a system of  $T + 1$  simultaneous equations spanning the bond life. In matrix notation this is given by:

$$\mathbf{N} \mathbf{B}_{T-k} = \frac{1}{(1 + R_{f,T-k})^k} \quad k = 0, 1, 2, \dots, T; \quad \mathbf{N} = (N_1, \dots, N_m);$$

$$\mathbf{B}_{T-k} = (B_{1,T-k}, \dots, B_{m,T-k})$$

as well as:

$$\mathbf{B}_{t-k} = \mathbf{c}_{t-k} + \mathbf{F}_t^{(h)} \mathbf{B}_{t+(k-h)}; \mathbf{B}_T = \mathbf{L},$$

$$\mathbf{F}_t^{(h)} = \left[ q_{j,T-k,h} p_{ij}^{(h)} \right]; h = 1, 2, \dots, k; k = 1, 2, \dots, T$$

This renders the estimation of the term structure of ratings discount grossly under-determined. However, some approximations can be made which may be acceptable practically. Such an approximation consists in assuming that the rates at a given time are assumed time-invariant and the term structure of risk-free rates is known and we only estimate the short ratings discount. We assume first the case of a maturity larger than the number of rating classes.

**Case  $T \geq 2m$**

When the bond maturity is larger than the number of ratings  $T \geq 2m$ , and  $q_{j,T-k,h} = q_{j,h}$ ,  $h = 1$  and  $q_{j,1} = q_j$ , the hedging portfolio of rated bonds is found by a solution of the system of linear equations above (with  $h = 1$ ), leading to the unique solution:

$$\mathbf{N}^* = \mathfrak{S}^{-1} \mathbf{\Omega}$$

where  $\mathfrak{S}$  is the matrix transpose of  $[B_{i,T-j+1}]$  and  $\mathbf{\Omega}$  is a column vector with entries  $[1/(1 + R_{f,T-s})^s]$ ,  $s = 0, 1, 2, \dots, m - 1$ . Explicitly, we have:

$$\begin{bmatrix} N_1 \\ N_2 \\ \dots \\ N_m \end{bmatrix} = \begin{bmatrix} B_{1,T} & B_{2,T} & B_{3,T} & \dots & B_{m,T} \\ B_{1,T-1} & B_{2,T-1} & B_{3,T-1} & \dots & B_{m,T-1} \\ \dots & \dots & \dots & \dots & \dots \\ B_{1,T-m} & B_{2,T-m} & B_{3,T-m} & \dots & B_{m,T-m} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1/(1 + R_{f,T-1})^1 \\ \dots \\ 1/(1 + R_{f,T-m})^m \end{bmatrix}$$

Thus, the *condition for no arbitrage* is reduced to satisfying a system of system of nonlinear equations:

$$\mathfrak{S}^{-1} \mathbf{\Omega} \mathbf{B}_{T-k} = \frac{1}{(1 + R_{f,T-k})^k}; k = m, m + 1, \dots, T$$

For example, for a zero-coupon rated bond and stationary short discounts, we have  $\mathbf{B}_{t-k} = (\mathbf{F})^k \mathbf{L}$  and therefore, the no-arbitrage condition becomes:

$$\mathfrak{S}^{-1} \mathbf{\Omega} (\mathbf{F})^k \mathbf{L} = \frac{1}{(1 + R_{f,T-k})^k}; k = m, m + 1, \dots, T$$

where  $\mathbf{F}$  has entries  $q_j p_{ij}$ . This provides, therefore,  $T + 1 - m$  equations applied to determining the bond ratings short (one period) discount rates  $q_j$ .

Our system of equations may be over- or under-identified for determining the ratings discount rates under our no-arbitrage condition. Of course, if  $T + 1 - m = m$ , we have exactly  $m$  additional equations we can use to solve the ratings discount rates uniquely (albeit, these are nonlinear equations and can be solved only numerically). If  $(T \geq 2m + 1)$  we can use the remaining equations to calculate some of the term structure discounts of bond ratings as well. For

example, for a bond with maturity three times the number of ratings,  $T = 3m$ , we have the following no-arbitrage condition:

$$\mathfrak{S}^{-1}\mathbf{\Omega}\mathbf{B}_{T-k} = \frac{1}{(1 + R_{f,T-k})^k}; \quad k = m, m + 1, \dots, T$$

and

$$B_{i,T-k} = c_{i,T-k} + \sum_{j=1}^m q_{j,1} p_{ij}^{(1)} B_{j,T-(k-1)}; \quad B_{i,T} = \ell_i,$$

$$B_{i,T-k} = c_{i,T-k} + \sum_{j=1}^m q_{j,2} p_{ij}^{(2)} B_{j,T-(k-2)}; \quad B_{i,T} = \ell_i,$$

$$i = 1, 2, \dots, m; \quad k = 1, 2, 3, \dots, T$$

Thus, when the bond maturity is very large (or if we consider a continuous-time bond), an infinite number of equations is generated which justifies the condition for no arbitrage stated by Jarrow *et al.* (1997).

When data regarding the risk-free term structure is limited, or for short bonds, we have,  $m \leq T < 2m$  and the Markov model is incomplete. We must, therefore, proceed to an approach that can, nevertheless, provide an estimate of the ratings discount rates. We use for convenience a sum of squared deviations from the rated bond arbitrage condition, in which case we minimize the following expression (for estimating the short discount rates only):

$$\text{Minimize}_{0 \leq q_1, q_2, \dots, q_{m-1}, q_m \leq 1} \sum_{k=m}^T \left( \mathfrak{S}^{-1}\mathbf{\Omega}\mathbf{B}_{T-k} - \frac{1}{(1 + R_{f,T-k})^k} \right)^2$$

subject to a number of equalities used in selecting the portfolio, namely:

$$B_{i,T-k} = c_{i,T-k} + \sum_{j=1}^m q_{j,1} p_{ij}^{(1)} B_{j,T-(k-1)}; \quad B_{i,T} = \ell_i, \quad k = 1, 2, \dots, T$$

Additional constraints, reflecting expected and economic rationales of the ratings discounts  $q_j$  might be added, such as:

$$0 \leq q_j \leq 1 \text{ as well as } 0 \leq q_m \leq q_{m-1} \leq q_{m-2} \leq q_{m-3}, \dots, \leq q_2 \leq q_1 \leq 1$$

These are typically nonlinear optimization problems, however. A simple two-ratings problem and other examples are considered to highlight the complexities in determining both the hedging portfolio and the ratings discounts provided the risk-free term structure is available.

### Example: Valuation of a two-rates rated bond

For a portfolio of two-rates bonds over one period where  $\ell_1 = 1$ ,  $\ell_2 = 0.2$  we have the following two equations that can be used to calculate the risk-free

portfolio composition:

$$N_1 + 0.2N_2 = 1 \quad \text{or} \quad N_2 = 5(1 - N_1)$$

$$N_1 B_{1,T-1} + 5(1 - N_1)B_{2,T-1} = \frac{1}{1 + R_{f,T-1}}$$

and,

$$N_1 = \frac{B_{2,T-1}(1 + R_{f,T-1}) - 0.2}{(1 + R_{f,T-1})(B_{2,T-1} - 0.2B_{1,T-1})}$$

If we assume a bond of maturity of three periods only, then only two additional equations are available ( $T - 2, T - 3$ ) providing a no-arbitrage estimate for the rated bond discounts and given by:

$$\begin{aligned} & \left( \frac{B_{2,T-1}(1 + R_{f,T-1}) - 0.2}{(1 + R_{f,T-1})(B_{2,T-1} - 0.2B_{1,T-1})} \right) \\ & \quad \times (B_{1,T-2} - 5B_{2,T-2}) + 5B_{2,T-2} = \frac{1}{(1 + R_{f,T-2})^2} \\ & \left( \frac{B_{2,T-1}(1 + R_{f,T-1}) - 0.2}{(1 + R_{f,T-1})(B_{2,T-1} - 0.2B_{1,T-1})} \right) \\ & \quad \times (B_{1,T-3} - 5B_{2,T-3}) + 5B_{2,T-3} = \frac{1}{(1 + R_{f,T-3})^3} \end{aligned}$$

$$B_{1,T-1} = c_{1,T-1} + q_1 p_{11} + 0.2q_2 p_{12};$$

$$B_{2,T-1} = c_{2,T-1} + q_1 p_{21} + q_2 p_{22} 0.2$$

$$B_{1,T-2} = c_{1,T-2} + q_1 p_{11} B_{1,T-1} + q_2 p_{12} B_{2,T-1};$$

$$B_{2,T-2} = c_{2,T-2} + q_1 p_{21} B_{1,T-1} + q_2 p_{22} B_{2,T-1}$$

$$B_{1,T-3} = [c_{1,T-3} + q_1 p_{11} c_{1,T-2} + q_2 p_{12} c_{2,T-2}]$$

$$+ [q_1^2 p_{11}^2 + q_1 q_2 p_{12} p_{21}] B_{1,T-1} + [q_2 q_1 p_{11} p_{12} + q_2^2 p_{12} p_{22}] B_{2,T-1}$$

$$B_{2,T-3} = [c_{2,T-3} + q_1 p_{21} c_{1,T-2} + q_2 p_{22} c_{2,T-2}]$$

$$+ [q_1 q_1 p_{21} p_{11} + q_1 q_2 p_{12} p_{21}] B_{1,T-1} + [q_2 q_1 p_{21} p_{22} + q_2^2 p_{22}^2] B_{2,T-1}$$

where,  $B_{1,T-1}, B_{1,T-2}, B_{1,T-3}$  are functions of the bond redemption values and the ratings discount rates  $q_1$  and  $q_2$ . That is to say, we have a system of two independent equations in two unknowns only that we can solve by standard numerical analysis. For example, consider a zero-coupon bond with a rating matrix given by:  $p_{11} = 0.8, p_{12} = 0.2, p_{21} = 0.1, p_{22} = 0.9$ . In addition, set  $\ell_1 = 1, \ell_2 = 0.6$  (and therefore a recuperation rate of 60 % on bond default)

and  $R_{f,T-1} = 0.07$ ,  $R_{f,T-2} = 0.08$  thus:

$$B_{1,T-1} = 0.8q_1 + 0.12q_2; B_{2,T-1} = 0.1q_1 + 0.54q_2$$

$$B_{1,T-2} = 0.8q_1 B_{1,T-1} + 0.2q_2 B_{2,T-1}; \quad B_{2,T-2} = 0.1q_1 B_{1,T-1} + 0.9q_2 B_{2,T-1}$$

$$B_{1,T-3} = [0.64q_1^2 + 0.02q_1q_2] B_{1,T-1} + [0.16q_2q_1 + 0.18q_2^2] B_{2,T-1}$$

$$B_{2,T-3} = [0.08q_1^2 + 0.02q_1q_2] B_{1,T-1} + [0.09q_2q_1 + 0.81q_2^2] B_{2,T-1}$$

and therefore

$$\begin{aligned} & \left( \frac{1.07B_{2,T-1} - 0.6}{1.07((0.8q_1 - 0.6)B_{1,T-1} + 0.2q_2B_{2,T-1})} \right) \\ & \quad \times (0.6334q_1B_{1,T-1} - 1.2294.2q_2B_{2,T-1}) \\ & \quad + (0.1666q_1B_{1,T-1} + 1.4994q_2B_{2,T-1}) = 0.8733 \\ & \left( \frac{1.07B_{2,T-1} - 0.6}{1.07((0.8q_1 - 0.6)B_{1,T-1} + 0.2q_2B_{2,T-1})} \right) \\ & \quad \times \left( [0.5072q_1^2 - 0.01332q_1q_2] B_{1,T-1} + \right. \\ & \quad \left. + [0.01006q_2q_1 - 1.16946q_2^2] B_{2,T-1} \right) \\ & \quad + ([0.13328q_1^2 + 0.03332q_1q_2] B_{1,T-1} + [0.14994q_2q_1 + 1.3446q_2^2] B_{2,T-1}) \\ & \quad = 0.7937 \end{aligned}$$

This is a system of six equations six unknowns that can be solved numerically by the usual methods.

#### 8.4.4 Valuation of default-prone rated bonds\*

We consider next the more real and practical case consisting in a bonds defaulting prior to maturity and generally we consider the first time  $n$ , a bond rated initially  $i$ , is rated  $j$  and let the probability of such an event be denoted by,  $f_{ij}(n)$ . This probability equals the probability of not having gone through a  $j$ th rating in prior transitions and be rated  $j$  at time  $n$ . For transition in one period, this is equal to the transition bond rating matrix (S&P or Moody's matrix), while for a transition in two periods it equals the probability of transition in two periods conditional on not having reached rating  $j$  in the first period. In other words, we have:

$$f_{ij}(1) = p_{ij}(1) = p_{ij}; f_{ij}(2) = p_{ij}(2) - f_{ij}(1)p_{jj}$$

By recursion, we can calculate these probabilities:

$$f_{ij}(n) = p_{ij}(n) - \sum_{k=1}^{n-1} f_{ij}(k)p_{jj}(n-k)$$

The probability of a bond defaulting prior to time  $n$  is thus,

$$F_{km}(n - 1) = \sum_{j=1}^{n-1} f_{km}(j)$$

while the probability that such a bond does not default is:

$$\bar{F}_{km}(n - 1) = 1 - F_{km}(n - 1)$$

At present, denote by  $\Phi_i(n)$  the probability that the bond is rated  $i$  at time  $n$ . In vector notation we write,  $\bar{\Phi}(n)$ . Thus, given the rating matrix,  $[\mathbf{P}]$  we have:

$$\bar{\Phi}(n) = [\mathbf{P}]' \bar{\Phi}(n - 1), n = 1, 2, 3, \dots, \quad \text{and} \quad \bar{\Phi}(0) \text{ given}$$

with  $[\mathbf{P}]'$ , the matrix transpose. Thus, at time  $n$ ,  $\bar{\Phi}(n) = [\mathbf{P}]'^n \bar{\Phi}(0)$  and the present value of a coupon payment (given that there was no default at this time) is therefore discounted at  $R_{j,n}$ ,  $q_{j,n} = 1/(1 + R_{j,n})$  if the bond is rated  $j$ . In other words, its present value is:

$$\sum_{j=1}^{m-1} c_{j,n} q_{j,n}^n \Phi_{j,n}; \quad \Phi_{j,n} = \sum_{i=1}^{m-1} \Phi_{i,0} p_{ij}^{(n)}$$

where  $p_{ij}^{(n)}$  is the  $ij$ th entry of the transpose power matrix  $[\mathbf{P}]'^n$  and  $\Phi_{i,0}$  is the probability that initially the bond is rated  $i$ .

When a coupon-bearing default bond rated  $i$  at time  $s$ , defaults at time,  $s + 1$ ,  $T - (s + 1)$  periods before maturity with probability  $f_{im}(s + 1 - s) = f_{im}(1)$ , we have a value:

$$V_{s,i} = (c_{i,T-s} + q_i^{\ell_{m,T-(s+1)}}) \text{ w.p. } f_{im}(1)$$

If such an even occurs at time,  $s + 2$ , with probability  $f_{im}(s + 2 - s) = f_{im}(2)$ , we have:

$$V_{s,i} = \left( c_{i,T-s} + \sum_{k=1}^{m-1} q_k c_{k,T-(s+1)} \Phi_{k,(s+1)-s} + q_i^2 \ell_{m,T-(s+2)} \right) \text{ w.p. } f_{im}(2)$$

where

$$\Phi_{k,1} = \sum_{i=1}^{m-1} \Phi_{i,0} p_{ik}^{(1)}$$

and  $\Phi_{i,0}$  is a vector whose entries are all zero except at  $i$  (since at  $s$  we conditioned the bond value at being rated  $i$ ). By the same token three periods hence and prior to maturity, we have:

$$V_{s,i} = \left( c_{i,T-s} + \sum_{k=1}^{m-1} q_k c_{k,T-(s+1)} \Phi_{i,0} p_{ik}^{(1)} + \sum_{k=1}^{m-1} q_k^2 c_{k,T-(s+2)} \Phi_{i,0} p_{ik}^{(2)} + q_i^3 \ell_{m,T-(s+3)} \right) \text{ w.p. } f_{im}(3)$$

and, generally, for any period prior to maturity,

$$V_{s,i} = \left( c_{i,T-s} + \sum_{\theta=1}^{\tau-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} p_{ik}^{(\theta)} + q_i^\tau \ell_{m,T-(s+\tau)} \right) \text{w.p. } f_{im}(\tau)$$

In expectation, if the bond defaults prior to its maturity, its expected price at time  $s$  is,

$$EB_{i,D}(s, T) = c_{i,T-s} + \sum_{\tau=1}^{T-s} \left( q_i^\tau \ell_{m,T-(s+\tau)} + \sum_{\theta=1}^{\tau-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} p_{ik}^{(\theta)} \right) f_{im}(\tau)$$

With  $\ell_{m,T-j}$ , the bond recovery when the bond defaults, assumed to be a function of the time remaining for the faultless bond to be redeemed. And therefore, the price of such a bond is:

$$B_{i,ND}(s, T) = \left( c_{i,T-s} + \sum_{k=1}^{m-1} q_k^{T-s} \ell_k \Phi_{i,0} p_{ik}^{(T-s)} + \sum_{\theta=1}^{T-s-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} p_{ik}^{(\theta)} \right) \left[ 1 - \sum_{u=1}^{T-s} f_{im}(u) \right]$$

where  $\ell_i$  denotes the bond nominal value at redemption when it is rated  $i$ . Combining these sums, we obtain the price of a default prone bond rated  $i$  at time  $s$ :

$$B_i(s, T) = c_{i,T-s} + \left( c_{i,T-s} + \sum_{k=1}^{m-1} q_k^{T-s} \ell_k \Phi_{i,0} p_{ik}^{(T-s)} + \sum_{\theta=1}^{T-s-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} p_{ik}^{(\theta)} \right) \left[ 1 - \sum_{u=1}^{T-s} f_{im}(u) \right] + \sum_{\tau=1}^{T-s} \left( q_i^\tau \ell_{m,T-(s+\tau)} + \sum_{\theta=1}^{\tau-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} p_{ik}^{(\theta)} \right) f_{im}(\tau)$$

For a zero-coupon bond, this is reduced to:

$$B_i(s, T) = \left( \sum_{k=1}^{m-1} q_k^{T-s} \ell_k \Phi_{i,0} p_{ik}^{(T-s)} \right) \left[ 1 - \sum_{u=1}^{T-s} f_{im}(u) \right] + \sum_{\tau=1}^{T-s} (q_i^\tau \ell_{m,T-(s+\tau)}) f_{im}(\tau)$$

To determine the (short) price discounts rates for a default-prone rated bond we can proceed as we have before by constructing a hedged portfolio consisting of  $N_1, N_2, \dots, N_{m-1}$  shares of bonds rated,  $i = 1, 2, \dots, m - 1$ . Again, let  $R_{f,T-u}$  be the risk-free rate when there are  $u$  periods left to maturity. Then, assuming no

arbitrage and given the term structure risk-free rate, we have:

$$\sum_{i=1}^{m-1} N_i B_i(s, T) = \frac{1}{(1 + R_{f,T-s})^s}, \quad s = 0, 1, 2, \dots$$

with  $B_i(s, T)$  defined above. Note that the portfolio consists of only  $m - 1$  rated bonds and therefore, we have in fact  $2m - 1$  variables to be determined based on the risk-free term structure. When the system is over-identified (i.e. there are more terms in the risk-free term structure than there are short ratings discount to estimate), additional equations based on the ratings discount term structure can be added so that we obtain a sufficient number of equations. Assuming that our system is under-determined (which is the usual case), i.e.  $T \leq 2m - 1$ , we are reduced to solving the following minimum squared deviations problem:

$$\text{Minimize} \quad \sum_{s=0}^T \left( \sum_{k=1}^{m-1} N_k B_k(s, T) - \frac{1}{(1 + R_{f,T-s})^s} \right)^2$$

$0 \leq q_1 \leq q_2 \leq \dots \leq q_{m-1} \leq 1;$   
 $N_1, N_2, N_3, \dots, N_{m-1}$

subject to:

$$B_i(s, T) = c_{i,T-s} + \left( c_{i,T-s} + \sum_{k=1}^{m-1} q_k^{T-s} \ell_k \Phi_{i,0} P_{ik}^{(T-s)} \right. \\ \left. + \sum_{\theta=1}^{T-s-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} P_{ik}^{(\theta)} \right) \left[ 1 - \sum_{u=1}^{T-s} f_{im}(u) \right] \\ + \sum_{\tau=1}^{T-s} \left( q_i^\tau \ell_{m,T-(s+\tau)} + \sum_{\theta=1}^{\tau-1} \sum_{k=1}^{m-1} q_k^\theta c_{k,T-(s+\theta)} \Phi_{i,0} P_{ik}^{(\theta)} \right) f_{im}(\tau)$$

This is, of course, a linear problem in  $N_k$  and a nonlinear one in the rated discounts which can be solved analytically with respect to the hedged portfolio (and using the remaining equations to calculate the ratings discount rates). Explicitly, we have:

$$\sum_{s=0}^T \sum_{k=1}^{m-1} N_k B_j(s, T) B_k(s, T) = \sum_{s=0}^T \frac{B_j(s, T)}{(1 + R_{f,T-s})^s}$$

This is a system of linear equation we can solve by:

$$\sum_{k=1}^{m-1} N_k A_{jk} = D_j; \quad j = 1, 2, 3, \dots, m - 1; \quad D_j = \sum_{s=0}^T \frac{B_j(s, T)}{(1 + R_{f,T-s})^s};$$

$$A_{jk} = \sum_{s=0}^T B_j(s, T) B_k(s, T)$$

and, therefore, in matrix notation

$$\mathbf{N}^* \mathbf{A} = \mathbf{D} \rightarrow \mathbf{N}^* = \mathbf{A}^{-1} \mathbf{D}$$

and obtain the replicating portfolio for a risk-free investment. This solution can be inserted in our system of equations to obtain the reduced set of equations for the ratings discount rates of the default bond. A solution can be found numerically.

**Example: A two-rated default bond**

Consider a two-rated zero-coupon bond and define the transition matrix:

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix} \quad \text{with} \quad \mathbf{P}^n = \begin{bmatrix} p^n & 1-p^n \\ 0 & 1 \end{bmatrix}$$

The probability of being in one of two states after  $n$  periods is  $(p^n, 1-p^n)$ . Further,

$$f_{12}(1) = 1-p; \quad f_{12}(2) = p_2^{(2)} - (1)f_{12}(1) = 1-p^2 - (1-p) = p(1-p)$$

Thus, for a no-coupon paying bond, we have:

$$B_i(s, T) = c_{i,T-s} + q^{T-s\ell} \left[ 1 - \sum_{u=1}^{T-s} f_{12}(u) \right] + \sum_{\tau=1}^{T-s} (q^\tau \ell_{2,T-(s+\tau)}) f_{12}(\tau)$$

In particular,

$$B_1(T, T) = \ell$$

$$B_1(T-1, T) = q\ell [1 - f_{12}(1)] + q\ell_{2,0} f_{12}(1)$$

$$B_1(T-2, T) = q^2\ell [1 - f_{12}(1) - f_{12}(2)] + q\ell_{2,1} f_{12}(\tau) + q^2\ell_{m,0} f_{12}(2)$$

$$B_1(T-3, T) = q^3\ell [1 - f_{12}(1) - f_{12}(2) - f_{12}(3)] + (q\ell_{2,2}) f_{12}(1) + q^2\ell_{2,1} f_{12}(2) + q^3\ell_{2,0} f_{12}(3)$$

If we have a two-year bond, then the condition for no arbitrage is:

$$NB_1(T, T) = N\ell = 1 \quad \text{and} \quad N = 1/\ell$$

$$NB_1(T-1, T) = \frac{1}{1 + R_{f,T-1}} \Rightarrow 1 + R_{1,T-1} = \frac{1 + R_{f,T-1}}{1 - (1 - \ell_{2,0}/\ell) f_{12}(1)}$$

If we set,  $\ell_{2,0} = 0$ , implying that when default occurs at maturity, the bond is a total loss, then:

$$1 + R_{1,T-1} = \frac{1 + R_{f,T-1}}{1 - f_{12}(1)} = 1 + \frac{(1-p) + R_{f,T-1}}{p} \quad \text{or}$$

$$R_{1,T-1} = \frac{(1-p) + R_{f,T-1}}{p}$$

This provides an explicit determination of the rated '1' bond in terms of the risk-free rate. If there is no loss ( $p = 1$ ), then,  $R_{1,T-1} = R_{f,T-1}$ . If we have a two-year bond, then the least quadratic deviation cost rating can be applied. Thus,

$$\text{Minimize } \mathbb{Q} = ((1/\ell)B(T-1, T) - (q_{f,T-1}))^2_{0 \leq q \leq 1} + ((1/\ell)B(T-2, T) - (q_{f,T-2}))^2$$

Subject to:

$$B_1(T - 1, T) = q\ell [1 - f_{12}(1)] + q\ell_{2,0} f_{12}(1)$$

$$B_1(T - 2, T) = q^2\ell [1 - f_{12}(1) - f_{12}(2)] + q\ell_{2,1} f_{12}(1) + q^2\ell_{2,0} f_{12}(2)$$

leading to a cubic equation in  $q$  we can solve by the usual methods. Rewriting the quadratic deviation in terms of the discount rate yields:

$$\begin{aligned} \text{Minimize } & (q[1 - f_{12}(1)(1 - (\ell_{2,0}/\ell))] - (q_{f,T-1}))^2 \\ & + (q^2[1 - f_{12}(1) - (1 - \ell_{2,0}/\ell)f_{12}(2)] + q(\ell_{2,1}/\ell)f_{12}(1) - (q_{f,T-2}))^2 \end{aligned}$$

Set

$$\begin{aligned} a &= [1 - f_{12}(1)(1 - (\ell_{2,0}/\ell))]; \quad b = [1 - f_{12}(1) - (1 - \ell_{2,0}/\ell)f_{12}(2)]; \\ c &= (\ell_{2,1}/\ell)f_{12}(1) \end{aligned}$$

Then an optimal  $q$  is found by solving the equation:

$$2q^3b^2 + 3q^2bc + q(a^2 - 2bq_{f,T-2} + c^2) - (aq_{f,T-1} + cq_{f,T-2}) = 0$$

Assume the following parameters,

$$R_{f,T-1} = 0.07; \quad R_{f,T-2} = 0.08, \quad p = 0.8, \quad \ell = 1, \quad \ell_{2,0} = 0.6, \quad \ell_{2,1} = 0.4$$

In this case,

$$f_{12}(1) = 1 - p = 0.2 \quad \text{and} \quad f_{12}(2) = p(1 - p) = 0.16$$

For a one-period bond, we have:

$$1 + R_{1,T-1} = \frac{1 + 0.07}{1 - (0.084)} = 1.168$$

and, therefore, we have a 16.8 % discount,  $R_{1,T-1} = 0.168$ .

**Problem**

An AAA-rated bond has a clause that if it is downgraded to an AA bond, it must increase its coupon payment by 12 % while if it is downgraded to a B bond, then the firm has to redeem the bond in its entirety. For simplicity, say that the firm has only credit-rating classes (AAA, AA, B) and that it cannot default. In this case, how would you value the bond if initially it were an AAA bond?

**Problem**

By how much should a coupon payment be compensated when the bond class is downgraded? Should we sell a bond when it is downgraded? What are the considerations to keep in mind and how can they be justified?

**Example**

For simplicity, consider a coupon-paying bond with two credit ratings, A and D. ‘D’ denotes default and at redemption at time  $T$ , \$1 is paid. Let  $q = 1 - p$  be a constant probability of default. Thus, if default occurs for the first time at  $n \leq T$ ,

the probability that default occurs at any time prior to redemption is given by the geometric distribution  $p^{n-1}(1-p)$  and therefore the conditional probability of default at  $n \leq T$  is given by:

$$f(n | n \leq T) = \frac{p^{n-1}(1-p)}{1-F(T)}, n = 1, 2, \dots, T$$

where  $F(T)$  is the probability of default before or at bond redemption, or

$$F(T) = (1-p) \left( \frac{1-p^{T+1}}{1-p} - 1 \right) = p(1-p^T)$$

As a result, if the yield of this bond is  $y$ , the bond price is given by:

$$B(0) = \left\{ c \sum_{j=1}^T \left( \frac{1}{1+y} \right)^j \left( \frac{p^{j-1}(1-p)}{1-p(1-p^T)} \right) \right\} + \left( \frac{1}{1+y} \right)^T [1 - p(1-p^T)]$$

While for a risk-free bond we have at the risk-free rate:

$$B_f(0) = \left\{ c \sum_{j=1}^{T-1} \left( \frac{1}{1+R_f} \right)^j \right\} + 1 \left( \frac{1}{1+R_f} \right)^T$$

The difference between the two thus measures the premium paid for a rated bond. These expressions can be simplified, however. This is left as an exercise.

### Problem

Following Enron's collapse, Standard and Poors Corp. has said that it plans to rank the companies in the S&P500 stock index for the quality of their public disclosures as investors criticize the rating agency for failing to identify recent bankruptcies. As a result, a complex set of criteria will be established to construct a 'reliability rating' of S&P's own rating. Say that the credit rating of a bond is now given both by its class (AAA, AA, etc.) and by a reliability index, meaning that to each rating, there is an associated probability with the complement probability associated to a bond with lower rating. How would you proceed to integrate this reliability in bond valuation?

### Example: Cash valuation of a rated firm

Earlier we noted that the cash flow of a firm can be measured by a synthetic sum of zero-coupon bonds. If these bonds are rated, then of course it is necessary for cash flow valuation to recognize this rating. Let  $k$ ,  $k = 1, 2, 3, \dots, m$  be the  $m$  rates a firm assumes and let  $q_k(t)$  be the probability that the bond is rated  $k$  at time  $t$ . In vector notation we write  $\bar{q}(t)$  which is given in terms of the rating matrix transpose  $\mathbf{P}'$ . In other words,  $\bar{q}(t) = [\mathbf{P}']^t \bar{q}(0)$ ,  $k = 1, 2, 3, \dots$  and  $\bar{q}(0)$  given. Thus, the NPV of a rated firm is given by:

$$NPV(t | \mathfrak{R}) = \sum_{s=t}^n \sum_{k=1}^m C_s B_k(t, s) q_k(s), \sum_{k=1}^m q_k(s) = 1$$

Our analysis can be misleading, however. Bonds entering in a given state may remain there for a certain amount of time before they switch to another state. Unstable countries and firms transit across rated states more often than say ‘stable countries’ and ‘firms’. Further, they will usually switch to adjacent states or directly to a default state rather than to ‘distant states’. For example, an AAA bond may be rated after some time to an AA bond while it is unlikely that it would transit directly to rating C. It is possible, however, that for some (usually external) reason the bond defaults, even if initially it is highly rated. These possibilities extend the Markov models considered and are topic for further empirical and theoretical study.

### 8.5 INTEREST-RATE PROCESSES, YIELDS AND BOND VALUATION\*

Bonds, derivative securities and most economic time series depend intimately on the interest-rate process. It is therefore not surprising that much effort has been devoted to constructing models that can replicate and predict reliably the evolution of interest rates. There are, of course, a number of such models, each expressing some economic rationale for the evolution of interest rates. So far we have mostly assumed known risk-free interest rates. In fact, these risk-free (discounting) interest rates vary over time following some stochastic process and as a function of the discount period applied. Generally, and mostly for convenience, an interest-rate process  $\{r(t), t \geq 0\}$  is represented by an Ito stochastic differential equation:

$$dr = \mu(r, t) dt + \sigma(r, t) dw$$

where  $\mu$  and  $\sigma$  are the drift and the diffusion function of the process, which may or may not be stationary. Table 8.8 summarizes a number of interest rates models. Note that while Merton’s model is nonstationary (letting the

**Table 8.8**

Author	Drift	Diffusion	Stationary
Merton (1973)	$\beta$	$\sigma$	no
Cox (1975)	0	$\sigma r^{3/2}$	yes
Vasicek (1977)	$\beta(\alpha - r)$	$\sigma$	yes
Dothan (1978)	0	$\sigma r$	yes
Brennan–Schwartz (1979)	$\beta r[\alpha - \ln(r)]$	$\sigma r$	yes
Courtadon (1982)	$\beta(\alpha - r)$	$\sigma r$	yes
March–Rosenfeld (1983)	$\alpha r^{-(1-\delta)} + \beta r$	$\sigma r^{\delta/2}$	yes
Cox–Ingersoll–Ross (1985)	$\beta(\alpha - r)$	$\sigma r^{1/2}$	yes
Chan et al. (1992)	$\beta(\alpha - r)$	$\sigma r^\lambda$	yes
Constantinidis (1992)	$\alpha + \beta r + \gamma r^2$	$\sigma + \gamma r$	yes
Duffie–Kan (1996)	$\beta(\alpha - r)$	$\sqrt{\sigma + \gamma r}$	yes

diffusion-volatility be time-variant), other models have attempted to model this diffusion coefficient. Of course, to the extent that such a coefficient can be modelled appropriately, the technical difficulties encountered when the coefficients are time-variant can be avoided and the model parameters estimated (even though with difficulty, since these are mostly nonlinear stochastic differential equations). Further, note that the greater part of these interest rate models are of the 'mean reversion' type. In other words, over time short-term interest rates are pulled back to some long-run average level. Thus when the short rate is larger than the average long rate, the drift coefficient is negative and vice versa. Black and Karasinski (1991) (see also Sandmann and Sonderman, 1993) have also suggested that interest models can be modelled as well as a lognormal process. Explicitly, let the annual effective interest rate be given by the nonstationary lognormal model:

$$\frac{dr_a(t)}{r_a(t)} = \beta(t) dt + \sigma(t) dW; r_a(0) = r_{a,0}$$

and consider the continuously compounded rate  $R(t) = \ln(1 + r_a(t))$ . An application of Ito's Lemma to this transformation yields also a diffusion process:

$$dR(t) = (1 - e^{-R(t)}) \left[ \theta(t) - \frac{1}{2}(1 - e^{-R(t)})\sigma^2 \right] dt + \sigma dW(t)$$

Another model suggested, and covering a broad range of distributional assumptions, includes the following (Hogan and Weintraub, 1993):

$$dR(t) = R(t) \left[ \theta(t) - a \ln R(t) + \frac{1}{2}\sigma^2 \right] dt + R(t)\sigma dW(t)$$

The valuation of a bond when interest rates are stochastic is difficult because we cannot replicate the bond value by a risk-free rate. In other words, when rates are stochastic there is no unique way to price the bond. Mathematically this means that there are 'many' martingales we can use for pricing the bond and determine its yield (the integral of the spot-rate process). The problem we are faced with is, therefore, to determine a procedure which we can use to select the 'appropriate martingale' which can replicate observed bond prices. Specifically, say that the interest-rate model is defined by a stochastic process which is a function of a vector parameters  $\Lambda$ . In other words, we write the stochastic process:

$$dr = \mu(r, t, \Lambda) dt + \sigma(r, t, \Lambda) dw$$

If this were the case, the theoretical price of a zero-coupon bond paying \$1 at time  $T$  is:

$$B_{Th}(0, T; \Lambda) = E^* \exp \left( - \int_0^T r(u, \Lambda) du \right) = E^* (e^{-y(0, T; \Lambda)T})$$

where  $y(0, T; \Lambda)$  is the yield, a function of the vector parameters  $\Lambda$ . Now assume

that these bond prices can be observed at time zero for a whole set of future times  $T$  and denote these observed values by,  $B_{obs}(0, T)$ . In order to determine the parameters set  $\Lambda$  we must find therefore some mathematical mechanism that would minimize in some manner some function of the ‘error’

$$\Delta_B = B_{obs}(0, T) - B_{Th}(0, T; \Lambda)$$

There are several alternatives to doing so, as well as numerous mathematical techniques we can apply to solving this problem. This is essentially a computational problem (see, for example, Nelson and Siegel, 1987; Wets *et al.*, 2002; Kortanek and Medvedev, 2001; Kortanek, 2003; Delbaen and Lorimier, 1992; Filipovic, 1999, 2000, 2001).

The Nelson and Siegel approach is applied by many banks and consists in estimating the zero-coupon yield curve by fitting for all available bonds data in a sector credit combination the yield curve:

$$\begin{aligned} B_{Th}(0, T; \Lambda) &= E(e^{-r(0, T; \Lambda)T}); r[0, T; \Lambda(\beta_i), i = 1, \dots, 4] \\ &= \beta_0 + (\beta_1 + \beta_2) \left( \frac{1 - e^{-\beta_3 T}}{\beta_3 T} \right) - \beta_2 e^{-\beta_3 T} \end{aligned}$$

where  $r(0, T; \Lambda(\beta_i), i = 1, \dots, 4)$  is the spot rate and  $\Lambda(\beta_i)$  are the model parameters. The Roger Wets approach ([www.episolutions.com](http://www.episolutions.com)) is based upon a Taylor series approximation of the discount function in integral form. It is based on an approximation, and in this sense it shares properties with purely spline methods. Kortanek and Medvedev (2001), however, use a dynamical systems approach for modelling the term structure of interest rates based on a stochastic linear differential equation by constructing perturbation functions on either the unobservable spot interest rate or its integral (the yield) as unknown functions. Functional parameters are then estimated by minimizing a norm of the error comparing computed yields against observed yields over an observation period, in contrast to using the expectation operator for a stochastic process. When applied to a future period, the solved-for spot-rate function becomes the forecast of the unobservable function, while its integral approximates the yield function to the desired accuracy.

Some prevalent methods for computing (extracting) the zeros, curve-fitting procedures, equating the yield curve to observed data in the central bank include, among others: in Canada using the Svensson procedure and David Bolder (Bank of Canada); in Finland the Nelson–Siegel procedure; in France, the Nelson–Siegel, Svensson procedures; in Japan and the USA the banks use smoothing splines etc. (see Kortanek and Medvedev, 2001; Filipovic, 1999, 2000, 2001). Explicit solutions can be found for selected models, as we shall see below when a number of examples are solved. In particular, we shall show that approaches based on the optimal control of selected models can also be used.

### 8.5.1 The Vasicek interest-rate model

The Vasicek model has attracted much attention and is used in many theoretical and empirical studies. Its validity is of course, subject to empirical verification. An analytical study of the Vasicek model is straightforward since it is a classical model used in stochastic analysis (also called the Ornstein–Uhlenbeck process, as we saw in Chapter 4). In Vasicek’s model the interest-rate change fluctuates around a long-run rate,  $\alpha$ . This fluctuation is subjected to random and normal perturbations of mean zero and variance  $\sigma^2 dt$  however.

$$dr = \beta(\alpha - r) dt + \sigma dw$$

This model’s solution at time  $t$  when the interest rate is  $r(t)$  is  $r(u; t)$ :

$$r(u; t) = \alpha + e^{-\beta(u-t)} (r(t) - \alpha) + \sigma \int_t^u e^{-\beta(u-\tau)} dw(\tau)$$

In this theoretical model we might consider the parameters set  $\Lambda \equiv (\alpha, \beta, \sigma)$  as determining a number of martingales (or bond prices) that obey the model above, namely bond prices at time  $t = 0$  can theoretically equal the following:

$$B_{th}(0, T; \alpha, \beta, \sigma) = E^* \exp \left( - \int_0^T r(u; \alpha, \beta, \sigma) du \right)$$

In this simple case, interest rates have a normal distribution with a known mean and variance (volatility) evolution. Therefore

$$\int_0^T r(u, \alpha, \beta, \sigma) du$$

has also a normal probability distribution with mean and variance given by:

$$m(r(0), T) = \alpha T + (1 - e^{-\beta T}) \frac{r(0) - \alpha}{\beta}$$

$$v(r(0), T) = v(T) = \frac{\sigma^2}{2\beta^3} (4e^{-\beta T} - e^{-2\beta T} + 2\beta T - 3)$$

In these equations the variance is independent of the interest rate while the mean is a linear function of the interest which we write by:

$$m(r(0), T) = \alpha \left[ T - \frac{(1 - e^{-\beta T})}{\beta} \right] + r(0) \frac{(1 - e^{-\beta T})}{\beta}$$

This property is called an affine structure and is of course computationally desirable for it will allow a simpler calculation of the desired martingale. Thus, the

theoretical zero-coupon bond price paying \$1  $T$  periods hence can be written by:

$$\begin{aligned}
 B_{th}(0, T; \alpha, \beta, \sigma) &= E \exp \left( \int_0^T r(u, \alpha, \beta, \sigma) du \right) \\
 &= e^{-m(r(0), T) + v(T)/2} = e^{A(T) - r_0 D(T)} \\
 A(T) &= -\alpha \left[ T - \frac{(1 - e^{-\beta T})}{\beta} \right] + \frac{\sigma^2}{4\beta^3} (4e^{-\beta T} - e^{-2\beta T} + 2\beta T - 3); \\
 D(T) &= \frac{(1 - e^{-\beta T})}{\beta}
 \end{aligned}$$

Now assume that a continuous series of bond values are observed and given by  $B_{obs}(0, T)$  which we write for convenience by,  $B_{obs}(0, T) = e^{-R_T T}$ . Without loss of generality we can consider the yield error term given by:

$$\Delta_T = R_T - (A(T) - r_0 D(T))$$

and thus select the parameters (i.e. select the martingale) that is closest in some sense to observed values. For example, a least squares solution of  $n$  observed bond values yields the following optimization problem:

$$\text{Min}_{\alpha, \beta, \sigma} \sum_{i=1}^n (\Delta_i)^2$$

When the model has time-varying parameters, the problem we faced above turns out to have an infinite number of unknown parameters and therefore the yield curve estimation problem we considered above might be grossly underspecified. Explicitly, let the interest rate model be defined by:

$$dr(t) = \beta [\alpha(t) - r(t)] dt + \sigma dw$$

The theoretical bond value has still an affine structure and therefore we can write:

$$B_{th}[t, T; \alpha(t), \beta, \sigma] = E^* \exp \left( - \int_0^T r(u; \alpha, \beta, \sigma) du \right) = e^{A(t, T) - r(t) D(t, T)}$$

The integral interest-rate process is still normal with mean and variance leading to:

$$A(t, T) = \int_t^T \left\{ \frac{1}{2} \sigma^2 D^2(s, T) - \beta \alpha(s) D(s, T) \right\} ds; \quad D(t, T) = \frac{1}{\beta} \{ 1 - e^{-\beta(T-t)} \}$$

or:

$$\frac{dA(t, T)}{dt} = \alpha(t) \{ 1 - e^{-\beta(T-t)} \} - \frac{\sigma^2}{2\beta^2} \{ 1 - e^{-\beta(T-t)} \}^2, \quad A(T, T) = 0$$

in which  $\alpha(t)$ ,  $\beta$ ,  $\sigma$  are unspecified. If we equate this equation to the available bond data we will obviously have far more unknown variables than data points and therefore the yield curve estimate will depend again on the optimization technique we use to generate the best fit parameters  $\beta^*$ ,  $\sigma^*$  and the function,  $\alpha^*(t)$ . Such problems can be formulated as standard problems in the calculus of variations (or optimal control theory). For example, if we consider the observed prices  $B_{obs}(t, T)$ ,  $t \leq T < \infty$ , for a specific time,  $T$ , and minimize the following squared error in continuous time, we obtain the following singular control problem:

$$\text{Min}_{\alpha(u)} = \int_0^t [A(u, T) - c(u, T)]^2 du$$

subject to:

$$\frac{dA(u, T)}{du} = \alpha(u)a(u, T) - b(u, T), \quad A(T, T) = 0$$

with

$$c(u, t) = y_{obs}(u, T) + r(u) \left[ \frac{1}{\beta} \{1 - e^{-\beta(T-u)}\} \right],$$

$$a(u, t) = \{1 - e^{-\beta(T-u)}\}; \quad b(u, t) = \frac{\sigma^2}{2\beta^2} \{1 - e^{-\beta(T-u)}\}^2$$

and  $\alpha(u)$  is the control and  $A(u, T)$  is the state which can be solved by the usual techniques in optimal control. The solution of this problem leads either to a bang-bang solution, or to a singular solution. Using the deterministic dynamic programming framework, the long-run (estimated) rate is given by solving:

$$-\frac{\partial J}{\partial u} = \text{Min}_{\alpha(u)} \left\{ [A(u, T) - c(u, T)]^2 + \frac{\partial J}{\partial A} [\alpha(u)a(u, T) - b(u, T)] \right\}$$

On a singular strip,  $\partial J / \partial A = 0$  where  $a(u, t) \neq 0$  and therefore in order to calculate  $\alpha(u)$ , we can proceed by a change of variables and transform the original control problem into a linear quadratic control problem which can be solved by the standard optimal control methods. Explicitly, set:

$$y(u) = [A(u, T) - c(u, T)]$$

$$\text{with } \frac{dw(u)}{du} = \alpha(u) \quad \text{and} \quad z(u) = y(u) - a(u, T)w(u)$$

Thus, the problem is reduced to:

$$\text{Min}_{w(u)} = \int_0^t [z(u) + a(u, T)w(u)]^2 du$$

subject to:

$$\frac{dz(u)}{du} = -\dot{a}(u, T)w(u) - b(u, T) - c(u, T); \quad \dot{a}(u, T) = da(u, T)/du$$

and at time T,

$$z(T) = -c(T, T) - a(T, T)w(T)$$

This is a linear control problem whose objective is quadratic in both the state and the control. As a result, the problem solution of this standard control problem is the linear feedback form:

$$w(u) = Q(u) + S(u)z(u) \quad \text{or} \quad \alpha(t) = \int_0^t w(u) du$$

The functions  $Q(u)$ ,  $S(u)$  can be found by inserting in the problem's conditions for optimality. This problem is left for self-study, however (see also Tapiero, 2003).

**Problem: The cox-ingersoll-Ross (CIR) model**

By changing the interest-rate model, we change naturally the results obtained. Cox, Ingersoll and Ross (1985), for example, suggested a model, called the *square root process*, which has a volatility given as a function of interest rates as well, namely, they assume that:

$$dr = \beta(\alpha - r) dt + \sigma\sqrt{r}dw$$

First show that the interest rate process is not normal but its mean and variance are given by:

$$E(r(t)|r_0) = c(t) \left[ \frac{4\beta\alpha}{\sigma^2} + \xi \right]; \text{Var}(r(t)|r_0) = c(t)^2 \left[ \frac{8\beta\alpha}{\sigma^2} + 4\xi \right]$$

where

$$c(t) = \frac{\sigma^2}{4\beta} [1 - e^{-\beta t}]; \quad \xi = \frac{4r_0\beta}{\sigma^2 [\exp(\beta t) - 1]}$$

Demonstrate then that this process has an affine structure as well by verifying that:

$$B(r, t, T) = E \exp \left( - \int_t^T r(T - u) du \right) = e^{A(t, T) - rD(t, T)}; \quad B(r, T, T) = 1$$

and at the boundary  $A(T, T) = 0$ ,  $D(T, T) = 0$ . Finally, calculate both  $A(t, T)$  and  $D(t, T)$  and formulate the numerical problem which has to be solved in order to determine the bond yield curve based on available bond prices.

**Problem: The nonstationary Vasicek model**

Show for the nonstationary model  $dr = \mu(t)(m(t) - r) dt + \sigma r dw$  that its solution is:

$$r(t) = \exp[-A(t)] \left[ y + \int_0^t \mu(s)m(s) \exp[-A(s)] ds \right]$$

$$A(t) = M(t) + \sigma^2 t/2 - \sigma \int_0^t dw \quad \text{and} \quad M(t) = \int_0^t \mu(s) ds$$

**8.5.2 Stochastic volatility interest-rate models**

Cotton, Fouque, Papanicolaou and Sircar (2000) have shown that a single factor model (i.e. with one source of uncertainty) driven by Brownian motion implies perfect correlation between returns on bonds for all maturities  $T$ , which is not seen in empirical analysis. They suggest, therefore, that the volatility in the Vasicek model ought to be stochastic as well. Their derivation, based on a mean reverting model in the short rate, shows an exponential decay in the short-term, (two weeks). This is small compared to bonds with maturities of several years. Denote the variance in an interest model by  $V = \sigma^2(r, t)$ , then an interest-rate ‘stochastic volatility model’ consists of two stochastic differential equations, with two sources of risk ( $w_1, w_2$ ) which may be correlated or not. An example would be:

$$dr = \mu(r, t) dt + \sqrt{V(r, t)} dw_1$$

$$dV = v(V, r, t) dt + \gamma(V, r) dw_2$$

where the variance  $V$  appears in both equations. Hull and White (1988) for example suggest that we use a square root model with a mean reverting variance model given by:

$$\frac{dr}{r} = \mu dt + \sqrt{V} dw_1; \quad dV = \alpha(\beta - V) dt + \gamma r V^\lambda dw_2, \quad \rho dt = E dw_1 dw_2$$

In this case, note that when stock prices increase, volatility increases. Further when volatility increases, interest rates (or the underlying asset we are modeling) increase as well. Cotton *et al.* (2000), in contrast, suggested that, in a CIR-type model such as  $dr = (\mu - r) dt + \sigma r^\gamma dW$ ,  $\gamma$  is not equal to a half but rather is equal to one and half and thereby certainly greater than one. The model they suggest turns out:

$$dr = \theta_r (\mu_r - r) dt + \sqrt{\alpha_r + \beta_r V} dW^*$$

$$dV = \theta_V (\mu_V - V) dt + \sqrt{\alpha_V + \beta_V V} dZ^*$$

where ( $dW^*, dZ^*$ ) are Brownian motion under the pricing measure. Note here that the volatility is a mean reverting driving process. The advantage in using such a model is that it also leads to an affine structure where the time-dependent

coefficients are given by the solutions of differential equations. In this case, estimation of the yield curve can be reached, as we have stated above, by the solution of an optimal control problem. In other words, once a theoretical estimate of the bond price is found, and observed bond prices are available, we can calculate the parameters of the model by solving the appropriate optimization problem.

### 8.5.3 Term structure and interest rates

Interest rates applied for known periods of time, say  $T$ , change necessarily over time. In other words, if  $r(t, T)$  is the interest rate applied at  $t$  for  $T$ , then at  $t + 1$ , the relevant rate for this period  $T$  would be  $r(t + 1, T - 1)$ , while the going interest for the same period would be  $r(t + 1, T)$ . If these interest rates are not equal, there may be an opportunity for refinancing. As a result, the evolution of interest rates for different maturity dates is important. For example, if a model is constructed for interest rates of maturity  $T$ , then we may write:

$$dr(t, T) = \mu(r, T) dt + \sigma(r, T) dw$$

The price of a zero-coupon bond is a function of such interest rates and is given by  $B(t, T) = \exp[-r(t, T)(T - t)]$  whose differential equation (see the mathematical Appendix to this chapter)

$$0 = \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} [\mu(r, T) - \lambda(r, t)] + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) - rB$$

$$B(r, T, T) = 1$$

where the price of risk, a known function of  $r$  and time  $t$ , is proportional to the returns standard deviation and given by:

$$\alpha(r, t, T) = r + \lambda(r, t) \frac{1}{B} \frac{\partial B}{\partial r}$$

The solution of this equation, although cumbersome, can in some cases be determined analytically, and in others it can be solved numerically. For example, if we set  $(\mu(r, T) - \lambda(r, t)) = \theta; \sigma^2(r, T) = \rho^2$  where  $(\theta, \rho)$  are constant then a solution of the partial differential equation of the bond price (see the Mathematical Appendix), we obtain an affine structure type:

$$B(r, t, T) = \exp \left\{ -r(T - t) - \frac{1}{2} \theta (T - t)^2 + \frac{1}{6} \rho^2 (T - t)^3 \right\}$$

#### Problem

Set the following equalities:  $\mu(r, T) - \lambda(r, t) = k(\theta - r); \sigma^2(r, T) = \rho^2 r$  (which is the CIR model seen earlier) and show that the solution for the bond price equation is of the following form:

$$B(r, t, T) = \exp \{ A(T - t) + rD(T - t) \}$$

A solution for the function  $A(\cdot)$  and  $D(\cdot)$  can be found by substitution.

## 8.6 OPTIONS ON BONDS\*

Options on bonds are compound options, traded popularly in financial markets. The valuation of these options requires both an interest-rate model and the valuation of term structure bond prices (which depend on the interest rates for various maturities of the bond). For instance, say that there is a  $T$  bond call option, which confers the right to exercise it at time  $S < T$ . The procedure we adopt in valuing a call option on a bond consists then in two steps. First we evaluate the term structure for a  $T$  and an  $S$  bond. Then we can proceed to value the call on the  $T$  bond with exercise at time  $S$  (used to replace the spot price at time  $S$  in the plain vanilla option model of Black–Scholes). The procedure is explicitly given by the following. First we construct a hedging portfolio consisting of the two bonds maturities  $S$  and  $T$  ( $S < T$ ). Such a portfolio can generate a synthetic rate, equated to the spot interest rate so that no arbitrage is possible. In this manner, we value the option on the bond uniquely. An extended development is considered in the Mathematical Appendix while here we summarize essential results. Let, for example, the interest process:

$$dr = \mu(r, t) dt + \sigma(r, t) dw$$

A portfolio ( $n_S, n_T$ ) of these two bonds has a value and a rate of return given by:

$$V = n_S B(t, S) + n_T B(t, T) \quad \text{and} \quad \frac{dV}{V} = n_S \frac{dB(t, S)}{B(t, S)} + n_T \frac{dB(t, T)}{B(t, T)}$$

The rates of return on  $T$  and  $S$  bonds are assumed given as in the previous section. Each bond with maturity  $T$  and  $S$  has at its exercise time a \$1 denomination, thus the value of each of these ( $S$  and  $T$ ) bonds is given by  $B_T(t, r)$  and  $B_S(t, r)$ . Given these two bonds, we define the option value of a call on a  $T$  bond with  $S < T$  and strike price  $K$ , to be:

$$X = \text{Max} [B(S, T) - K, 0]$$

with  $B(S, T)$  the price of the  $T$  bond at time  $S$ . The bond value  $B(S, T)$  is of course found by solving for the term structure equation and equating  $B(r, S, T) = B(S, T)$ . To simplify matters, say that the solution (value at time  $t$ ) for the  $T$  bond is given by  $F(t, r, T)$ , then at time  $S$ , this value is:  $F(S, r, T)$  to which we equate  $B(S, T)$ . In other words,

$$X = \text{Max} [F(S, r, T) - K, 0]$$

Now, if the option price is  $B(\cdot)$ , then, as we have seen in the plain vanilla model in Chapter 6, the value of the bond is found by solving for  $P(\cdot)$  in the following partial differential equation:

$$0 = \frac{\partial B}{\partial t} + \mu(r, t) \frac{\partial B}{\partial r} + \sigma^2 \frac{1}{2} \frac{\partial^2 B}{\partial r^2} - rB, \quad B(S, r) = \text{Max} [F(S, r, T) - K, 0]$$

A special case of interest consists again in using an affine term structure (ATS) model as shown above in which case:

$$F(t, r, T) = e^{A(t,T) - rD(t,T)}$$

where  $A(\cdot)$  and  $D(\cdot)$  are calculated by the term structure model. The price of an option of the bond is thus given by the solution of the bond partial differential equation, for which a number of special cases have been solved analytically. When this is not the case, we must turn to numerical or simulation techniques.

### 8.6.1 Convertible bonds

Convertible bonds confer the right to the bond issuer to convert the bond into stock or into certain amounts of money that include the conversion cost. For example, if the bond can be converted against  $m$  shares of stock, whose price dynamics is:

$$dS = \mu S dt + \sigma S dw$$

Then, the bond price is necessarily a function of the stock price and given by  $V(S, t)$ . To value such a bond we proceed 'as usual' by constructing an equivalent risk-free and replicating portfolio. Let this risk-free portfolio be:

$$\pi = V + \alpha S \quad \text{and} \quad \text{therefore} \quad d\pi = dV + \alpha dS$$

For this portfolio to be risk-free, we equate it to a portfolio whose rate of return is the risk-free rate  $R_f$ . Thus,  $d\pi = dV + \alpha dS = R_f \pi dt = R_f (V + \alpha S) dt$  and  $dV = R_f (V + \alpha S) dt - \alpha dS$ . Using Ito's Lemma, we calculate  $dV$  leading to:

$$\frac{\partial V}{\partial t} dt + \left( \frac{\partial V}{\partial S} \right) dS + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} (dS)^2 = R_f (V + \alpha S) dt - \alpha dS$$

which can be rearranged to:

$$\left[ \frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 V}{\partial S^2} + \alpha \mu S - R_f V - \alpha R_f S \right] dt + \sigma S \left( \frac{\partial V}{\partial S} + \alpha \right) dw = 0$$

A risk-free portfolio has no volatility and therefore we require:

$$\sigma S \left( \frac{\partial V}{\partial S} + \alpha \right) dw = 0 \quad \text{or} \quad \alpha = -\frac{\partial V}{\partial S}$$

Inserting  $\alpha = -\partial V/\partial S$  yields the following partial differential equation:

$$\frac{\partial V}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 V}{\partial S^2} + R_f S \frac{\partial V}{\partial S} - R_f V = 0, \quad V(S, T) = 1$$

where at redemption the bond equals \$1. If the conversion cost is  $C(S, t) = mS$ , the least cost is  $\text{Min} \{V(S, t), C(S, t)\}$ . Therefore, in the continuation region (i.e. as long as we do not convert the bonds into stocks), we have:  $V(S, t) \geq C(S, t) = mS$  while in the stopping region (i.e. at conversion) we have:  $V(S, t) \leq mS$ . In

other words, the convertible option has the value of an American option which we solve as indicated in Chapter 6. It can also be formulated as a stopping time problem, but this is left as an exercise for the motivated reader.

### 8.6.2 Caps, floors, collars and range notes

A *cap* is a contract guaranteeing that a floating interest rate is capped. For example, let  $r_\ell$  be a floating rate and let  $r_c$  be an interest rate cap. If we assume that the floating rate equals approximately the spot rate,  $r \approx r_\ell$  then a simple *caplet* is priced by:

$$\frac{\partial V}{\partial t} + [\mu(r, t) - \lambda\sigma] \frac{\partial P}{\partial r} + \sigma^2 \frac{1}{2} \frac{\partial^2 V}{\partial r^2} - rV, \quad V(r, T) = \text{Max} [r - r_c, 0]$$

with the cap being a series of caplets. By the same token, a *floor* ensures that the interest rate is bounded below by the rate floor:  $r_f$ . Thus, the rate at which a cash flow is valued is:  $\text{Max} (r_f - r_\ell, 0)$ ,  $r_\ell \geq r_c$ . Again, if we assume that the floating rate equals the short rate, we have a *floorlet* price given by:

$$\frac{\partial V}{\partial t} + [\mu(r, t) - \lambda\sigma] \frac{\partial P}{\partial r} + \sigma^2 \frac{1}{2} \frac{\partial^2 V}{\partial r^2} - rV, \quad V(r, T) = \text{Max} [r_f - r, 0]$$

while the floor is a series of floorlets. A *collar*, places both an upper and a lower bound on interest payments, however. A collar can thus be viewed as a long position on a cap, with a given strike  $r_c$  and a short position on a floor with a lower strike  $r_f$ . If the interest rate falls below  $r_f$ , the holder is forced into paying the higher rate of  $r_f$ . The strike price of the call is often set up so that the cost of the cap is exactly subsidized by the revenue from the sale of the floor. When the interest rate on a notional principal is bounded above and below, then we have a *range* note. In this case, the value of the range note can be solved by using the differential equations framework as follows:

$$\begin{aligned} & \frac{\partial V}{\partial t} + [\mu(r, t) - \lambda\sigma] \frac{\partial P}{\partial r} + \sigma^2 \frac{1}{2} \frac{\partial^2 V}{\partial r^2} - rV + \Xi(r) \\ & = 0; \quad \Xi(r) = \begin{cases} r & \text{if } r < r < \bar{r} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This is only an approximation since, in practice, the relevant interest rate will have a finite maturity (Wilmott, 2000).

### 8.6.3 Swaps

An *interest rate swap* is a private agreement between two parties to exchange one stream of cash flow for another on a specific amount of principal for a specific period of time. Investors use swaps to exchange fixed-rate liabilities/assets into floating-rate liabilities/assets and vice versa.

Interest rate swaps are most important in practice. They emerged in the 1980s and their growth has been spectacular ever since. They are essentially customized

commodity exchange agreements between two parties to make periodic payments to each other according to well-defined rules. In the simplest of interest rate swaps, one part periodically pays a cash flow determined by a fixed interest rate and receives a cash flow determined by a floating interest rate (Ritchken, lecture notes, 2002).

For example, consider Company A with \$50 000 000 of floating-rate debt outstanding on which it is paying LIBOR plus 150 bps (basis points), i.e. if LIBOR is 4 %, the interest rate would be 5.5 %. The company thinks that interest rates will rise, i.e. company's interest expense will rise, and the company decides to convert its debt from floating-rate into fixed-rate debt. Now consider Company B which has \$50 000 000 of fixed-rate 6 % debt. The company thinks that interest rates will fall, which would benefit the company if it has floating-rate debt instead of fixed-rate debt, since its interest expense will be reduced. By entering into an interest rate swap with Company A, both parties can effectively convert their existing liabilities into the ones they truly want. In this swap, Company A might agree to pay Company B fixed-rate interest payments of 5 % and Company B might agree to pay Company A floating-rate interest payments of LIBOR. Therefore Company A will pay LIBOR + 150 to its original lender and 5 % in the swap, giving a total of LIBOR + 6.5 %; it receives LIBOR in the swap. This leaves an all-in cost of funds of 6.5 %, a fixed rate. In the case of Company B, it pays 6 % to its original lender and LIBOR in the swap, giving a total of LIBOR + 6 %. In return it receives 5 % in the swap, leaving an all-in cost of LIBOR + 1 %, a floating rate (see Figure 8.7).

There are four major components to a swap: the notional principal amount, the interest rate for each party, the frequency of cash exchange and the duration of the swap. A typical swap in swap jargon might be \$20m, two-year, pay fixed, receive variable, semi. Translated, this swap would be for \$20m notional principal, where one party would pay a fixed interest-rate payment for every 6 months based on the \$20m and the counterparty would pay a variable rate payment every 6 months based on the \$20m. The variable-rate payment is usually based on a specific short-term interest rate index such as the 6-months LIBOR. The time period specified by the variable rate index usually coincides with the frequency of swap payments. For example, a swap that is fixed versus 6 months LIBOR would have semiannual payments. Of course there can be exceptions to this rule.

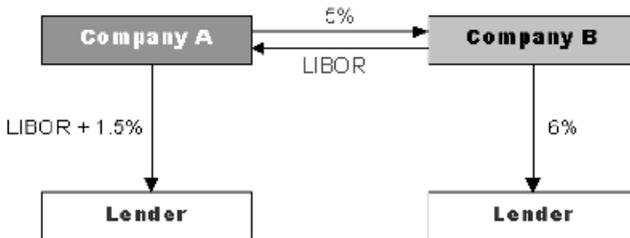


Figure 8.7 A swap contract.

For example, the variable-rate payment could be linked to the average of all T-bill auction rates during the time period between settlements.

Most interest rate swaps have payment date arrears. That is, the net cash flow between parties is established at the beginning of the period, but is actually paid out at the end of the period. The fixed rate for a generic swap is usually quoted as some spread over benchmark US treasuries. For example a quote of '20 over' for a 5-year swap implies that the fixed rate on a 5-year swap will be set at the 5-year Treasury yield that exists at the time of pricing plus 20 basis point. Usually, swap spreads are quoted against the two, three, five, seven and ten benchmark maturities. The yield used for other swaps (such as a 4-year swap) is then obtained by averaging the surrounding yields.

Finally, a 'swaption' is an option to swap. It confers the right to enter into a swap contract at a predetermined future date at a fixed-rate and be a payer at the fixed rate. 'Captions' and 'floortions' are similarly options on caps and floors respectively.

A swap price can be defined by a cap–floor parity, where

$$\text{Cap} = \text{Floor} + \text{Swap}$$

In other words, the price of a swap equals the price difference between the cap and the floor. A caplet can be shown to be equivalent to the price of a put option expiring at a time  $t_{i-1}$  prior to a bond's maturity at time  $t_i$ . Set the payoff  $\text{Max}(r_\ell - r_c, 0)$  a period hence, which is discounted to  $t_{i-1}$  and yielding:

$$\frac{1}{1+r_\ell} \text{Max}(r_\ell - r_c, 0) = \text{Max}\left(\frac{r_\ell - r_c}{1+r_\ell}, 0\right) = \text{Max}\left(1 - \frac{1+r_c}{1+r_\ell}, 0\right)$$

However  $(1+r_c)/(1+r_\ell)$  is the price of paying  $1+r_c$  a period hence. Thus a caplet is equivalent to a put option expiring at  $t_{i-1}$  on a bond with maturity  $t_i$ .

## REFERENCES AND ADDITIONAL READING

- Arrow, K.J. (1953) Le role des valeurs boursières pour la repartition la meilleur des risques, in econometric, *Colloquia International du CNRS*, **40**, 41–47. In English, The role of securities in the optimal allocation of risk bearing, *Review of Economic Studies*, **31**, 91–96, 1963.
- Augros, Jean-Claude (1989) *Les Options sur Taux d'Intérêt*, Economica, Paris.
- Belkin, B., S. Suchower and L. Forest Jr (1998) A one-parameter representation of credit risk and transition matrices, *CreditMetrics® Monitor*, third quarter, 1998. (<http://www.riskmetrics.com/cm/pubs/index.cgi>)
- Bingham, N.H. (1991) Fluctuation theory of the Ehrenfest urn, *Advances in Applied Probability*, **23**, 598–611.
- Black F. and J.C. Cox, (1976), Valuing corporate securities: Some effects of bond indenture provisions, *Journal of Finance*, **31**(2), 351–367.
- Black, F., E. Derman and W. Toy (1990) A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts Journal*, January–February, 133–139.
- Black, F., and P. Karasinski (1991) Bond and options pricing when short rates are lognormal, *Financial Analysts Journal*, **47**, 52–59.

- Black, F., and M. Scholes (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**(3), 637–654.
- Brennan, M.J., and E.S. Schwartz (1977) Convertible bonds: valuation and optimal strategies for call and conversion, *Journal of Finance*, **32**, 1699–1715.
- Brennan, M.J., and E.S. Schwartz (1979) A continuous time approach to the pricing of corporate bonds, *Journal of Banking and Finance*, **3**, 133–155.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders (1992) An empirical comparison of alternative models of the short term interest rate, *Journal of Finance*, **47**, 1209–1227.
- Chance, D. (1990) Default risk and the duration of zero-coupon bonds, *Journal of Finance*, **45**(1), 265–274.
- Cotton, Peter, Jean-Pierre Fouque, George Papanicolaou and K. Ronnie Sircar (2000) Stochastic volatility corrections for interest rate derivatives, Working Paper, Stanford University, 8 May.
- Courtadon, G. (1982) The pricing of options on default free bonds, *Journal of Financial and Quantitative Analysis*, **17**, 75–100.
- Cox, J.C., J.E. Ingersoll and S.A. Ross (1985) A theory of the term structure of interest rates, *Econometrica*, **53**, 385–407.
- Delbaen, F., and S. Lormier (1992) Estimation of the yield curve and forward rate curve starting from a finite number of observations. *Insurance: Mathematics and Economics*, **11**, 249–258.
- Duffie, D., and D. Lando (2001) Term structures of credit spreads with incomplete accounting information, *Econometrica*, **69**, 633–664.
- Duffie, D. and K.J. Singleton (1997) Modeling term structures of defaultable bonds, *Review of Financial Studies*, **12**(4), 687–720.
- Duffie, D. and K.J. Singleton (1999) Modelling term structures of defaultable bonds, *Review of Financial Studies*, **12**(4), 687–720.
- Duffie, G.R. (1998) The relation between Treasury yields and corporate bond yield spreads, *Journal of Finance*, **53**, 2225–2242.
- Duffie, G.R. (1999) Estimating the price of default risk, *Review of Financial Studies*, **12**, 197–226.
- Duffie, J.D., and R. Kan (1996) A yield-factor model of interest rates, *Mathematical Finance*, **6**, 379–406.
- Fabozzi, F.J. (1996) *Bond Markets: Strategies and Analysis*, Prentice Hall, Englewood Cliffs, NJ.
- Fama, E., and K. French (1993) Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, **33**(1), 3–56.
- Filipovic, D. (1999) A note on the Nelson–Siegel family, *Mathematical Finance*, **9**, 349–359.
- Filipovic, D. (2000) Exponential-polynomial families and the term structure of interest rates, *Bernoulli*, **6**, 1–27.
- Filipovic, D. (2001) Consistency problems for Heath–Jarrow–Morton interest rate models, Number 1760 in *Lecture Notes in Mathematics*, Series Editors J.-M. Morel, F. Takens and B. Teissier. Springer Verlag, New York.
- Fisher, I. (1906) *The Nature of Capital and Income*, Sentry Press, New York. Reprinted by Augustus M. Kelly, New York, 1965.
- Fisher, I. (1907), *The Rate of Interest*, Macmillan, New York.
- Fisher, I. (1930) *The Theory of Interest*, Macmillan, New York. Reprinted by Augustus M. Kelly, New York, 1960.
- Flajollet, F., and F. Guillemin (2000) The formal theory of birth–death processes, lattice path combinatorics and continued fractions, *Advances Applied in Probability*, **32**, 750–778.
- Geske, R. (1977) The valuation of corporate liabilities as compound options, *Journal of Financial and Quantitative Analysis*, **12**(4), 541–552.
- Heath, D., R. Jarrow and A. Morton (1990) Contingent claim valuation with a random evolution of interest rates, *The Review of Future Markets*, 54–76.

- Heath, D., R. Jarrow and A. Morton (1992) Bond pricing and the term structure of interest rates: A new methodology for contingent claims evaluation, *Econometrica*, **60**, 77–105.
- Ho, T.S.Y., and S.B. Lee (1986) Term structure movements and pricing of interest rate contingent claims, *Journal of Finance*, **41**, December, 1011–1029.
- Hogan, M., and K. Weintraub (1993) The lognormal interest model and Eurodollars futures, Working Paper, Citibank, New York.
- Hull J. and A. White (1988) An analysis of the bias in option pricing caused by a stochastic volatility, *Advances in Futures and Options Research*, **3**(1), 29–61
- Hull J. and A. White (1993) Bond Option pricing based on a model for the evolution of bond prices, *Advances in Futures and Options Research*, **6**, 1–13.
- Jarrow, Robert A. (1996) *Modelling Fixed Income Securities and Interest Rate Options*, McGraw Hill, New York.
- Jarrow, R., and S. Turnbull (1995) Pricing derivatives on financial securities subject to credit risk, *Journal of Finance*, **50**, 53–86.
- Jarrow, R.A., D. Lando and S. Turnbull (1997) A Markov model for the term structure of credit spreads, *Review of Financial Studies*, **10**, 481–523.
- Karlin, S., and J.L. McGregor (1957) The differential equation of birth–death processes, and the Stieltjes moment problem, *Transactions of the American Mathematical Society*, **85**, 489–546.
- Karlin, S., and H.M. Taylor (1975) *A First Course in Stochastic Processes*, 2nd edn, Academic Press, New York.
- Karlin, S., and H.M. Taylor (1981) *A Second Course in Stochastic Processes*, 2nd edn, Academic Press, New York.
- Kim, J. (1999) Conditioning the transition matrix, *Credit Risk*, a special report by Risk, October.
- Kortanek, K.O. (2003) Comparing the Kortanek & Medvedev GP approach with the recent Wets approach for extracting the zeros, April 26 (Internet paper).
- Kortanek, K.O., and V. G. Medvedev (2001) *Building and Using Dynamic Interest Rate Models*, John Wiley & Sons, Ltd., Chichester.
- Lando D. (1998) Oncox processes and credit risky securities, *Review of Derivatives Research*, **2**(2, 3), 99–120.
- Lando, D. (2000) Some elements of rating-based credit risk modeling, in: N. Jegadeesh and B. Tuckman (Eds), *Advanced Fixed-Income Valuation Tools*, John Wiley & Sons, Inc., New York.
- Leland H. E. (1994) Corporate debt value, bond covenants and optimal capital structure, *Journal of Finance*, **49**(4), 1213–1252.
- Longstaff, F., and E. Schwartz (1995) A simple approach to valuing risky fixed and floating rate debt, *Journal of Finance*, **50**, 789–819.
- Merton, R. (1974) On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance*, **29**, 449–470.
- Moody's Special Report (1992) Corporate bond defaults and default rates, *Moody's Investors Services*, New York.
- Nelson, C.R., and A. F. Siegel (1987) Parsimonious modeling of yield curves, *Journal of Business*, **60**, 473–489.
- Nickell, P., W. Perraudin and S. Varotto (2000) Stability of rating transitions, *Journal of Banking and Finance*, **24** (1–2), 203–227.
- Rebonato, R. (1998) *Interest-rate option models*, 2nd edn, John Wiley & Sons, Ltd, Chichester.
- Sandmann, K., and D. Sondermann (1993) A term structure model and the pricing of interest rates derivatives, *The Review of Future Markets*, **12**(2), 391–423.
- Sandmann, K., and D. Sondermann (1993) In the stability of lognormal interest rate models, SFB 303, Universität Bonn, Working Paper B-263.
- Standard & Poors Special Report (1998) Corporate defaults rise sharply in 1998, Standard & Poors, New York.
- Standard & Poors (1999) Rating performance 1998, stability and transition, Standard & Poors, New York.

Tapiero, C.S. (2003) Selecting the optimal yield curve: An optimal control approach, Working Paper, ESSEC, France.  
 Vasicek, O. (1977) An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, 177–188.  
 Walras, L. (1874) *Elements d'Economie Politique Pure*, Corbaz, Lausanne. (English translation, *Elements of Pure Economy*, Irwin, Homewood, IL, 1954.  
 Wets, R.J.B., S.W. Bianchi and L. Yang (2002) Serious zero curves. Technical report, EpiSolutions, Inc, El Cerrito, California,  
 Wilmott, P., (2000), Paul Wilmott on Quantitative Finance, John Wiley & Sons Ltd., Chichester.

## MATHEMATICAL APPENDIX

### A.1: Term structure and interest rates

Let the interest-rate process:

$$dr(t, T) = \mu(r, T) dt + \sigma(r, T) dw$$

The price of a zero-coupon bond is a function of such interest rates which we can assume to be:

$$\frac{dB(t, T)}{B(t, T)} = \alpha(r, t, T) dt + \beta(r, t, T) dw$$

with  $(\alpha(r, t, T), \beta(r, t, T))$  the parameters denoting the drift and diffusion of the bond's return. To determine these parameters, we apply Ito's Lemma's to:

$$B(t, T) = \exp[-r(t, T)(T - t)]$$

leading to:

$$dB(t, T) = \left[ \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} \mu(r, T) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) \right] dt + \frac{\partial B}{\partial r} \sigma(r, T) dw$$

and therefore, by equating the two equations, the bond and the term structure interest-rate model, we have:

$$\alpha(r, t, T)B = \left[ \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} \mu(r, T) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) \right]; \beta(r, t, T)B = \frac{\partial B}{\partial r} \sigma(r, T)$$

Now assume that the risk premium is proportional to the returns standard deviation. Assuming that the price of risk is a known function of  $r$  and time  $t$ , we have thus:

$$\alpha(r, t, T) = r + \lambda(r, t) \frac{1}{B} \frac{\partial B}{\partial r}$$

which we insert in the Bond equation derived above leading to:

$$rB + \lambda(r, t) \frac{\partial B}{\partial r} = \left[ \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} \mu(r, T) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) \right]$$

and finally we obtain a partial differential equation whose solution provides the price of a zero-coupon bond with maturity  $T$ ,

$$0 = \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} (\mu(r, T) - \lambda(r, t)) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) - rB$$

$$B(r, T, T) = 1$$

## A.2: Options on bonds

Let for example, the interest process:

$$dr = \mu(r, t) dt + \sigma(r, t) dw$$

A synthetic portfolio ( $n_S, n_T$ ) of two  $S$  and  $T$  bonds has a value and a rate of return given by:

$$V = n_S B(t, S) + n_T B(t, T) \quad \text{and} \quad \frac{dV}{V} = n_S \frac{dB(t, S)}{B(t, S)} + n_T \frac{dB(t, T)}{B(t, T)}$$

The rates of return on the  $T$  and  $S$  bonds are (as shown previously):

$$\frac{dB(t, T)}{B(t, T)} = \alpha_T(r, t) dt + \beta_T(r, t) dw$$

$$\alpha_T(r, t) = \frac{1}{B(t, T)} \left[ \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} \mu(r, T) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, T) \right];$$

$$\beta_T(r, t) = \frac{1}{B(t, T)} \frac{\partial B}{\partial r} \sigma(r, T)$$

$$\frac{dB(t, S)}{B(t, S)} = \alpha_S(r, t) dt + \beta_S(r, t) dw$$

$$\alpha_S(r, t) = \frac{1}{B(t, S)} \left[ \frac{\partial B}{\partial t} + \frac{\partial B}{\partial r} \mu(r, S) + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \sigma^2(r, S) \right];$$

$$\beta_S(r, t) = \frac{1}{B(t, S)} \frac{\partial B}{\partial r} \sigma(r, S)$$

We replace these terms in the synthetic bond portfolio leading to:

$$\frac{dV}{V} = (n_S \alpha_S + n_T \alpha_T) dt + (n_S \beta_S + n_T \beta_T) dw$$

A risk-free portfolio has no volatility, however. If the portfolio initial value is one dollar ( $V = 1$ ), we can then specify two equations in the two unknown portfolio parameters ( $n_S, n_T$ ) which we can solve simply. Explicitly, these equations are:

$$\begin{cases} n_S \beta_S + n_T \beta_T = 0 \\ n_S + n_T = 1 \end{cases} \quad \text{and} \quad \begin{cases} n_S = \frac{\beta_T}{\beta_T - \beta_S} \\ n_T = -\frac{\beta_S}{\beta_T - \beta_S} \end{cases}$$

The risk-free (synthetic) portfolio has thus a rate of growth, called the synthetic rate  $k(t)$ , explicitly given by:

$$\frac{dV}{V} = \left( \frac{\beta_T \alpha_S - \beta_S \alpha_T}{\beta_T - \beta_S} \right) dt = k(t) dt$$

This rate is equated to the spot rate  $r(t)$ , providing thereby the following equality:

$$r(t) = \frac{\beta_T \alpha_S - \beta_S \alpha_T}{\beta_T - \beta_S} \quad \text{or} \quad \lambda(t) = \frac{r(t) - \alpha_S}{\beta_S} = \frac{r(t) - \alpha_T}{\beta_T}$$

where  $\lambda(t)$  denotes the price of risk per unit volatility. Each bond with maturity  $T$  and  $S$  has at its exercise time a \$1 denomination, thus the value of each of these ( $S$  and  $T$ ) bonds is:

$$0 = \frac{\partial B_T}{\partial t} + \frac{\partial B_T}{\partial r} [\mu(r, T) - \lambda \beta_T] + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \beta_T^2 - r B_T, \quad B(r, T) = 1$$

$$0 = \frac{\partial B_S}{\partial t} + \frac{\partial B_S}{\partial r} [\mu(r, S) - \lambda \beta_S] + \frac{1}{2} \frac{\partial^2 B}{\partial r^2} \beta_S^2 - r B_S, \quad B(r, S) = 1$$

Given a solution to these two equations, we define the option value of a call on a  $T$  bond with  $S < T$  and strike price  $K$ , to be:  $X = \text{Max} [B(S, T) - K, 0]$  with  $B(S, T)$  the price of the  $T$  bond at time  $S$ . The bond value  $B(S, T)$  is of course found by solving for the term structure equation and equating  $B(r, S, T) = B(S, T)$ . To simplify matters, say that the solution (value at time  $t$ ) for the  $T$  bond is given by  $F(t, r, T)$ , then at time  $S$ , this value is:  $F(S, r, T)$  to which we equate  $B(S, T)$ . In other words,

$$X = \text{Max} [F(S, r, T) - K, 0]$$

Now, if the option price is  $B(\cdot)$ , then as we have seen in the plain vanilla model in the previous chapter, the value of the bond is found by solving for  $B(\cdot)$  in the following partial differential equation:

$$0 = \frac{\partial B}{\partial t} + \mu(r, t) \frac{\partial B}{\partial r} + \sigma^2 \frac{1}{2} \frac{\partial^2 B}{\partial r^2} - r B, \quad B(S, r) = \text{Max} [F(S, r, T) - K, 0]$$

as indicated in the text.

# Incomplete Markets and Stochastic Volatility

## 9.1 VOLATILITY DEFINED

Volatility pricing, estimation and analysis are topics of considerable interest in finance. The value of an option, for example, depends on the volatility, which cannot be observed directly but must be estimated or guessed – the larger the volatility the larger the value of an option. Thus, trading in options requires that volatility be predicted and positions taken to profit from forthcoming high volatility and vice versa from forthcoming low volatility. In many instances, attempts are also made to manage volatility, either by using derivative-based strategies or by some other creative means, such as ‘certification’. In a past issue of *The Economist* (18 August 2001, p. 56), an article on ‘Fishy Math’ pointed out that salmon certification may stabilize prices and thereby profit Alaska’s fishermen. To do so, options were used by the MSC (the Marine Stewardship Council, a not-for-profit agency that campaigns for sustaining fishing), to value the certification of Alaska salmon, claimed to ensure a certain standard of fishery and environmental management which customers are said to value. For fishermen, a long-term benefit would be to reduce the volatility of salmon prices and thereby increase the value of their catch. The valuation of such profits was found by the MSC using Black–Scholes options. That is to say, the options prices implied by those two levels of volatility – what a reasonable person would expect to pay to hedge the price risk before and after certification – were calculated and compared, indicating a profit for fishermen, a profit sufficient to cover the cost of certification. Choosing a model of volatility is critical in the valuation of derivatives, however. In a stable economic environment it makes sense to use plain vanilla models. However, there is ample historical evidence that this may not be the case and therefore volatility, and in particular stochastic volatility, can be the cause of market incompleteness and create appreciable difficulties in pricing assets and their derivatives. The study of volatility is thus important, for both these and many other reasons. For example, the validation of fundamental financial theory presumes both the ‘predictability’ of future prices and interest rates, as well as other relevant

time series. Financial markets and processes where the underlying uncertainty is modelled by 'random walks' are such an instance, since they can provide future predictions, albeit characterized by a known probability distribution. The random walk hypothesis further implies, as we saw earlier; independent increments, independently and identically distributed Gaussian random variables with mean zero and a linear growth of variance. Statistically independent increments imply in fact, 'a linear growth of uncertainty'. Technically, this is shown by noting that the functional relationship, implying independence,  $f(t + s) = f(t) + f(s)$  implies a linear growth since it is uniquely given by the linear time function,  $f(t + s) = (t + s)f(0)$ .

This facet of 'linear growth of uncertainty' has been severely criticized as too simplistic, ignoring the long-term dependence of financial time series. Further, empirical evidence has shown that financial series are not always 'well-behaved' and thus, cannot be always predicted. For this reason, extensive research has been initiated seeking to explain, for example, the leptokurtic character of rates of returns distributions, the 'chaotic behaviour' of time series, underscoring the 'unpredictability of future asset prices'. These approaches characterize 'nonlinear science' approaches to finance. Practically, 'bursts' of activity, 'feedback volatility' and broadly varying behaviours by stock market agents, 'memory' etc., are contributing to processes which do not exhibit predictable price processes and therefore violate the presumptions of fundamental finance. The study of these series has motivated a number of approaches falling under a number of themes spanning: fat tails (or Pareto–Levy stable) distribution analysis characterized by infinite variance; long-term memory and dependence characterized by explosive growth of volatility; chaotic analysis; Lyapunov stability analysis; complexity analysis; fractional Brownian motion; multifractal time series analysis; R/S (range to scale) analysis etc. Extensive study has been devoted to these methods (see, for example, the review papers of Mandelbrot (1997a) and Lo (1997)).

Volatility modelling and estimation is often specialized to the second moment evolution of a price process, but it is much more. Generally, we say that a random variable, say the returns  $x$  is more volatile than a random variable  $y$  if for all  $a > 0$ , the cumulative density functions of the returns distributions  $F_X(\cdot)$ ,  $F_Y(\cdot)$  satisfies,  $F_X(a) > F_Y(a)$ . The mathematics of 'stochastic ordering' consisting in comparing and ordering distributions, as above, has focused financial managers' attention on such measurements using terms such as 'stochastic dominance' (or first, second and third degree), 'hazard rate dominance', convex dominance, etc. These techniques have the advantage of being utility-free, but they are not easy to apply, nor is it always possible to do so. A practical measurement of volatility is thus problematic. When the underlying distribution of a process is Normal, consisting of two parameters, the mean and the variance, it makes sense to accept the standard deviation as a measure of volatility. However, when the underlying distribution is not Normal (as with leptokurtic distributions, expressing asymmetry in the distribution), the definition of what constitutes volatility has to be dealt with carefully. An appropriate measure of volatility is thus far from

being unique, albeit a process standard deviation is often used and will be used in this chapter. There are other indicators of volatility, such as the range  $R$ , the semi-variance, R/S statistics (see the last section in this chapter for a development and explanations of such statistics) etc. providing thereby more than one approach and more than one statistical measurement to express the volatility of a series.

Given the importance of volatility, a broad number of approaches and techniques have been applied to measure and model it. The simplest case is, of course, the constant (variance) volatility model implied in random walk models. When the variance changes over time (whether it is stochastic or not), models of volatility are needed that are both economically acceptable and statistically measurable. We shall provide a brief overview of these techniques in this chapter since they are currently a 'workhorse' of financial statistics.

## 9.2 MEMORY AND VOLATILITY

'Memory' represents quantitatively the effects of past states on the current one and how we use it to construct forecasts of future states. A temporal 'independence' is equivalent to a 'timeless' situation in which the events reached at one point in time are independent of past and future states. In this circumstance, there is no 'memory' and volatility tends to be smaller. A temporal dependence induces time correlations, however, and thereby a process variance (volatility) growth. Time and memory, in both psychological and quantitative senses, can also form the basis for distinguishing among past, present and future. Objectively, the present is now; subjectively, however, the present also consists of past and future. This idea has been stated clearly by St Augustine (*Confessions*, Book XI, xx):

yet perchance it might be properly said, 'there be three times; a present of things past, a present of things present, and a present of things future.' For these three do exist in some sort, in the soul, but otherwise I do not see them; present of things past, memory; present of things present, sight; present of things future, expectation.

We are thus always in the present. But the present has three dimensions:

- (1) The present of the past.
- (2) The present of the present.
- (3) The present of the future.

Technically, we construct the past with experiences and empirical observations of the (price) process as it unfolds over time; our construction of the future (prices), on the other hand, must be in terms of indeterminate and uncertain events which are our best assessment of the future (price) at a given (filtered) present time. To a large extent, 'technical analysis' in finance uses such an approach. We

have different mechanisms for establishing things past and establishing things future. Our ability to relate the past and the future to each other – i.e. to make sense of temporal change – by means of a temporal ‘sequentiality’ is the prime reason for studying memory processes. For example, ‘remembering that stock markets behave cyclically’ might induce a cyclical behaviour of prices (which need not, of course, be the case). ‘Remembering’, i.e. recording the claims history of an insured over the last years, may be used to determine a premium payments schedule. The ‘health’ history of a patient might provide important clues to determining the probabilities of his survival over time as he approaches ages where a population has a tendency to be depleted. In finance, these issues are particularly relevant. Rational expectations and its risk-neutral pricing framework squarely states that ‘there is no memory of the past’ since all current price values are ‘an estimate of future prices’. In this sense, in a rational expectations framework, the ‘present is the anticipation of the future at the known risk-free rate’. By the same token, the SDF (stochastic discount factor) claims essentially the same but without specifying a deterministic kernel for discounting future states. By contrast, charting approaches in finance state that there is a memory of the past which is used through modelling based on past data to determine current prices. The financial dilemma regarding rational expectations and charting is thus reduced to a memory issue and how it affects the process of price formation. Potential approaches can be summarized by:

- (1) No memory in which case the past and the future have no effect on current prices.
- (2) Anticipative (rational expectations or SDF) memory in which current prices are defined in terms of a predictable ‘expectation’ of future prices.
- (3) Long-run memory, expressing the inter-relationship of past events and current prices and therefore the omnipresent effects of the past in any present.

For example, if speculative prices exhibit dependency, then the existence of such dependency would be inconsistent with rational expectations and would thus make a strong case for technical forecasting on stock prices (contrary to the conventional assumption that prices fluctuate randomly and are thus unpredictable). In addition, the notion of market efficiency is dependent on ‘market memory’. Fama (1970) defines explicitly an efficient market as one in which information is instantly reflected in the market price. This means that, provided all the past information  $\mathbf{F}(t)$  at time  $t$  is used, a market is efficient if its expected price conditioned by this information equals the current price. Thus for a given time  $t + T$  and price  $p(t)$ , we have, as seen previously:  $p(t) = E[p(t + T) | \mathbf{F}(t)]$ . As time goes by, additional information is obtained and  $\mathbf{F}(t)$  grows to include more information (a new filtration)  $\mathbf{F}(t + 1)$  and thus  $p(t + 1) = E[p(t + 1 + T) | \mathbf{F}(t + 1)]$ . This property of markets efficiency (assuming that it exists) underlies the martingale approach to finance, as we saw earlier. Without it, markets have a measure of ‘predictability’ and can thus lead to some investor making arbitrage profits.

### 9.3 VOLATILITY, EQUILIBRIUM AND INCOMPLETE MARKETS

Volatility, and in particular stochastic volatility, is an increasingly important issue dealt with by financial managers. The *Financial Times* for example, reported in 1997 (although it could be any year):

The New York Stock Exchange (NYSE) has been swinging far more wildly than in previous years. The events of 1997 that struck stock markets throughout Asia and subsequently in Europe and in the US are additional proof that volatility (stochastic) is becoming a determinant factor of stock values. This has an important effect on investments and investing behaviour. Some investors, for example, are 'tiptoeing' away from the stock market and sitting on cash rather than stocks. Others are lulled by the swings in stock values and as a result are becoming less sensitive to these variations (which may be a costly strategy to follow if the stock market were to decline significantly). This year for example, there were daily price drops of more than 3 %, a phenomenon which in years past would have attracted a great deal of attention and warnings. Past experience has also indicated that when the volatility increases, it may signal a downturn on the stock market (although, it has also signalled upturns on the stock market – but less often). In any case, a growth of volatility makes investors rethink their strategy and thereby, to change their portfolio holdings.

Stochastic volatility is often used as a proof that markets are incomplete (since the former implies the latter). In other words, it implies an underlying departure from conventional approaches to economics and finance that invalidates risk-neutral pricing. Incompleteness, thus, reflects our inability to explain uniquely prices' formation. Ever since the Second World War change has been plentiful, providing an opportunity to explain why volatility may have grown or changed. Some factors contributing to an appreciable change in economics and finance theories that seek to explain the behaviour of financial markets include among others:

- The demise of Bretton Woods.
- The liberalization of the financial sector worldwide.
- Globalization through the growth of multinational firms, cross-boundary capital flows etc.
- The growth of derivatives and related products that have enriched financial theories and financial markets but at the same time have allowed the use of financial products on an unprecedented scale.

Explicitly, economic theory has changed! Classical equilibrium precepts, coined by the Arrow–Debreu–Mackenzie studies have diverted attention to disequilibrium theories, information asymmetries, organization and the effects of contracts on economic behaviour. Economic and finance theories have recognized these changes that led to new approaches – both theoretical and practical and underlie to a large extent fundamental finance. The assumption of the rational expectations hypothesis that markets clear (i.e. decision and expectations are compatible both in current and derivatives markets in the present and the future), the assumption that decision makers are homogeneous, self-interested, rational

and informed, with common knowledge of the market statistics came in some cases to be doubtful. As a result, the study of incompleteness and situations involving bounded rationality, information asymmetry, utility maximizing decision makers etc. have also become important elements to reckon with in devising a mechanism for the valuation and pricing of assets.

Although financial economics has greatly contributed to finance practice, both through its approach to valuation by risk-neutral pricing and in a better understanding of financial market mechanisms, there are some problems to be reckoned with. First, financial theory is based on assumptions that are not always right. In this case, we ought to develop other theories to compensate for theoretical imperfections. For these reasons, making sure that financial theory assumptions are validated is essential for making money using ‘complete market models’.

### 9.3.1 Incomplete markets

Markets are incomplete when any random cash flow cannot be generated by some portfolio strategy. The market is then deemed ‘not rich enough’. Technically, this means that the number of assets that make up a portfolio is smaller than the number of market risk sources plus one, or:

$$\text{number of assets} \leq \text{Number of risk sources} + 1$$

When this is not the case, we cannot replicate, for example, an option’s implied cash flow and thus, are unable to value the option uniquely. For this, as well as other reasons, incompleteness, implying non-uniqueness in pricing, is particularly important. Non-uniqueness can arise for many reasons, however, including for example issues:

- Due to pricing, rationality and psychology.
- Due to information asymmetries and networking.
- Due to transaction costs.
- Due to stochastic volatility.

If markets are not complete or close to it, financial markets have problems to value assets and investments. Some cases are well studied, however (transaction costs for certain types of assets, some problems associated with stochastic volatility), where one uses additional sources of information to replicate a derived financial asset. Financial markets may be perceived as too risky, perhaps ‘chaotic’, and therefore profits may be too volatile; the risk premium would then be too high and investment horizons smaller, thereby reducing investments. Finally, contingent claims may have an infinite number of prices (or equivalently an infinite number of martingale measures). As a result, valuation becomes forcibly utility-based, which is ‘subjective’ rather than based on the market mechanism. In these circumstances, the SDF (stochastic discount factor) framework presented in Chapter 3 is

particularly useful, providing an empirical approach to pricing (risk-discounting) financial assets.

**Example: Sources of incompleteness**

- (1) Incompleteness can arise in many circumstances. Below, a few are summarized briefly:
  - Because of lack of liquidity (leading to market-makers' and bid/ask spreads – for which trading micro-models are constructed).
  - Because of excessive friction defined in terms of: taxes; indivisibility of assets; varying rates for lending and borrowing, such as no short sales and various portfolio constraints.
  - Because of transaction costs leading to 'friction' in market transactions.
  - Because of insiders trading introducing a risk originating in information asymmetries and leading thereby to assets mis-pricing.
- (2) *Arbitrage*: The existence of arbitrage opportunities implies nonviable markets rendering the unique determination of contingent claim prices impossible. If there is arbitrage, there will be trade only out of equilibrium and thus the fundamental theory of finance will not be again applicable and risk-neutral pricing cannot be applied.
- (3) *Network and information asymmetries*: Networks of hedge funds, communicating with each other and often coordinated explicitly and implicitly and herding into speculative activities can lead to market inefficiencies, thus contradicting a basic hypothesis in finance assuming that agents are price takers. In networks, the information exchange provides a potential for information asymmetries or at least delays in information. In this sense, the existence of networks in their broadest and weakest form may also cause market incompleteness.
- (4) *Pricing and classical contract theory*: Transaction costs, informational asymmetries in the Arrow–Debreu paradigm, lead to significant amendments of classical analysis. For example, analysis of competition in the presence of moral hazard and adverse selection lead to stressing substantial differences between trade on 'contracts' and trade on contingent commodities. The profit associated with the sale of one unit of a (contingent) good depends then only on its price. Further, the profitability of the sale of one contract may also depend on the identity of the buyer. Identity matters, either because the buyer has bought other contracts (the exclusivity problem) or because profitability of the sales depends on the buyer's characteristics, also known as the screening problem. These issues relate to financial intermediation too, where special attention must be given to the effects of informational asymmetries to better understand prices and how they differ from the 'social values of commodities'.
- (5) *Psychology and rationality*: The *Financial Times* has pointed out that some investment funds seek to capitalize on human frailties to make money. For example: Are financial managers human? Are they always rational, mimicking *Star Trek's* Mr Spock? Are they devoid of emotions and irrationality?

Psychological decision-making processes integrated in economic rationales have raised serious concerns regarding the rationality axioms of DM processes, as was discussed in Chapters 2 and 3. There are, of course, many challenges to reckon with in understanding human behaviour. Some of these include:

- Thought processes based on decision-making approaches focusing on the big picture versus compartmentalization.
- The effects of under- and over-confidence on decision making.
- The application of heuristics of various sorts applied to trading.

These psychological aspects underpin an important trend in finance called ‘Behavioral Finance’ and at the same they provide and presume important sources of incompleteness, stimulating research to bridge observed and normative economic behaviour.

#### 9.4 PROCESS VARIANCE AND VOLATILITY

Say that a stock price has a time-variant mean and standard deviation given by  $(\mu_t, \sigma_t)$ . In other words, if we let  $z_t$  be a standard random variable then the record of the series can be written as follows:  $x_t = \mu_t + \sigma_t z_t$ . When the standard deviation is known, the time series can be used to estimate the mean parameter (even if it is time-variant). When the variance is not known, it is necessary to estimate it as well. Such estimation is usually difficult and requires that specific models describing the evolution of the variance be constructed. For example, if we standardize the time series, we obtain a standard normal probability random variable for the error as seen below,

$$z_t = \frac{x_t - \mu_t}{\sigma_t} \sim N(0, 1)$$

We can rewrite this model by setting  $\varepsilon_t = \sigma_t z_t$  where the error has a zero mean (usually obtained by de-trending the time series). If the standard deviation is not known, then of course the error is no longer normal and therefore there are statistical problems associated with its estimation. Models of the type ARCH and GARCH seek to estimate this variance by using the residual squared deviations. There are many ways to proceed, however, from both a modelling and a statistical point of view, rendering volatility modelling a challenging task. Empirical finance research has sought to explain volatility in terms of the randomness of incoming information and trading processes. In the first instance, volatility is explained by the effects of external events which were not accounted for initially, while in the latter instance it is based on the behaviour of traders, buyers and sellers that induce greater volatility (such as herd or other systematic and unsystematic behaviours). The number of approaches and statistical techniques one may use for estimating volatility vary as well. For this reason, we shall consider some simple cases, although numerous studies, both methodological and empirical, abound. Many references related to these topics are included as well in the ‘References and additional reading at the end of the chapter.

**Example**

Let  $R_{t+1}$ , the returns of a firm at time  $t + 1$ , be unknown at time  $t$  and assume that mean returns forecast at time  $t$  are given by the next period expectation  $\mu_t = E_t(R_{t+1})$ . This means that the conditional expectation of the one-period returns ‘forecast’ can be calculated. Such a model assumes rational expectations since current returns are strictly an expectation of future ones. At present, hypothesize a model for the error, given by  $\varepsilon_t$  – also called the innovation. Thus, a one-period ahead return can be written by:

$$R_{t+1} = \mu_t + \varepsilon_{t+1}$$

The volatility (or the return variance) is by definition:

$$\sigma_t^2 = E_t(R_{t+1}^2) - \mu_t^2 = E_t(\varepsilon_{t+1}^2)$$

which is presumed either known or unknown, in which case it is a stochastic volatility model. A simple variance estimate can be based on statistical historical averages. That is to say, using closing daily financial prices  $P_t$  (spot on stocks for example) and in particular using daily proportional price change:  $R_t = \ln P_t - \ln P_{t-1}$ , we obtain (historical) estimates for the mean and the variance:

$$\mu = \frac{1}{T} \sum_{t=1}^T R_t; \quad \sigma^2 = \frac{1}{T-1} \sum_{t=1}^T (R_t - \mu)^2$$

By the same token we can use the daily range (or a Hi, Lo statistic) for volatility estimation. This is justified by the fact that for identically and independently distributed (iid) large sample statistics, the range and the variance have, approximately, equivalent distributions. Then,

$$\sigma = \frac{0.627}{T-1} \sum_{t=1}^T \ln(H_t/L_t)$$

where  $(H_t, L_t)$  are the high and low prices of the trading day respectively.

Historical estimation can be developed further by building weighted estimation schemes, giving greater prominence to recent data compared to past data. In other words, say that a volatility estimate is given by a weighted sum of squares of past returns:

$$\hat{\sigma}_t^2 = E_t(R_{t+1}^2) = w_0 + \sum_{i=1}^k w_i(t) R_{t+i-1}^2$$

where  $w_i(t)$  denotes the weight at time  $t$  associated to past returns. Variance models may be differentiated then by the weighting schemes we use. For the naïve historical model, we have:

$$\hat{\sigma}_t^2 = E_t(R_{t+1}^2) = \frac{1}{T} \sum_{i=1}^T R_{t+i-1}^2; \quad w_i(t) = \frac{1}{T}$$

For an exponential smoothing of volatility forecasts, as done by Riskmetrics, we

have:

$$\hat{\sigma}_t^2 = E_t(R_{t+1}^2) = \sum_{i=1}^{\infty} \theta^{i-1} R_{t+i-1}^2; w_i(t) = \theta^{i-1}, 0 \leq \theta^{i-1} \leq 1$$

since:

$$\hat{\sigma}_t^2 = \sum_{i=1}^{\infty} \theta^{i-1} R_{t+i-1}^2 = R_{t-1}^2 + \theta \sum_{i=1}^{\infty} \theta^{i-1} R_{t-1+i-1}^2$$

and we obtain the recursive scheme:

$$\hat{\sigma}_t^2 = R_{t-1}^2 + \theta \hat{\sigma}_{t-1}^2$$

Extensions were suggested by Engle (1987, 1995) (ARCH models) and Bollerslev (GARCH models). There are other estimation techniques such as nonparametric models that are harder to specify. In these cases, the weighting function  $w(x_{t-i})$  expresses a memory based on a number of state variables. Such approaches are in general difficult to estimate. The importance of ARCH and GARCH modelling in financial statistics cannot be overestimated, however. Econometric software makes it possible to perform such statistical analyses with great ease, using general models of the variance. For further study we refer to Bollerslev (1986), Nelson and Foster (1994), Taylor (1986), and Engle and Bollerslev (1986).

### Example: Stochastic volatility and process discretization

A stochastic volatility model can be obtained by discretization of a plain vanilla continuous-time model. This demonstrates that in handling theoretical models for practical ends and discretizing the model we may also introduce problems associated with stochastic volatility. Say that an asset price is given by the often-used lognormal model:

$$\frac{dS}{S} = \mu dt + \sigma dW$$

where  $\mu$  is asset rate of return and  $\sigma$  is its volatility. An application of Ito's differential rule to  $Y = \ln S$ , yields:

$$dY = \left( \mu - \frac{\sigma^2}{2} \right) dt + \sigma dW$$

A simple discretization with a time interval  $\Delta k$ , needed for estimation purposes, yields:

$$Y_k - Y_{k-1} = \left( \mu - \frac{\sigma^2}{2} \right) \Delta k + \left( \sigma \sqrt{\Delta k} \right) Z_k; Z_k \sim N(0, 1); k = 1, 2, \dots$$

A linear regression provides an estimate of  $(\mu - \sigma^2/2)$ , requiring that the volatility be presumed known and constant for the estimate to be meaningful. If the volatility is not known but it is also estimated by the data at hand, then another regression is needed, supplied potentially by the ARCH-GARCH apparatus and providing a simultaneous estimation of the model's parameters. Such estimation

subsumes, however, a stochastic volatility (since the volatility is error-prone and estimated using historical values). As a result, discretization, even when it is properly done, can lead to estimation problems that imply stochastic volatility.

### 9.5 IMPLICIT VOLATILITY AND THE VOLATILITY SMILE

It is possible to estimate volatility for traded stocks, exchange rates and other financial instruments using the Black–Scholes (BS) equation. Note that the BS equation is given as a function of volatility and a number of other variables which are recorded easily or market-specified. As a result, we can use recorded option prices to calculate, other things being equal, the corresponding volatility. This is also called the implicit volatility. When there is no arbitrage (and the BS equation provides the option price), the implicit volatility corresponds to the actual volatility. Otherwise, there may be some opportunity for arbitrage profit. Ever since the stock market crash of 1987, it has been noted that options’ implicit volatility with the same maturity are a function of the strike. This is known as the volatility smile, shown graphically in Figure 9.1. It is believed that this effect is due to some extent to agents’ willingness to pay to hedge their position in case of sudden and unpredictable market reversal. Of course, such a ‘smile’ has a direct effect on return distributions which may no longer be normal but rather be defined by a skewed distribution.

Explicitly, say that we use the BS formula, specified in Chapter 6:

$$W = F(p, t; T, K, R_f, \sigma) = p(t)\Phi(d_1) - K e^{-R_f(T-t)}\Phi(d_2)$$

where  $\sigma$  is the volatility,  $T$  is the exercise (maturity) date,  $K$  is the exercise price and  $R_f$  is the risk-free interest rate and, of course,  $W$ , is the option price of the underlying asset whose price a time  $t$  is equal to  $p(t)$  with:

$$\Phi(y) = (2\pi)^{-1/2} \int_{-\infty}^y e^{-u^2/2} du,$$

$$d_1 = \left[ \frac{\log(p(t)/K) + (T - t)(R_f + \sigma^2/2)}{\beta\sqrt{T - t}} \right], \quad d_2 = d_1 - \sigma\sqrt{T - t},$$

A solution for  $\sigma$ , leads by implicit numerical techniques to a function which is



Figure 9.1

given by:

$$\sigma = \Psi(p, t; T, K, R_f, W)$$

In this manner, and using data for the option price, the volatility can be calculated. This analysis presumes, of course, that the BS option is the proper function for valuing an option on the stock exchange.

## 9.6 STOCHASTIC VOLATILITY MODELS

Stochastic volatility models presume that a process's volatility (variance) varies over time following some stochastic process, usually well specified. As a result, it is presumed that volatility growth increases market unpredictability, thereby rendering the application of the rational expectations hypothesis, at best, a tenuous one. Modelling volatility models might require then a broad number of approaches not falling under the 'random walk hypothesis'. Techniques such as ARCH and GARCH, we referred to, might be used to estimate empirically the volatility in such cases. Below, we consider a number of problems and issues associated with stochastic volatility in the valuation of financial assets.

Stochastic volatility introduces another 'source of risk', a volatility risk, when we model an asset's price (or returns). This leads to incompleteness and thus to non-unique asset prices. Risk-neutral pricing is no longer applicable since the probabilities calculated by the application of rational expectations (i.e. hedging to eliminate all sources of risk and using the risk-free rate as a mechanism to replicate assets) do not lead to risk-neutral valuation. For this reason, unless some other asset can be used to 'enrich' a hedging portfolio (for the volatility risk as well), we are limited to using approximations based on an economic rationale or on some other principles so that our process can be constructed (and on the basis of which risk-neutral pricing can be applied). A number of approaches can be applied including:

- time contraction,
- approximate replication,
- approximate risk-neutral pricing valuation,
- bounding.

These approaches and related ones are the subject of much ongoing research. Again, we shall consider some simple cases and, in some cases, define only a quantitative framework of the problem at hand.

### 9.6.1 Stochastic volatility binomial models\*

Stochastic volatility has an important effect on the process underlying uncertainty, altering the basic assumption of 'normal' or 'binomial' driving disturbances. To see these effects we consider the simple binomial model we have used repeatedly

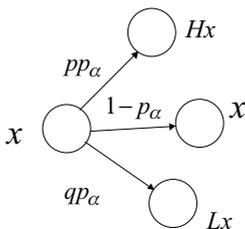


Figure 9.2

and given below:

$$\frac{x_{t+1} - x_t}{x_t} = \alpha \varepsilon_t; \varepsilon_t = \begin{cases} H & p \\ L & q \end{cases};$$

Here  $\alpha$  denotes the process constant volatility. Now assume a mild stochastic volatility. Namely, we let the volatility  $\alpha$  assume a value of 1 and zero with probabilities  $(p_\alpha, q_\alpha)$  leading to:

$$\frac{x_{t+1} - x_t}{x_t} = \tilde{\alpha}_t \varepsilon_t; \varepsilon_t = \begin{cases} H & p \\ L & q \end{cases}; \tilde{\alpha}_t = \begin{cases} 1 & p_\alpha \\ 0 & q_\alpha \end{cases}$$

In this case, the random (binomial) volatility is reduced to a trinomial model (see Figure 9.2) where  $p$  is the probability of the constant volatility model and  $p_\alpha$  is the probability that volatility equals one and  $q_\alpha$  is the probability that there is no volatility. For this simple case, already it is not possible to construct a perfect hedge for, say, an option as we have done earlier. This is because there are two sources of risk – one associated with the price and the other with volatility. Assuming one asset only, the number of risk sources is larger than the number of assets and therefore we have an incomplete market situation where prices need not be unique.

When volatility is constant, note that  $\alpha_t \varepsilon_t$  is a random walk, but when volatility is a random variable, the process  $\tilde{\alpha}_t \varepsilon_t$  is no longer a random walk. Let  $z_t = \tilde{\alpha}_t \varepsilon_t$  have a density function  $F_{z_t}(\cdot)$  and assume that the random walk and the volatility are statistically independent (which is a strong assumption). Using elementary probability calculations, we have:

$$F_{z_t}(z) = \int_0^\infty F_\alpha(\alpha) F_\varepsilon\left(\frac{z}{\alpha}\right) \frac{d\alpha}{\alpha} - \int_{-\infty}^0 F_\alpha(\alpha) F_\varepsilon\left(\frac{z}{\alpha}\right) \frac{d\alpha}{\alpha}$$

For example, if the random walk  $\varepsilon_t = (+1, -1)$  is biased and with probabilities  $(p, \bar{p})$ , while volatility assumes two values  $\tilde{\alpha} = (a, b)$  with probabilities  $(q, \bar{q})$ , the following quadrinomial process results:

$$z_t = \begin{cases} +a & w.p. & pq \\ +b & w.p. & p\bar{q} \\ -b & w.p. & \bar{p}q \\ -a & w.p. & \bar{p}\bar{q} \end{cases}$$

In other words, stochastic volatility has generated incompleteness in the form of

a quadrinomial process. By enriching the potential states volatility may assume we are augmenting the ‘volatility stochasticity’. Note, that it is not the size of volatility that induces incompleteness but its uncertainty. We shall see below, using a simple example, that the option of a ‘mild volatility’ process is larger than a larger (constant) volatility process – hence, emphasizing the effects of incompleteness (stochastic volatility) on option prices which in some cases can be more important to greater (but constant) volatility.

**Time contraction**

The underlying rationale of ‘time contraction’ is a reverse discretization. In other words, *assuming* that at the continuous-time limit, the underlying price process can be represented by a stochastic differential equation of the Ito type, it is then reasonable to assume that there is some binomial process that approximates the underlying process. Of course, there may be more than one way to do so (thus leading, potentially, to multiple prices) and therefore, this approach has to be applied carefully to secure that the limit makes economic sense as well. In this approach, a multinomial process is replaced by a binomial tree, consisting of as many stages (discretized time) as are needed to replicate the underlying model. For example, the trinomial process considered earlier can be reduced to a two-stage tree as shown in Figure 9.3, where  $(p_1, p_2, p_3)$  are assumed to be risk-neutral probabilities, appropriately selected by replication. Note that we have necessarily:

$$\begin{aligned}
 p_1 p_2 &= p p_\alpha \\
 p_1 \bar{p}_2 + \bar{p}_1 p_3 &= 1 - p_\alpha \\
 \bar{p}_1 \bar{p}_3 &= q \bar{p}_\alpha
 \end{aligned}$$

Since there are only two independent equations, we have in fact a system of three variables in two equations that can be solved in a large number of ways (for example as a function of  $\bar{p}_3$ ). This means also that there is no unique price. In this case,

$$\begin{aligned}
 p_2 &= [p p_\alpha / (p \bar{p}_\alpha / \bar{p}_3)] \\
 p_1 &= 1 - (p \bar{p}_\alpha / \bar{p}_3)
 \end{aligned}$$

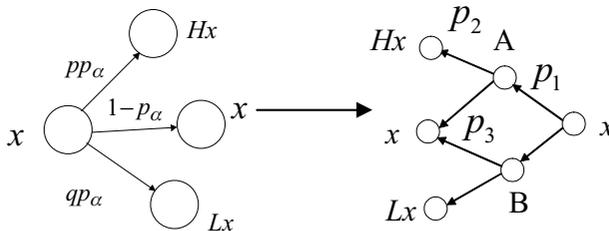


Figure 9.3

However, if we assume rational expectations, then:

$$A = \frac{1}{1 + R_f} [p_2(1 + H)x + \bar{p}_2x] \text{ and } B = \frac{1}{1 + R_f} [p_3x + \bar{p}_3(1 + L)x]$$

and

$$1 = \left( \frac{1}{1 + R_f} \right)^2 [p_1 p_2 (1 + H) + p_1 \bar{p}_2 + p_3 \bar{p}_1 + \bar{p}_1 \bar{p}_3 (1 + L)]$$

which provides a third equation in  $(p_1, p_2, p_3)$ . Of course, for  $p_\alpha = 1$  we have nonstochastic volatility and, therefore, we can calculate the approximate risk-neutral probabilities as a function of  $0 < p_\alpha < 1$ . For simplicity, set  $p_1 = p_2 = p_3$ , then we have a quadratic equation:

$$0 = \left[ p^2 - 2 \frac{L}{(H + L)} p - \frac{(1 + R_f)^2 - 1 - L}{(H + L)} \right]$$

whose solution is given by:

$$p = \frac{L}{(H + L)} \pm \sqrt{\left( \frac{L}{(H + L)} \right)^2 + \frac{(1 + R_f)^2 - 1 - L}{(H + L)}}$$

If we use the following parameters as an example,  $1 + H = 1.4$ ,  $1 + L = 0.8$ ,  $R_f = 0.04$ , the only feasible solution is:

$$p = \sqrt{1 + \frac{(1 + 0.04)^2 - 0.8}{(0.2)}} - 1 \text{ or } p = 0.55177$$

Inserting in our equations:

$$\begin{aligned} A/x &= 0.9615 [(0.55177)(1.4) + 0.44823] = 1.1737 \\ B/x &= 0.9615 [(0.55177) + 0.44823(0.8)] = 0.8753 \end{aligned}$$

In this particular case, the option price is given by:

$$C = \frac{1}{(1 + r)^2} [p^{*2}(H - K)x] = 0.9245[0.3(0.55177)^2x] = 0.084439x$$

We consider next the problem with mild volatility (see Figure 9.4) and set:  $p' = p^{*2}/p_\alpha = 0.3044/p_\alpha$ . If  $p_\alpha = 0.6$ ,  $p' = 0.50733$ . If we assume no stochastic volatility but  $\alpha = 1$  with probability 1 (that is a process more volatile than the previous one), then the value of an option is calculated by  $p' = p^{*2}/p_\alpha = 0.3044/p_\alpha$ . In addition, since  $p_\alpha = 1$  we have  $p' = 0.3044$ . As a result, the option price is:

$$C = \frac{1}{1 + R} [p'(H - K)x] = 0.9245 [0.3044(1.4 - 1.1)x] = 0.08442x$$

compared with an option price with mild volatility given by  $C = 0.084439x$ .

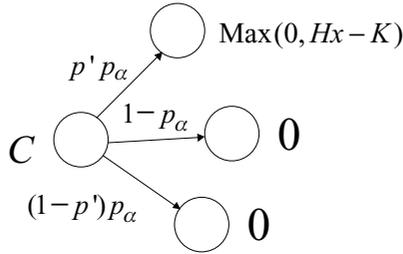


Figure 9.4

Thus, the difference due to the stochastic volatility growth is equal to  $0.08442x - 0.084439x = 0.00003x$ . In other words the value of an option increases both with volatility and with stochastic volatility.

We can generalize this approach further. For example, if the volatility can assume a number of potential values, say,  $\tilde{\alpha} = (0, 1, 2, 3, 4, 5)$ , then it is possible to reduce the ten-nomial process to a ten-stage binomial process as shown below. Mathematically, this is given by:

$$x_{t+1} = x_t + \tilde{\alpha}\varepsilon_t \quad \text{and} \quad \varepsilon_t = \begin{cases} +1 & 1/2 \\ -1 & 1/2 \end{cases}; \quad \tilde{\alpha} = \begin{cases} 0 & w.p. p_0 \\ \dots & \dots \dots \\ 5 & w.p. p_5 \end{cases}$$

An analysis similar to the previous one provides the mechanism to calculate the approximate binomial risk-neutral probability. Of course, in this model, the single stage ten-nomial process is transformed into a nine-stage binomial process (see Figure 9.5). The number of ways to do so might be very large, however. Additional information and assumptions might then be needed to reduce the number of possibilities and thereby constrain the set of prices the financial asset can assume.

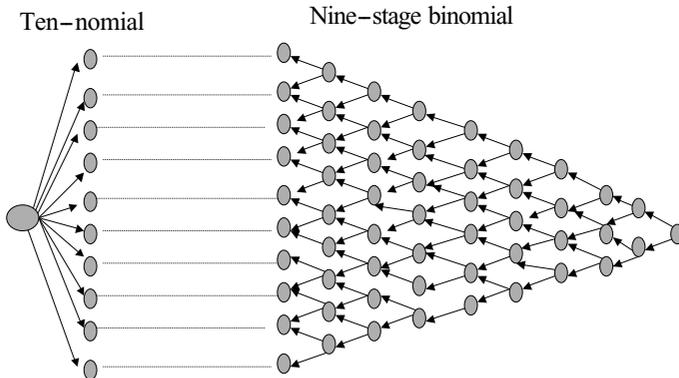


Figure 9.5

**Problem**

Consider the following stochastic volatility process:

$$x_{t+1} = x_t + \tilde{\alpha}\varepsilon_t; \quad \varepsilon_t = \begin{cases} +1 & 1/2 \\ -1 & 1/2 \end{cases}; \quad \tilde{\alpha} = \begin{cases} 2 & q \\ 3 & 1-q \end{cases}$$

Construct the equivalent binomial process and find the risk-neutral probabilities coherent with such a process.

$$x_{t+1}^\alpha = x_t^\alpha + z_t; \quad z_t = \begin{cases} +3 & w.p. (0.5)(1-q) \\ +2 & w.p. (0.5)q \\ -2 & w.p. (0.5)q \\ -3 & w.p. (0.5)(1-q) \end{cases}$$

This approach can be extended to continuous-time models. For example, assume a mean reversion interest-rate model:

$$dS = \mu(\alpha - S)dt + V dW \quad \text{or} \quad S(t) = \alpha + [S(0) - \alpha \pm V\varepsilon(t)] e^{-\mu t}$$

where  $\varepsilon(t)$  is a standard random walk and  $V$  is a stochastic volatility given by:

$$dV = \theta(\beta - V)dt + \kappa dW \quad \text{or} \quad V(t) = \beta + [V(0) - \beta \pm \kappa\eta(t)] e^{-\theta t}$$

We assume that  $V \geq 0$  for simplicity. Note that the interest-rate process combines two sources of risk given by  $(\eta(t), \varepsilon(t))$  and therefore:

$$S(t) = S(0)e^{-\mu t} + \alpha(1 - e^{-\mu t}) \pm [V(0)e^{-(\theta+\mu)t} + \beta(e^{-\mu t} - e^{-(\theta+\mu)t})]\varepsilon(t) \pm \kappa e^{-(\theta+\mu)t}\eta(t)\varepsilon(t)$$

The mean rate is therefore:

$$\hat{S}(t) = S(0)e^{-\mu t} + \alpha(1 - e^{-\mu t}) + \kappa e^{-(\theta+\mu)t} E[\eta(t)\varepsilon(t)]$$

Since  $[\eta(t), \varepsilon(t)]$  are standard random walks their covariation  $E[\eta(t)\varepsilon(t)]$  is equal to  $1/4$ , thus:

$$\hat{S}(t) = \alpha + e^{-\mu t}[S(0) - \alpha + \kappa e^{-\theta t}/4]$$

Since the resulting process is given explicitly by a quadrinomial process:

$$\begin{aligned} S(t) = S(0)e^{-\mu t} + \alpha(1 - e^{-\mu t}) \pm & \\ + e^{-(\mu+\theta)t}[V(0) + \beta(e^{\theta t} - 1) + \kappa] w.p. & 1/4 \\ + e^{-(\mu+\theta)t}[V(0) + \beta(e^{\theta t} - 1) - \kappa] w.p. & 1/4 \\ - e^{-(\mu+\theta)t}[V(0) + \beta(e^{\theta t} - 1) - \kappa] w.p. & 1/4 \\ - e^{-(\mu+\theta)t}[V(0) + \beta(e^{\theta t} - 1) + \kappa] w.p. & 1/4 \end{aligned}$$

The interest-rate variance can be calculated easily since:

$$\text{Var}(S(t)) = e^{-2(\theta+\mu)t}[V(0) + \beta(e^{\theta t} - 1) - (\kappa)/4]^2 + \kappa^2 e^{-2(\mu+\theta)t}$$

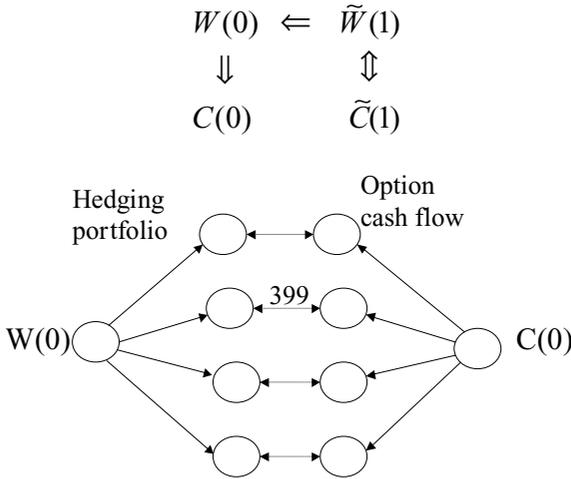


Figure 9.6

This process can be replaced by an approximate risk-neutral pricing process by time contraction, by mean–variance hedging or by applying the principle of least divergence as we shall see below.

**(2) Mean variance replication hedging**

This approach consists in the construction of a hedging portfolio in an incomplete (stochastic volatility) market by equating ‘as much as possible’ the cash flows resulting from the hedging portfolio and the option’s contract. We seek to do so, while respecting the basic rules of rational expectations and risk-neutral pricing. In particular, say that we consider a two-stage model (Figure 9.6; see also Chapter 6). Thus, assuming that a portfolio and an option have the same cash flow, (i.e.  $\tilde{W}(1) \equiv \tilde{C}(1)$ ), their current price are necessarily equal, implying that:  $W(0) = C(0)$ . Since, by risk-neutral pricing,

$$C(0) = \frac{1}{1 + R_f} E\tilde{C}(1); \quad W(0) = \frac{1}{1 + R_f} E\tilde{W}(1) \text{ and } C(0) = W(0)$$

or equivalently:

$$E\tilde{W}(1) = \tilde{E}C(1) \quad \text{and further,} \quad E\tilde{W}^2(1) = \tilde{E}C^2(1)$$

These equations provide only four equations while the number of parameters might be large. However, since a hedging portfolio can involve a far greater number of parameters, it might be necessary to select an objective to minimize. A number of possibilities are available.

A simple quadratic optimization problem consisting in the minimization of the squared difference of probabilities associated to the binomial tree might be used. Alternatively, the minimization of a hedging portfolio and the option ex-post values of some option contract leads to the following:

$$\text{Min}_{p_1, \dots, p_n} \Phi = E(\tilde{W}(1) - \tilde{C}(1))^2$$

subject to:  $W(0) = C(0)$  or

$$\frac{1}{1 + R_f} [E\tilde{W}(1)] = \frac{1}{1 + R_f} [E\tilde{C}(1)] \text{ and } E\tilde{W}^2(1) = E\tilde{C}^2(1)$$

Of course the minimization objective can be simplified further to:

$$\text{Min}_{p_1, \dots, p_n} \Phi = E\tilde{C}^2(1) - E\tilde{W}(1)\tilde{C}(1) \text{ or } \text{Min}_{p_1, \dots, p_n} \Phi = \sum_{i=1}^n p_i(C_{1i}^2 - W_{1i}C_{1i})$$

where  $W_{1i}$ ,  $C_{1i}$  are the hedging portfolio and option outcomes associated with each of the events  $i$  that occurs with probability  $p_i$ ,  $i = 1, 2, \dots, n$ . For example, consider the four-states model given above and assume a portfolio consisting of stocks and bonds  $aS + B$ . Let  $K$  be the strike price with  $S_i - K > 0$ ,  $i = 1, 2$  and  $S_i - K < 0$ ,  $i = 3, 4$  then the cash flows at time '1' for the portfolio and the option are respectively  $aS_i + (1 + R_f)B$ ,  $i = 1, 2, 3, 4$  and  $(S_1 - K, S_2 - K, 0, 0)$  as shown in Figure 9.7. The following and explicit nonlinear optimization problem results:

$$\text{Min}_{p_1, \dots, p_4} \Phi = \sum_{i=1}^2 p_i(S_i - K)[(1 - a)S_i - K - (1 + R_f)B]$$

Subject to:

$$a \sum_{i=1}^4 p_i(S_i - S) + R_f B = \sum_{i=1}^2 p_i(S_i - K) \text{ and}$$

$$S = \frac{1}{1 + R_f} \sum_{i=1}^4 p_i S_i, \sum_{i=1}^4 p_i = 1; p_i \geq 0$$

Explicitly, say that we have the following parameters:

$$S_1 = 110, S_2 = 100, S_3 = 90, S_4 = 80, S = 90, K = 95, R_f = 0.12$$

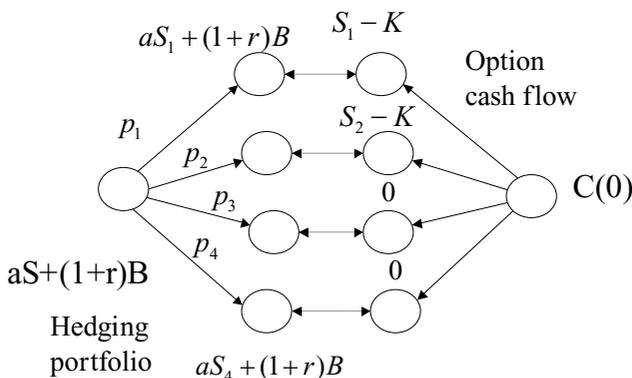


Figure 9.7

Then, our problem is reduced to a nonlinear optimization problem:

$$\text{Min}_{p_1, \dots, p_4} \Phi = 3p_1(110a - 15 - 1.12B) + p_2(100a - 5 - 1.12B)$$

Subject to:

$$4ap_1 + 2ap_2 - 2ap_4 + 0.024B = 3p_1 + p_2$$

$$1.09126p_1 + 0.99206p_2 + 0.8928p_3 + 0.79365p_4 = 1$$

$$p_1 + p_2 + p_3 + p_4 = 1; p_i \geq 0$$

The problem can of course be solved easily by standard nonlinear optimization software.

### Example

Let  $S_j$ ,  $j = 1, 2, \dots, n$  be the  $n$  states a stock can assume at the time an option can be exercised. We set,  $S_0 < S_1 < S_2 < \dots < S_n$  and define the buying and selling prices of the stock by:  $S^a$ ,  $S^b$  respectively. By the same token, we define the corresponding observed call option prices  $C^a$ ,  $C^b$ . Let  $p$  be the probability of a price increase. Of course, if the ex-post price is,  $S_n$ , this will correspond to the stock increasing each time period with probability given by:  $P_n = \binom{n}{n} p^n (1-p)^{n-n} = p^n$ . By the same token, the probability of the stock having a price  $S_j$  corresponding to the stock increasing  $j$  times and decreasing  $n-j$  times is given by the binomial probability:

$$P_j = \binom{n}{j} p^j (1-p)^{n-j}$$

As a result, we have, under risk-neutral pricing:

$$S = \frac{1}{1+R_f} \sum_{j=0}^n P_j S_j = \frac{1}{1+R_f} \sum_{j=0}^n S_j \binom{n}{j} p^j (1-p)^{n-j}; S^a \leq S \leq S^b$$

and the call option price is:

$$C = \sum_{j=0}^n C_j = \frac{1}{(1+R_f)^n} \sum_{j=0}^n \binom{n}{j} p^j (1-p)^{n-j} \text{Max}(S_j - K, 0)$$

subject to an appropriate constraint on the call options values,  $C^a \leq C \leq C^b$ . Note that  $S$  and  $C$ , as well as  $p$ , are the only unknown values so far. While the buy and sell values for stock and option, the strike time  $n$  and its price  $K$ , as well as the risk-free discount rate and future prices  $S_j$ , are given. Our problem at present is to select an objective to optimize and calculate the risk-neutral probabilities. We can do so by minimizing the quadratic distance between a portfolio of a unit of stock and a bond  $B$ . At time  $n$ , the portfolio is equal  $aS_j + (1+r)^n B$  if the

price is,  $S_j$ . Of course, initially, the portfolio equals:

$$S = \frac{1}{(1 + R_f)^n} \left( a \sum_{j=0}^n P_j S_j + (1 + R_f)^n B \right)$$

As a result, the least squares replicating portfolio is given by:

$$\Phi = \sum_{j=1}^n P_j (a S_j + (1 + R_f)^n B - \text{Max}(S_j - K, 0))^2$$

leading to the following optimization problem:

$$\text{Min}_{1 \geq p \geq 0, C, S} \Phi = \sum_{j=1}^n \binom{n}{j} p^j (1 - p)^{n-j} [a S_j + (1 + R_f)^n B - \text{Max}(S_j - K, 0)]^2$$

Subject to:

$$S = \frac{1}{(1 + R_f)^n} \sum_{j=0}^n \binom{n}{j} p^j (1 - p)^{n-j} S_j; S^a \leq S \leq S^b$$

$$C = \frac{1}{(1 + r)^n} \sum_{j=0}^n \binom{n}{j} p^j (1 - p)^{n-j} \text{Max}(S_j - K, 0); C^a \leq C \leq C^b$$

$$a S + B = C$$

This is again a tractable nonlinear optimization problem.

**(3) Divergence and entropy**

Divergence is defined in terms of directional discrimination information (Kullback, 1959) which can be measured in discrete states by two probability distributions say,  $(p_i, q_i), i = 1, 2, \dots, m$ , as follows.

$$I(p, q) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i}; I(q, p) = \sum_{i=1}^m q_i \log \frac{q_i}{p_i}$$

while the divergence is:

$$J(p, q) = I(p, q) + I(q, p) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i} + \sum_{i=1}^m q_i \log \frac{q_i}{p_i}$$

and finally,

$$J(p, q) = \sum_{i=1}^m (p_i - q_i) \log \frac{p_i}{q_i}$$

This expression measures the ‘distance’ between these two probability distributions. When they are the same their value is null and therefore, given a distribution ‘p’, a distribution ‘q’ can be selected by the minimization of the divergence, subject to a number of conditions imposed on both distributions ‘p’ and ‘q’ (such as expectations, second moments, risk-neutral pricing etc.). For example, one distribution

may be an empirical distribution while the other may be theoretical, providing the parameters needed for the application of approximate risk-neutral pricing.

This approach can be generalized further to a bivariate state, involving time as well as states. In this case, we have for each time period:

$$I(p, q) = \sum_{t=1}^T \sum_{i=1}^n p_{it} \log \left( \frac{p_{it}}{q_{it}} \right); \quad J(p, q) = \sum_{t=1}^T \sum_{i=1}^n (p_{it} - q_{it}) \log \left( \frac{p_{it}}{q_{it}} \right)$$

where the following constraints must be satisfied:

$$\sum_{i=1}^n p_{it} = 1; \quad \sum_{i=1}^n q_{it} = 1; \quad p_{it} \geq 0, \quad q_{it} \geq 0$$

Moments conditions as well as other constraints may be imposed as well, providing a 'least divergence' risk-neutral pricing approximation to the empirical (incomplete) distribution.

### Example

Consider the following random volatility process

$$x_{t+1}^\alpha = x_t^\alpha + z_t; \quad z_t = \begin{cases} +3 & \text{w.p. } (0.5)(1-q) \\ +1 & \text{w.p. } (0.5)q \\ -1 & \text{w.p. } (0.5)q \\ -3 & \text{w.p. } (0.5)(1-q) \end{cases}$$

A three-stage standard binomial process leads to:

$$x_{t+1}^\alpha = x_t^\alpha + z_t; \quad z_t = \begin{cases} +3 & \text{w.p. } (0.5)(1-q) \leftrightarrow \pi^3 \\ +1 & \text{w.p. } (0.5)q \leftrightarrow 3\pi^2(1-\pi) \\ -1 & \text{w.p. } (0.5)q \leftrightarrow 3\pi(1-\pi)^2 \\ -3 & \text{w.p. } (0.5)(1-q) \leftrightarrow (1-\pi)^3 \end{cases}$$

As a result, we can calculate the probability  $\pi$  by minimizing the divergence  $J$  which is given by an appropriate choice of  $\pi$ :

$$\begin{aligned} J &= [\pi^3 - (0.5)(1-q)] \log \left( \frac{2\pi^3}{1-q} \right) + [3\pi^2(1-\pi) - (0.5)q] \\ &\quad \times \log \left( \frac{6\pi^2(1-\pi)}{q} \right) + [3\pi(1-\pi)^2 - (0.5)q] \log \left( \frac{6\pi(1-\pi)^2}{q} \right) \\ &\quad + [(1-\pi)^3 - (0.5)(1-q)] \log \left( \frac{2(1-\pi)^3}{1-q} \right) \end{aligned}$$

Of course other constraints may be imposed as well. Namely, if the current price is \$1, we have by risk-neutral pricing the constraint:

$$1 = \left( \frac{1}{1+R_f} \right)^2 [3\pi^3 + \pi^2(1-\pi) - \pi(1-\pi)^2 - 3(1-\pi)^3]; \quad 0 \leq \pi \leq 1$$

As a result, the least divergence parameter  $\pi$  is found by solving the optimization problem above.

## 9.7 EQUILIBRIUM, SDF AND THE EULER EQUATIONS\*

We have seen, throughout Chapters 5, 6 and 7, the importance of the rational expectations hypothesis as a concept of equilibrium for determining asset prices. In Chapter 3, we have also used the maximization of the expected utility of consumption to determine a rationality leading to a pricing mechanism we have called the SDF (stochastic discount factor). In other words, while in rational expectations we have an asset price determined by:

$$\text{Current price} = \frac{1}{1 + R_f} E^* \{ \text{Future Prices} \}$$

where  $E^*$  denotes expectation with respect to a 'subjective' probability (in J. Muth, 1961 words) which we called the risk-neutral probability and  $R_f$  is the risk-free rate. In the SDF framework, we had:

$$\text{Current price} = E \{ \text{Kernel}^* \text{Future Prices} \}$$

In this section, we extend the two-period framework used in Chapter 3 to multiple periods. To do so, we shall use Euler's equation, providing the condition for an equilibrium based on a rationality of expected utility of consumption. Let an investor maximizing the expected utility of consumption:

$$V_t = \text{Max} \sum_{j=0}^{T-1} \rho^j u(c_{t+j}) + \rho^T G(W_T)$$

where  $u(c_{t+j})$  is the utility of consumption at time  $t + j$ ,  $T$  is the final time and  $G(W_T)$  is the terminal wealth state at time  $T$ . At time  $t$ , the change rate in the wealth is:

$$W_t - W_{t-1} = \Delta W_t = q_t c_t - R_t \quad \text{and therefore } c_{t+j} = \frac{\Delta W_{t+j} + R_{t+j}}{q_{t+j}}$$

We insert this last expression in the utility to be maximized:

$$V_t = \text{Max} \sum_{j=0}^{T-1} \rho^j u \left( \frac{\Delta W_{t+j} + R_{t+j}}{q_{t+j}} \right) + \rho^T G(W_T)$$

Application of Euler's equation, a necessary condition for value maximization, yields:

$$\frac{\partial V_t}{\partial W_{t+j}} - \Delta \left( \frac{\partial V_t}{\partial \Delta W_{t+j}} \right) = 0$$

Since

$$\frac{\partial V_t}{\partial W_{t+j}} = 0, \Delta \left( \frac{\partial V_t}{\partial \Delta W_{t+j}} \right) = 0$$

and therefore we have the following ‘equilibrium’:

$$\frac{\partial V_t}{\partial \Delta W_{t+j}} = \frac{\rho^j}{q_{t+j}} \frac{\partial u(c_{t+j})}{\partial \Delta W_{t+j}} = \text{constant or } \frac{\partial u(c_{t+j-1})}{\partial \Delta W_{t+j-1}} = E \left\{ \rho \frac{q_{t+j-1}}{q_{t+j}} \frac{\partial u(c_{t+j})}{\partial \Delta W_{t+j}} \right\}$$

In other words, the marginal utility of wealth increments (savings) equals the discounted inflation-adjusted marginal utilities of consumption. If wealth is invested in a portfolio of assets such that:

$$\Delta W_t = (N_t - N_{t-1})p_t = p_t \Delta N_t \quad \text{and} \quad \frac{\partial V_t}{\partial \Delta W_{t+j}} = p_t \frac{\partial u(c_{t+j})}{\partial \Delta N_{t+j}}$$

and therefore,

$$\frac{\partial u(c_{t+j-1})}{\partial \Delta N_{t+j-1}} p_{t-1} = E \left\{ \rho \frac{q_{t+j-1}}{q_{t+j}} \frac{\partial u(c_{t+j})}{\partial \Delta N_{t+j}} p_t \right\}$$

and

$$p_{t-1} = E \left\{ \rho \frac{q_{t+j-1}}{q_{t+j}} \frac{u'(c_{t+j})}{u'(c_{t+j-1})} p_t \right\}; \quad u'(c_{t+j}) = \frac{\partial u(c_{t+j})}{\partial \Delta N_{t+j}}$$

since at time  $t-1$ , the future price at time  $t$  is random, we have:

$$p_{t-1} = E \{ M_t p_t \}; \quad M_t = \rho \frac{q_{t+j-1}}{q_{t+j}} \frac{u'(c_{t+j})}{u'(c_{t+j-1})}$$

where  $M_t$  is the kernel, or the stochastic discount factor, expressing the ‘consumption impatience’. This equation can also be written as follows:

$$1 + R_t = \frac{p_t}{p_{t-1}}; \quad 1 = E \left\{ M_t \frac{p_t}{p_{t-1}} \right\} \rightarrow 1 = E \{ M_t (1 + R_t) | \Phi_t \}$$

which is the standard form of the SDF equation.

### Example: The risk-free rate

If  $p_t$  is a bond worth \$1 at time  $t$ , then for a risk-free discount rate:

$$\frac{1}{1 + R_f} = E \{ M_t \} (1) \quad \text{and therefore } E \{ M_t \} = \frac{1}{1 + R_f}$$

This leads to:

$$\frac{M_t}{E(M_t)} = \rho(1 + R_f) \frac{q_{t+j-1}}{q_{t+j}} \frac{u'(c_{t+j})}{u'(c_{t+j-1})} \quad \text{and finally to } p_{t-1} = E_t^* \left( \frac{p_t}{1 + R_f} \right)$$

where  $E_t^*$  is a modified (subjective) probability distribution.

**Example: Risk premium and the CAPM beta**

For a particular risky asset, the CAPM provides a linear discount mechanism which is:

$$M_{t+1} = a_t + b_t R_{M,t+1}$$

In other words, for a given stock, whose rate of return is  $1 + R_{t+1} = p_{t+1}/p_t$ , we have:

$$1 = E \{M_{t+1}(1 + R_{t+1})\} \rightarrow E(1 + R_{t+1}) = \frac{1}{E(M_{t+1})} - \frac{\text{cov}(M_{t+1}, 1 + R_{t+1})}{E(M_{t+1})}$$

and therefore, upon introducing the linear SDF, we have:

$$E(1 + R_{t+1}) = \frac{1}{E(M_{t+1})} - \frac{\text{cov}(M_{t+1}, 1 + R_{t+1})}{E(M_{t+1})}$$

After we insert the linear model for the kernel we have:

$$\begin{aligned} E(1 + R_{t+1}) &= (1 + R_{f,t}) [1 - \text{cov}(M_{t+1}, 1 + R_{t+1})] \quad \text{and} \\ E(1 + R_{t+1}) &= (1 + R_{f,t+1}) [1 - \text{cov}(a + bR_{M,t+1}, 1 + R_{t+1})] \end{aligned}$$

which is reduced to:

$$E(R_{t+1} - R_{f,t+1}) = \frac{\text{cov}(R_{M,t+1} - R_{f,t+1}, R_{t+1} - R_{f,t+1})}{\text{var}(R_{M,t+1} - R_{f,t+1})} E_t(R_{M,t+1} - R_{f,t+1})$$

or to

$$E(R_{t+1} - R_{f,t+1}) = \beta E_t(R_{M,t+1} - R_{f,t+1})$$

However, the hypothesis that the kernel is linear may be limiting. Recent studies have suggested that we use a quadratic measurement of risk with a kernel given by:

$$M_{t+1} = a_t + b_t R_{M,t+1} + c_t R_{M,t+1}^2$$

In this case, the skewness of the distribution also enters into the determination of the value of the stock.

## 9.8 SELECTED TOPICS\*

When a process has more sources of risk than assets, we are, as stated earlier, in an incomplete market situation. In such cases it is possible to proceed in two ways. Either find additional assets to use (for example, another option with different maturity and strike price) or approximate the stochastic volatility process by another risk-reduction process. There are two problems we shall consider in detail, including (stochastic) jumps and stochastic volatility continuous type models.

Problems are of three types: first, how to construct a process describing reliably the evolution of the variance; second, what are the sources of uncertainty of volatility; and, third, how to represent the stochastic relationship between the

underlying process and its volatility. These equations are difficult to justify analytically and therefore we shall be satisfied with any model that practically can be used and can replicate historical statistical data. There are, however, a number of continuous-time models for stocks, returns, interest rates and other prices and their volatility that are often used. We shall consider such a model in the appendix to this chapter in detail to highlight some of the technical problems we must resolve in order to deal with such problems.

### 9.8.1 The Hull and White model and stochastic volatility

Hull and White (1987) have suggested a stochastic volatility model in which volatility is a geometric Brownian motion. This is written as follows:

$$dS/S = \alpha dt + \sqrt{V} dw, \quad S(0) = S_0; \quad dV/V = \mu dt + \xi dz, \quad V(0) = v_0; \\ E(dw dz) = \rho dt$$

where  $V$  is the volatility while  $dw$  and  $dz$  are two Wiener processes, with correlation  $\rho$ . A call option would in this case be a function of both  $S$  and  $V$ , or  $C(t, S, V)$ . Since there are two sources of risk, the hedging (replicating) portfolio must reflect this multiplicity of risks. Hull and White assume that the volatility risk is perfectly diversifiable, consequently the volatility risk premium is null. Using a Taylor series development of the option's price allows the calculation of the value of a call option as a function of small perturbations in volatility. The resulting solution turns out to be (see the Appendix for a mathematical development):

$$\frac{1}{dt} E \left( \frac{dC}{C} \right) = R_f + \left[ (\alpha - R_f) \frac{S}{C} \frac{\partial C}{\partial S} + \frac{\lambda_V}{C} \frac{\partial C}{\partial V} \right] \\ = \frac{1}{dt} E \left\{ \frac{\partial C}{C \partial t} dt + \frac{\partial C}{C \partial S} dS + \frac{\partial C}{C \partial V} dV + \frac{1}{2} \frac{\partial^2 C}{C \partial S^2} [dS]^2 + \frac{1}{2} \frac{\partial^2 C}{C \partial V^2} [dV]^2 \right. \\ \left. + \frac{\partial^2 C}{C \partial S \partial V} [dS dV] \right\}$$

After some additional manipulations, we obtain a partial differential equation in two variables:

$$-R_f - \frac{\partial C}{C \partial t} = \frac{\partial C}{C \partial S} R_f S + \frac{\partial C}{C \partial V} (\mu V - \lambda_V) + \frac{1}{2} \frac{\partial^2 C}{C \partial S^2} S^2 V \\ + \frac{1}{2} \frac{\partial^2 C}{C \partial V^2} \xi^2 V^2 + \frac{\partial^2 C}{C \partial S \partial V} \rho \xi S V^{3/2} C(S, V, T) = \text{Max}(S(T) - K, 0)$$

where  $K$  is the strike price,  $\lambda_V = (\mu - R_f)V\beta_V$  and while  $\beta_V$  is the beta of the volatility. The analytical treatment of such problems is clearly difficult. In 1976, Cox introduced a model represented generally by:

$$dS = \mu(S, t) dt + V(S, t) dW$$

Additionally a volatility state that  $V(S, t) = \sigma S^\delta$  with  $\delta$  a real number between 0 and 1. Application of Ito's Lemma, as seen in Chapter 4, leads to:

$$dV = \frac{\partial V}{\partial S} \left[ \mu(S, t) + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} V^2(S, t) \right] dt + V(S, t) dW$$

and in this special case, we have:

$$V(S) = \sigma S^\delta, \quad \frac{\partial V}{\partial S} = \delta \sigma S^{\delta-1}, \quad \frac{\partial^2 V}{\partial S^2} = \delta(\delta - 1) \sigma S^{\delta-2}$$

which we insert in the equation above to obtain a stochastic volatility model:

$$dV/V = \delta S^{-1} \left[ \mu(S, t) + \frac{1}{2} \delta(\delta - 1) S^{-2} V(S, t) \right] dt + dW$$

A broad number of other models can be constructed. In particular, for interest rate models we saw in Chapter 8, mean reversion models. For example, Ornstein–Uhlenbeck models of stochastic volatility are used with both additive and geometric models for the volatility equation. The additive model is given by:

$$\frac{dS}{S} = \mu dt + \sqrt{V} dW_t \quad \text{and} \quad dV = \alpha(\theta - V) dt + \xi dW'_t$$

while the geometric model is:

$$\frac{dS}{S} = \mu dt + \sqrt{V} dW_t \quad \text{and} \quad \frac{dV}{V} = \alpha(\theta - V) dt + \xi dW'_t$$

In both cases the process is mean-reverting where  $\theta$  corresponds to a volatility, a deviation from which induces a volatility movement. It can thus be interpreted as the long-run volatility.  $\alpha$  is the mean reversion driving force while  $\xi$  is the stochastic effect on volatility. The study of these models is in general difficult, however.

### 9.8.2 Options and jump processes (Merton, 1976)

We shall consider next another ‘incomplete’ model with two sources of risk where one of the sources is a jump. We treat this model in detail to highlight as well the treatment of models with jumps. Merton considered such a problem for the following price process:

$$\frac{dS}{S} = \alpha dt + \sigma dw + K dQ$$

where  $dQ$  is an adapted Poisson process with parameter  $q \Delta t$ . In other words,  $Q(t + \Delta t) - Q(t)$  has a Poisson distribution function with mean  $q \Delta t$  or for infinitesimal time intervals:

$$dQ = \begin{cases} 1 & \text{w.p. } q dt \\ 0 & \text{w.p. } (1 - q) dt \end{cases}$$

Let  $F = F(S, t)$  be the option price. When a jump occurs, the new option price is  $F[S(1 + K)]$ . As a result,

$$dF = [F(S(1 + K)) - F] dQ$$

When no jump occurs, we have a process evolving according to the diffusion process:

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial S} dS + \frac{1}{2} \frac{\partial^2 F}{\partial S^2} (dS)^2$$

Letting  $\tau = T - t$  be the remaining time to the exercise date, we have:

$$dF = \left[ -\frac{\partial F}{\partial \tau} + \alpha S \frac{\partial F}{\partial p} + \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 F}{\partial S^2} \right] dt + S\sigma \frac{\partial F}{\partial S} dw$$

Combining these two equations, we obtain:

$$dF = a dt + b dw + c dQ$$

$$a = \left[ -\frac{\partial F}{\partial \tau} + \alpha S \frac{\partial F}{\partial S} + \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 F}{\partial S^2} \right]; \quad b = S\sigma \frac{\partial F}{\partial S}; \quad c = F[S(1 + K)] - F$$

with

$$E(dF) = [a + qc] dt \quad \text{since} \quad E(dQ) = q dt$$

To eliminate the stochastic elements (and thereby the risks implied in the price process) in this equation, we construct a portfolio consisting of the option and a stock. To eliminate the ‘Wiener risk’, i.e. the effect of ‘dw’, we let the portfolio  $Z$  consist of a future contract whose price is  $S$  for which a proportion  $v$  of stock options is sold (which will be calculated such that this risk disappears). In this case, the value of the portfolio is:

$$dZ = S\alpha dt + S\sigma dw + SK dQ - [va dt + vb dw + vc dQ]$$

If we set  $v = S\sigma/b$  and insert in the equation above (as done by Black–Scholes), then we will eliminate the ‘Wiener risk’ since:

$$dZ = S(\alpha - \sigma a/b) dt + (S\sigma - vb) dw + S(K - \sigma c/b) dQ$$

or

$$dZ = S(\alpha - \sigma a/b) dt + S(K - \sigma c/b) dQ$$

In this case, if there is no jump, the evolution of the portfolio follows the differential equation:

$$dZ = S(\alpha - \sigma a/b) dt$$

However, if there is a jump, then the portfolio evolution is:

$$dZ = S(\alpha - \sigma a/b) dt + S(K - \sigma c/b) dQ$$

Since the jump probability equals,  $q dt$ , we obviously have:

$$\frac{E(dZ)}{dt} = S(\alpha - \sigma a/b) + Sq(K - \sigma c/b)$$

There remains a risk in the portfolio due to the jump. To eliminate it we can construct another portfolio using an option  $F'$  (with exercise price  $E'$ ) and a future contract such that the terms in  $dQ$  are eliminated as well. Then, constructing a combination of the first ( $Z$ ) portfolio and the second portfolio ( $Z'$ ), both sources of uncertainty will be eliminated. Applying an arbitrage argument (stating that there cannot be a return to a riskless portfolio which is greater than the riskless rate of return) we obtain the proper proportions of the riskless portfolio.

Alternatively, finance theory (and in particular, application of the CAPM (capital asset pricing model) state that any risky portfolio has a rate of return in a small time interval  $dt$  which is equal the riskless rate plus a premium for the risk assumed. Thus, using the CAPM we can write:

$$E \frac{dZ}{Z dt} = R_f + \lambda \frac{S(K - \sigma c/b)}{Z}$$

where  $\lambda$  is assumed to be a constant and expresses the 'market price' for the risk associated with a jump. This equation can be analysed further, leading to the following partial differential equation which remains to be solved (once the boundary conditions are specified):

$$-\frac{\partial F}{\partial \tau} + (\lambda - q) \left\{ SK \frac{\partial F}{\partial S} - (F[S(1 + K) - F]) \right\} + \frac{1}{2} \frac{\partial^2 F}{\partial S^2} S^2 \sigma^2 - R_f F = 0$$

with boundary condition:

$$F(T) = \text{Max}[0, S(T) - E]$$

Of course, for an American option, it is necessary to specify the right to exercise the option prior to its final exercise date, or

$$F(t) = \text{Max}[F^*(t), S(t) - E]$$

where  $F^*(t)$  is the value of the option which is not exercised at time  $t$  and given by the solution of the equation above. The solution of this equation is of course much more difficult than the Black–Scholes partial differential equation. Specific cases have been solved analytically, while numerical techniques can be applied to obtain numerical solutions.

## 9.9 THE RANGE PROCESS AND VOLATILITY

The range process of a time series is measured by the difference between the largest and the lowest values the time series assumes within a given time interval. It provides another indication for a process volatility with some noteworthy differences between the range and the process standard deviation (or variance). Explicitly, when a series becomes more volatile, the series standard deviation estimate

varies more slowly than that of the range. Thus, a growth surge in volatility might be detected more quickly using the range. By the same token, when the volatility declines, the range process will be stabilized. These properties have been used, for example, in the *R/S* (range to standard deviation statistic) applied in financial analysis to detect volatility shifts. Both the variance and the range processes are therefore two sources of information which are important. The Bloomberg, for example, provides such a statistic for financial time series, also named the Hurst exponent (Hurst, 1951) or the *R/S* index. This index is essentially a parameter that seeks to quantify the statistical bias arising from self-similarity power laws in time series. In other words, it expresses the degree of power nonlinearity in the variance growth of the series. It is defined through rescaling the range into a dimensionless factor.

Calculations for the range and the *R/S* statistic are made as follows. Samples are of fixed length  $N$  are constructed, and thus the sample range is given by:

$$R_{t,N} = \text{Max}\{y_{t,N}\} - \text{Min}\{y_{t,N}\}$$

while the sample standard deviation is calculated by:

$$S_{t,N} = \sqrt{\frac{\sum_{i=1}^N (y_{t,i} - \hat{y}_{t,N})^2}{N - 1}}$$

where  $\hat{y}_{t,N}$  is the sample average. A regression,  $(R/S) = (\text{Const} * N)^H$  provides an estimate of  $H$ , the Hurst exponent, or using a logarithmic transformation:

$$\ln\left(\frac{R_N}{S_N}\right) = a + bH; \quad b = \log(\alpha N);$$

With the notation:  $H$  = Hurst exponent,  $R$  = sample's range,  $S$  = sample's standard deviation and finally  $\alpha$  = a constant. For random (Normal) processes, the Hurst index turns out to equal 0.5. While for any values larger than 0.5 obtained in a regression, it may indicate 'long-term dependence'. Use of the Hurst index should be made carefully and critically, however. The origins of the Hurst exponent are due to Hurst who began working on the Nile River Dam project and studied the random behaviour of the dam and the influx of water from rainfall over the thousand years data have been recorded. The observation was made that if the series were random, the range would increase with the square root of time – A result confirmed by many time series as well as theoretically for normal processes. Hurst noted explicitly that most natural phenomena follow a biased random walk and thus characterized it by the parameter  $H$  expressing as well a series' dependence called by Mandelbrot the 'Joseph effect' (Joseph interpreted Pharaoh's dream as seven years of plenty followed by seven years of famine). Explicitly, a correlation  $C$  between disjoint increments of the series is given by  $C = 2^{2H-1} - 1$ . Thus, if  $H = 0.5$ , the disjoint intervals are uncorrelated. For  $H > 0.5$ , the series are correlated, exhibiting a memory effect as stated above (which tends to amplify patterns in time series). For  $H < 0.5$ , these are called

'anti-persistent' time series. Such analyses require large samples  $N$ , however, which might not be always available. For this reason, such analyses are used when series are long, such as sunspots, water levels of rivers, intra-day trading stock market ticker data etc.

An attempt to represent these series, expressing a persistent behaviour (or alternatively a nonlinear variance growth) was reached by Mandelbrot who introduced a fractional Brownian motion, denoted by  $B_H(t)$  (see also Greene and Fielitz (1977, 1980) for an application in finance). A particular relationship for fractional Brownian motion which is pointed out by Mandelbrot and Van Ness (1968) is based on the self-similarity of the power law for such processes which means that the increment for a time interval  $s$  are in distribution proportional to  $s^H$ , or:

$$B_H(t+s) - B_H(t) \xrightarrow[\text{i.d.}]{s^H} [B_H(t+1) - B_H(t)]$$

where i.d. means in distribution. Furthermore, the increments variance is:

$$E [B_H(t+s) - B_H(t)]^2 = s^{2H} E [B_H(t+1) - B_H(t)]^2$$

which means that the variance for any time interval  $s$  is equal to  $s^{2H}$  times the variance for the unit interval. Of course, it is now obvious that for  $H = 0.5$ , the variance is linear (as is the case for random walks and for Brownian motion) and it is nonlinear otherwise. In this sense, assuming a relationship between the Hurst exponent (which is also a power law for the series) and the notion of long-run dependence of series (modelled by fractional Brownian motion), an estimate of the one is indicative of the other. From the finance point of view, such observations are extremely important. First and foremost, long-run dependence violates the basic assumptions made regarding price processes that are valued under the assumption of complete markets. As such, they can be conceived as statistical tests for 'fundamental' assumptions regarding the underlying process. Second, the Hurst index can be used as a 'herd effect' index applied to stocks or other time series, meaning that series volatility that have a tendency to grow, will grow faster over time if the index is greater than 0.5 and vice versa if the index is smaller than 0.5. For these reasons, the  $R/S$  index has also been associated to 'chaos', revealing series that are increasingly unpredictable.

## REFERENCES AND ADDITIONAL READING

- Adelman, I. (1965) Long cycles – fact or artifact ? *American Economic Review*, **60**, 440–463.  
 Amin, K. (1993) Jump diffusion option valuation in discrete time, *Journal of Finance*, **48**, 1833–1863.  
 Amin, K.I., and V.K. Ng (1993) Option valuation with systematic stochastic volatility, *Journal of Finance*, **48**, 881–909.  
 Andersen, T.G. (1994) Stochastic autoregressive volatility: A framework for volatility modeling, *Mathematical Finance*, **4**, 75–102.  
 Anis, A.A., and E.H. Lloyd (1976) The expected values of the adjusted rescaled Hurst range of independent normal summands. *Biometrika*, **63**, 111–116.

- Baillie, R.T., and T. Bollerslev (1990) A multivariate generalized ARCH approach to modeling risk premia in forward foreign rate markets, *Journal of International Money and Finance*, **9**(7), 309–324.
- Ball, C., and W. Torous (1985) On jumps in common stock prices and their impact on call option prices, *Journal of Finance*, **40**, 155–173.
- Bera, A.K., and M.L. Higgins (1993) ARCH models: Properties, estimation and testing, *Journal of Economic Surveys*, **7**, 305–366.
- Beran, Jan (1994) *Statistics for Long-Memory Processes*, Chapman & Hall, London.
- Bhattacharya, R.N., V.K. Gupta and E. Waymire (1983) The Hurst effect under trends, *Journal of Applied Probability*, **20**, 649–662.
- Biagini, F., P. Guasoni and M. Pratelli (2000) Mean–variance hedging for stochastic volatility models, *Mathematical Finance*, **10**(2), 109–123.
- Blank, S.C. (1991) ‘Chaos’ in futures markets? A nonlinear dynamical analysis, *The Journal of Futures Markets*, **11**, 711–728.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–328.
- Bollerslev, T. (1990) Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH model, *The Review of Economics and Statistics*, **72**, 498–505.
- Bollerslev, T., and R.F. Engle (1993) Common persistence in conditional variances, *Econometrica*, **61**, 167–186.
- Bollerslev, T., R.Y. Chou and K. F. Kroner (1992) ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, **52**, 5–59.
- Bollerslev, T., R.F. Engle and D. B. Nelson (1994) ARCH models, in *Handbook of Econometrics*, Vol. 4, R. F. Engle and D. McFadden (Eds), North Holland, Amsterdam.
- Booth, G., F. Kaen and P. Koveos (1982) R/S analysis of foreign exchange rates under two international monetary regimes, *Journal of Monetary Economics*, **10**, 407–415.
- Breedon, Douglas T. (1979) An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics*, **7**(3), 265–296.
- Breedon, Douglas T., and Litzenberger, Robert H. (1978) Prices of state-contingent claims implicit in option prices, *Journal of Business*, **51**, 621–651.
- Brock, W.A., and P.J. de Lima (1996) Nonlinear time series, complexity theory and finance, in G. Maddala and C. Rao (Eds), *Handbook of Statistics*, Vol. 14, *Statistical Methods in Finance*, North Holland, Amsterdam.
- Brock, W.A., and M.J. P. Magill (1979) Dynamics under uncertainty, *Econometrica*, **47** 843–868.
- Brock, W.A., D.A. Hsieh and D. LeBaron (1991) *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, MIT Press, Cambridge, MA.
- Cao, M., and J. Wei (1999) Pricing weather derivatives: An equilibrium approach, Working Paper, Queens University, Ontario.
- Campbell, J., and J. Cochrane (1999) By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy*, **107**, 205–251.
- Cecchetti, S., P. Lam and N. Mark (1990) Mean reversion in equilibrium asset prices, *American Economic Review*, **80**, 398–418.
- Cecchetti, S., P. Lam and N. Mark (1990) Evaluating empirical tests of asset pricing models, *American Economic Review*, **80**(2), 48–51.
- Cheung, Y.W. (1993) Long memory in foreign exchange rates, *Journal of Business Economics and Statistics*, **11**, 93–101.
- Cotton, P., J.P. Fouque, G. Papanicolaou and K.R. Sircar (2000) *Stochastic Volatility Correction for Interest Rates Derivatives*, May, Stanford University.
- Cox, D.R. (1991) Long range dependence, nonlinearity and time irreversibility, *Journal of Time Series Analysis*, **12**(4), 329–335.

- Cvitanic, J., W. Schachermayer and H. Wang (2001) Utility maximization in incomplete markets with random endowment, *Finance Stochastics*, **5**(2), 259–272.
- Davis, M. H. A. (1998) Option pricing in incomplete markets, in *Mathematics of Derivatives Securities*, M.A.H. Dempster and S. R. Pliska (Eds), Cambridge University Press, Cambridge.
- Diebold, F., and G. Rudebusch (1989) Long memory and persistence in aggregate output, *Journal of Monetary Economics*, **24**, 189–209.
- Diebold, F., and G. Rudebusch (1991) On the power of the Dickey–Fuller test against fractional alternatives, *Economic Letters*, **35**, 155–160.
- El Karoui, N., and M. C. Quenez (1995) Dynamic programming and pricing contingent claims in incomplete markets, *SIAM Journal of Control and Optimization*, **33**, 29–66.
- El Karoui, N., C. Lepage, R. Myneni, N. Roseau and R. Viswanathan (1991) The valuation of hedging and contingent claims Markovian interest rates, Working Paper, Université de Paris 6, Jussieu, France.
- Engle, R.F. (1995) *ARCH Selected Reading*, Oxford University Press, Oxford.
- Engle, R.F., and T. Bollerslev (1986) Modeling the persistence of conditional variances, *Econometric Reviews*, **5**, 1–50.
- Feller, W. (1951) The asymptotic distribution of the range of sums of independent random variables, *Annals of Mathematical Statistics*, **22**, 427–432.
- Feller, W. (1957, 1966) *An Introduction to Probability Theory and its Applications*, Vols I and II, John Wiley & Sons, Inc., New York.
- Fouque, J. P., G. Papanicolaou and K. R. Sircar (2000) *Stochastic Volatility*, Cambridge University Press, Cambridge.
- Frank, M., and T. Stengos (1988) Chaotic dynamics in economic time series, *Journal of Economic Surveys*, **2**, 103–133.
- Fung, H.G. and W.C. Lo (1993) Memory in interest rate futures, *The Journal of Futures Markets*, **13**, 865–873.
- Fung, Hung-Gay, Wai-Chung Lo, John E. Peterson (1994) Examining the dependency in intraday stock index futures, *The Journal of Futures Markets*, **14**, 405–419.
- Geman, H. (Ed.) (1998) *Insurance and Weather Derivatives: From Exotic Options to Exotic Underlying*, Risk Books, London.
- Geske, R., and K. Shastri (1985) Valuation by approximation: A comparison of alternative option valuation techniques, *Journal of Financial and Quantitative Analysis*, **20**, 45–71.
- Ghysels, E., A.C. Harvey and E. Renault (1996) Stochastic volatility, in C. R. Rao and G. S. Maddala (Eds), *Statistical Methods in Finance*, North-Holland, Amsterdam.
- Gourieroux, C. (1997) *ARCH Models and Financial Applications*, Springer Verlag, New York.
- Granger, C.W., and T. Trasvirta (1993) *Modeling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Green, M.T., and B. Fielitz (1977) Long term dependence in common stock returns, *Journal of Financial Economics*, **4**, 339–349.
- Green, M.T., and B. Fielitz (1980) Long term dependence and least squares regression in investment analysis, *Management Science*, **26**(10), 1031–1038.
- Harvey, A.C., E. Ruiz and N. Shephard (1994) Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247–264.
- Helms, B., F. Kaen and R. Rosenman (1984) Memory in commodity futures contracts, *Journal of Futures Markets*, **4**, 559–567.
- Hobson, D., and L. Rogers (1998) Complete models with stochastic volatility, *Mathematical Finance*, **8**, 27–48.
- Hsieh, D.A. (1989) Testing for nonlinear dependence in daily foreign exchange rates, *Journal of Business*, **62**, 339–368.
- Hsieh, D.A. (1991) Chaos and nonlinear dynamics application to financial markets, *Journal of Finance*, **46**, 1839–77.

- Hull, J., and A. White (1987) The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **42**, 281–300.
- Hurst, H.E. (1951) Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, 770–808.
- Imhoff, J. P. (1985) On the range of Brownian motion and its inverse process, *Annals of Probability*, **13**(3), 1011–1017.
- Imhoff, J. P. (1992) A construction of the Brownian motion path from BES (3) pieces, *Stochastic Processes and Applications*, **43**, 345–353.
- Kullback, S. (1959) *Information Theory and Statistics*, Wiley, New York.
- LeBaron, B. (1994) Chaos and nonlinear forecastability in economics and finance, *Philosophical Transactions of the Royal Society of London, A*, **348**, 397–404.
- Liu, T., C.W.J. Granger and W. P. Heller (1992) Using the correlation exponent to decide whether an economic series is chaotic, *Journal of Applied Econometrics*, **7**, Supplement, S23–S39.
- Lo, Andrew W. (1992) Long term memory in stock market prices, *Econometrica*, **59**(5), 1279–1313.
- Lo, Andrew W. (1997) Fat tails, long memory and the stock market since 1960's, *Economic Notes*, **26**, 213–245.
- Mandelbrot, B. (1971) When can price be arbitrated efficiently? A limit to the the validity of the random walk and martingale models, *Review of Economics and Statistics*, **53**, 225–236.
- Mandelbrot, B. (1972) Statistical methodology for non-periodic cycles: From the covariance to R/S analysis, *Annals of Economic and Social Measurement*, **1**, 259–290.
- Mandelbrot, B.B. (1971) Analysis of long run dependence in economics: The R/S technique, *Econometrica*, **39**, 68–69.
- Mandelbrot, B. (1997a) Three fractal models in finance: Discontinuity, concentration, risk, *Economic Notes*, **26**, 171–212.
- Mandelbrot, B.B. (1997b) *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer Verlag, New York.
- Mandelbrot, B., and J. W. Van Ness (1968) Fractional Brownian motions, fractional noises and applications, *SIAM Review*, **10** 422–437.
- Mandelbrot, B., and M. Taqqu (1979) Robust R/S analysis of long run serial correlation, *Bulletin of the International Statistical Institute*, **48**, Book 2, 59–104.
- Merton, R. (1976) Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics*, **3**, 125–144.
- Naik, V., and M. Lee (1990) General equilibrium pricing of options on the market portfolio with discontinuous returns, *Review of Financial Studies*, **3**, 493–521.
- Nelson, Charles, and Charles Plosser (1982) Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics*, **10**, 139–162.
- Nelson, Daniel B., and D. P. Foster (1994) Asymptotic filtering theory for univariate ARCH model, *Econometrica*, **62**, 1–41.
- Otway, T.H. (1995) Records of the Florentine proveditori degli cambiatori: An example of an antipersistent time series in economics, *Chaos, Solitons and Fractals*, **5**, 103–107.
- Peter, Edgar E. (1995) *Chaos and Order in Capital Markets*, John Wiley & Sons, Inc., New York.
- Renault, E. (1996) Econometric models of option pricing errors, in D.M. Kreps and K.F. Wallis (Eds), *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge.
- Sandmann, K., and D. Sondermann (1993) A term structure model and the pricing of interest rates derivatives, *The Review of Futures Markets*, **12**(2), 391–423.
- Scheinkman, J.A. (1994) Nonlinear dynamics in economics and finance, *Philosophical Transactions of the Royal Society of London*, **346**, 235–250.

- Scheinkman, J.A., and B. LeBaron (1989) Nonlinear dynamics and stock returns, *Journal of Business*, **62**, 311–337.
- Schlogl, Erik, and D. Sommer (1994) On short rate processes and their implications for term structure movements, Discussion Paper B-293, University of Bonn, Department of Statistics, Bonn.
- Siebenaler, Yves (1997) Etude de l’amplitude pour certains processus de Markov, Thesis, Department of Mathematics, University of Nancy I, France (under supervision of Professor Pierre Vallois).
- Tapiero, C.S., and P. Vallois (1997) Range reliability in random walks, *Mathematical Methods of Operations Research*, **45**, 325–345.
- Taqqu, M. S. (1986) A bibliographical guide to self similar processes and long range dependence, in *Dependence in Probability and Statistics*, E. Eberlein and M.S. Taqqu (Eds), Birkhäuser, Boston, pp. 137–165.
- Vallois, P. (1995) On the range process of a Bernoulli random walk, in *Proceedings of the Sixth International Symposium on Applied Stochastic Models and Data Analysis*, Vol. II., J. Janssen and C.H. Skiadas (Eds), World Scientific, Singapore, pp. 1020–1031.
- Vallois, P. (1996) The range of a simple random walk on  $\mathbb{Z}$ , *Advances in Applied Probability*, **28**, 1014–1033.
- Vallois, P., and C. S. Tapiero (1996) The range process in random walks: Theoretical results and applications, in *Advances in Computational Economics*, H. Ammans, B. Rustem and A. Whinston (Eds), Kluwer, Dordrecht.
- Vallois, P., and C.S. Tapiero (1996) Run length statistics and the Hurst exponent in random and birth–death random walks, *Chaos, Solutions and Fractals*, September. **7**(9), 1333–1341.
- Vallois, P., and C.S. Tapiero (2001) The inter-event range process in birth–death random walks, *Applied Stochastic Models in Business and Industry*.
- Wiggins, J. (1987) Option values under stochastic volatility: theory and empirical estimates, *Journal of Financial Economics*, **5**, 351–372.

**APPENDIX: DEVELOPMENT FOR THE HULL AND WHITE MODEL (1987)\***

Consider the stochastic volatility model in which volatility is a geometric Brownian motion. This is written as follows:

$$\begin{aligned} dS/S &= \mu dt + \sqrt{V} dw, S(0) = S_0 \\ dV/V &= \alpha dt + \beta\sqrt{V} dz, V(0) = V_0 \end{aligned}$$

where  $(w, z)$  are two Brownian motions with correlation  $\rho$ . A call option price  $C(t, S, V)$  would in this case be a function of time,  $S$  and  $V$ . Since there are two sources of risk, the hedging (replicating) portfolio must reflect this multiplicity of risks. Use Ito’s Lemma and obtain:

$$\begin{aligned} dC &= \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial S} dS + \frac{\partial C}{\partial V} dV + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} [dS]^2 + \frac{1}{2} \frac{\partial^2 C}{\partial V^2} [dV]^2 \\ &\quad + \frac{\partial^2 C}{\partial S \partial V} [dS dV] \end{aligned}$$

It is a simple exercise to show that we have:

$$dC = \left[ \begin{array}{l} \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} \alpha S + \frac{\partial C}{\partial V} \mu V \\ + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} [\sigma S]^2 + \frac{1}{2} \frac{\partial^2 C}{\partial V^2} [\xi V]^2 \\ + \frac{\partial^2 C}{\partial S \partial V} (\mu V dt + \rho \xi \sigma S V) \end{array} \right] dt + \frac{\partial C}{\partial S} [\sigma S dW] + \frac{\partial C}{\partial V} [\xi V dZ]$$

The first term in the brackets is a deterministic component while the remaining ones are stochastic terms that we seek to reduce by hedging. For example, if we construct a replicating portfolio consisting a riskless asset and a risky one, we have a portfolio whose value is  $X = nS + (X - nS)$  where the investment in a bond is  $B = (X - nS)$ . Further, we equate the replicating portfolio  $X$  and its differential  $dX$  with the option value and its differential ( $C, dC$ ), leading to:

$$dX = n dS + dB = dC$$

Since for the riskless bond  $dB = R_f B dt$ , we have also  $dC - n dS = R_f B dt$  which leads to:

$$\begin{aligned} 0 &= dC - n dS - R_f B dt \\ &= \left[ -R_f B + \left( \frac{\partial C}{\partial S} - n \right) \alpha S + \frac{\partial C}{\partial V} \mu V dt + \frac{\partial C}{\partial t} + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} [\sigma S]^2 \right. \\ &\quad \left. + \frac{1}{2} \frac{\partial^2 C}{\partial V^2} [S \xi]^2 + \frac{\partial^2 C}{\partial S \partial V} [\rho S V \sigma \xi] \right] dt + \left( \frac{\partial C}{\partial S} - n \right) \sigma S dW + \frac{\partial C}{\partial V} \xi V dZ \end{aligned}$$

For a hedging portfolio we require:

$$\left( \frac{\partial C}{\partial S} - n \right) \sigma S = 0 \rightarrow n = \frac{\partial C}{\partial S} \quad \text{as well as} \quad \frac{\partial C}{\partial V} \xi V \neq 0$$

which is clearly a nonreplicating portfolio. Of course, if the volatility is constant then,  $\partial C / \partial V = 0$ . In this case we have a replicating portfolio, however. For this reason it is essential to seek another asset. There are a number of possibilities to consider but for our current purpose we shall select another option. The replication portfolio is then:

$$X = n_1 S + n_2 C_2 + (X - n_1 S - n_2 C_2), \quad B = (X - n_1 S - n_2 C_2)$$

where  $n_1, n_2$  are the number of stock shares and an option of different maturity. In this case, proceeding as before and replicating the option cash process by the portfolio, we have  $dC_1 = dX$  which implies:

$$dC_1 - n_1 dS - n_2 dC_2 = R_f B dt = R_f (C_1 - n_1 S - n_2 C_2) dt$$

and therefore,

$$(dC_1 - R_f C_1 dt) - n_1(dS - R_f S dt) - n_2(dC_2 - R_f C_2 dt) = 0$$

This provides the equations needed to determine a hedging portfolio. Set,

$$d\Phi_1 = (dC_1 - R_f C_1 dt) \rightarrow \Phi_1 = e^{-R_f t} C_1$$

$$d\Phi_2 = (dS - R_f S dt) \rightarrow \Phi_2 = e^{-R_f t} S$$

$$d\Phi_3 = (dC_2 - R_f C_2 dt) \rightarrow \Phi_3 = e^{-R_f t} C_2$$

and

$$d\Phi_1 = n_1 d\Phi_2 + n_2 d\Phi_3 \quad \text{or} \quad d\left(\frac{C_1}{e^{R_f t}}\right) = n_1 d\left(\frac{S}{e^{R_f t}}\right) + n_2 d\left(\frac{C_2}{e^{R_f t}}\right)$$

Further, if we write:

$$\begin{aligned} \frac{dC_1}{C_1} &= \mu_1 dt + \sigma_1 dW_1 \text{ and set } \lambda_1 = \frac{(\mu_1 - R_f)}{\sigma_1} \text{ or } \frac{dC_1}{C_1} - R_f dt \\ &= \sigma_1 (\lambda_1 dt + dW_1) = \sigma_1 dW_1^* \end{aligned}$$

With  $(\lambda_1 dt + dW_1) = dW_1^*$ , the risk-neutral measure. Applying a CAPM risk valuation, we have:

$$\frac{1}{dt} E \left( \frac{dC_1}{C_1} - R_f \right) = \sigma_1 \lambda_1 = [(R_p - R_f)\beta_{cp} + (R_V - R_f)\beta_{cV}]$$

where  $(R_p)$  is the stock mean return,

$$\beta_{cp} = \frac{p}{C_1} \frac{\partial C_1}{\partial p} \beta_p$$

is the stock beta,  $(R_V)$  is the volatility drift, while

$$\beta_{cV} = \frac{V}{C_1} \frac{\partial C_1}{\partial V} \beta_V$$

is the beta due to volatility. We therefore obtain the following equations:

$$\begin{aligned} \frac{1}{dt} E \left( \frac{dC_1}{C_1} \right) &= R_f + \left[ (\alpha - R_f) \frac{S}{C_1} \frac{\partial C_1}{\partial S} \beta_p + (\mu - S) \beta_V \frac{V}{C_1} \frac{\partial C_1}{\partial V} \right] \\ R_p &= \alpha; \quad R_V = \mu, \quad \lambda_V = (\mu - R_f) V \beta_V \end{aligned}$$

where  $\lambda_V$  is the risk premium associated to the volatility. Thus,

$$\frac{1}{dt} E \left( \frac{dC_1}{C_1} \right) = R_f + \left[ (\alpha - R_f) \frac{S}{C_1} \frac{\partial C_1}{\partial S} + \frac{\lambda_V}{C_1} \frac{\partial C_1}{\partial V} \right]$$

which we equate to the option to value. Since

$$\frac{1}{dt} E \left( \frac{dC}{C} \right) = R_f + \left[ (\alpha - R_f) \frac{S}{C} \frac{\partial C}{\partial S} + \frac{\lambda_V}{C} \frac{\partial C}{\partial V} \right]$$

we obtain at last:

$$\begin{aligned} \frac{1}{dt} E \left( \frac{dC}{C} \right) = \frac{1}{dt} E \left\{ \frac{\partial C}{C \partial t} dt + \frac{\partial C}{C \partial S} dS + \frac{\partial C}{C \partial V} dV + \frac{1}{2} \frac{\partial^2 C}{C \partial S^2} [dS]^2 \right. \\ \left. + \frac{1}{2} \frac{\partial^2 C}{C \partial V^2} [dV]^2 + \frac{\partial^2 C}{C \partial S \partial V} [dS dV] \right\} \end{aligned}$$

We equate these last two expressions and replace all terms for  $(dS, dV)$  leading to a partial differential equation in  $(t, S, V)$  we might be able to solve numerically. This equation is given in the text.

# Value at Risk and Risk Management

## 10.1 INTRODUCTION

The definition of risk and its ‘translation’ into a viable set of indices is of practical importance for risk management. Risk is generally defined in terms of the returns variance. This is not the only approach, however. The range, semi-variance (also called the downside risk), duration (measuring, rather, an exposure to risk as we saw in Chapter 8) and the value at risk (VaR or the quantile risk) are some additional examples. Stone (1973) presented a class of risk measures which include most of the empirical and theoretical class of risk measures. It is specified by the following:

$$R_t(W_0, k, A; f) = \left( \int_{-\infty}^A |W - W_0|^k f_t(W) dW \right)^{1/k}$$

Note that  $R_t(\hat{W}, 2, \infty; f)$  denotes the standard deviation. The parameter  $A$  can be used, however, to define semi-variance and other potential measures. This class of risk measures is closely related to Fishburn’s (1977) measure, widely used in economics and finance, or

$$R_{\alpha,t}(f) = \int_{-\infty}^t (t - W)^\alpha f(W) dW$$

It includes the variance and the semi-variance as special cases and  $\alpha$  and  $t$  are two parameters used to specify the attitude to risk. Practical measurements of risk are extremely important for financial risk management. Artzner *et al.* (1997, 1999) and Cvitanic and Karatzas (1999), for example, have sought to provide a systematic and coherent definition of risk measurement, consistent with applications to financial products, and related to issues of:

- diversification (through portfolio design),
- the use of derivatives for risk management,
- real options valuation,
- risk transfer and sharing (in insurance for example).

Artzner *et al.* (1997, 1999) have proposed a definition in terms of the required reserve for the wealth state  $W$  of an insurance firm (and thus appropriate for a regulator), denoted by  $R(W)$ . The required properties from such a 'reserve function' are as follows:

- (1) Monotonicity, meaning that the riskier the wealth, the larger the required reserve.
- (2) Invariance by drift, meaning that if wealth increases by a fixed quantity, then the required reserve remains the same.
- (3) Homogeneity, meaning that reserves are proportional to wealth, or in mathematical terms  $R_t(aW) = aR_t(W)$ .
- (4) Sub-additivity, and therefore  $R_t(W + W^*) \leq R_t(W) + R_t(W^*)$ .

In particular, properties (3) and (4) imply convexity of the reserve function while (4) implies the economic usefulness of merging portfolio holdings (i.e. diversification). These axioms can be problematic, however (Shiu, 1999). For example, homogeneity is not always reasonable. An insurance company may charge  $X$  dollars to insure a person for a million dollars. It will not charge  $1000X$  to insure the same person for a billion dollars. Another way to put it is that an insurance company would rather insure one-tenth of each of ten ships rather than insure all of one ship.

Although VaR or 'value at risk', is a widely applied measure of risk, it does not satisfy all the properties specified by Artzner *et al.* (1997). It is essentially a quantile measure of risk expressing the expected loss resulting from potential adverse market movements with a specified probability over a period of time. The advantage of VaR is that it provides a single number which encapsulates the portfolio risk and which can be applied easily by non-technically minded financial risk managers. Its origin can be traced to the '4:15 report' of Dennis Weatherstone, chairman of JP Morgan who demanded that a one-page report be delivered to him every day summarizing the company's market exposure and providing an estimate of the potential loss over the next trading day. By the same token, CAR, or capital adequacy ratio (Jorion, 1997) is also used to compensate for market risk exposure. In this case, financial institutions set aside a certain amount of capital so that the probability that the institution will not survive adverse market conditions remains very small.

Although the concept of VaR has its origin in the investment banking sector, the recent generalization of its use by financial institutions is largely due to regulatory authorities. In April 1995, the Basle Committee announced that commercial banks could use the results given by their internal model to compute the level of regulatory capital corresponding to their market risk. The Basle Committee officially recognized VaR as sound risk-management practice as it adopted the

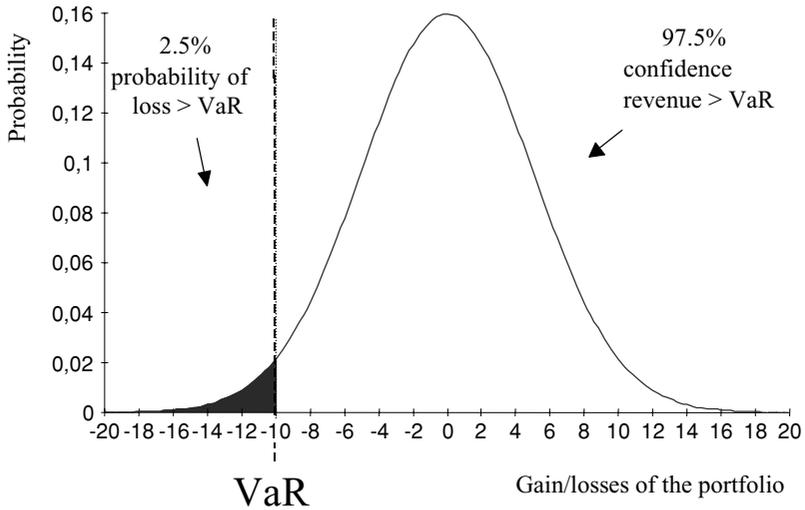


Figure 10.1 The VaR.

formula for the level of capital  $C$  given by:

$$C_{t+1} = \text{Max} \left[ \text{VaR}_t, (M + m) \frac{1}{60} \sum_{j=t-60}^{t-1} \text{VaR}_j \right]$$

where  $M$  is a factor whose value is arbitrarily set to 3 to provide near-absolute insurance for bankruptcy,  $m$  is an additional factor whose value is between 0 and 1 and depends on the quality of the prediction of the internal model developed by the institution. Finally,  $\text{VaR}_j$  is the VaR calculated in the  $j$ th day. Recently, Basle capital 2 has suggested a uniform specification (.99, 10), meaning a safety factor of 99 % over 10 days.

A graphical representation of the VaR can be shown by the probability distribution of Figure 10.1, expressing the potential (and probabilistic) variations in assets holdings over a specified period of time. Practically, it is not simple to assess such a distribution, since many adverse and unpredictable events can be rare and, therefore, mostly ignored. In addition, risks may arise from several independent sources that may be difficult to specify in practice and the interaction effects of which may be even more difficult to predict. Extensive research is pursued that seeks to provide better probability estimates of the risks firms are subject to.

## 10.2 VaR DEFINITIONS AND APPLICATIONS

VaR defines the loss in market value of say, a portfolio, over the time horizon  $T$  that is exceeded with probability  $1 - P_{\text{VaR}}$ . In other words, it is the probability

that returns (losses), say  $\xi$ , are smaller than  $-\text{VaR}$  over a period of time (horizon)  $T$ , or:

$$P_{\text{VaR}} = P(\xi < -\text{VaR}) = \int_{-\infty}^{-\text{VaR}} P_T(\xi) d\xi$$

where  $P_T(\cdot)$  is the probability distribution of returns over the time period  $(0, T)$ . When the returns are normal, VaR is equivalent to using the variance as a risk measure. When risk is sensitive to rare events and extreme losses, we can fit extreme distributions (such as the Weibull and Frechet distributions) or build models based on simulation of VaR. When risks are recurrent, VaR can be estimated by using historical time series while for new situations, scenarios simulation or the construction of theoretical models are needed.

The application of VaR in practice is not without pitfalls, however. In the financial crises of 1998, a bias had made it possible for banks to lend more than they ought to, and still seem to meet VaR regulations. If market volatility increases, capital has to be put up to meet CAR requirements or shrink the volume of business to remain within regulatory requirements. As a result, reduction led to a balance sheet leverage that led to a vicious twist in credit squeeze. In addition, it can be shown, using theoretical arguments, that in nonlinear models (such long-run memory and leptokurtic distributions), the exclusive use of the variance-covariance of return processes (assuming the Normal distribution) can lead to VaR understating risk. This may be evident since protection is sought from unexpected events, while the use of the Normal distribution presumes that risk sources are stable. This has led to the use of scenario and simulation techniques in calculating the VaR. But here again, there are some problems. It is not always possible to compare the VaR calculated using simulation and historical data. The two can differ drastically and therefore, they can, in some cases, be hardly comparable. Nonstationarities of the underlying return processes are also a source of risk since they imply that return distributions can change over time. Thus, using stationary parameters in calculating the VaR can be misleading again. Finally, one should be aware that VaR is only one aspect of financial risk management and therefore over-focusing on it, may lead to other aspects of risk management being neglected. Nonetheless, VaR practice and diffusion and the convenient properties it has, justifies that we devote particular attention to it. For example, the following VaR properties are useful in financial risk management:

- Risks can be aggregated over various instruments and assets.
- VaR integrates diversification effects of portfolios, integrating a portfolio risk properties. This allows the design of VaR-efficient portfolios as we shall see later on.
- VaR provides a common language for risk management, applicable to portfolio management, trading, investment and internal risk management.
- VaR is a simple tool for the selection of strategic risk preferences by top management that can be decomposed in basic components and applied at

**Table 10.1** VaR methods and parameters in a number of banks.

Bank	Technique applied	Confidence interval	Holding period
Abbey National	Variance/covariance	95 %	1 day
Bank Paribas	Monte Carlo simulations	99 %	1 day
BNP	Variance/covariance	99 %	1 day
Deutsche Bank Paris	Variance/covariance	99 %	1 day
Société Générale	Historical simulation	99 %	10 days
Basle Committee	Variance/covariance, historical simulation, Monte Carlo	99 %	10 days or 1 (days* $\sqrt{10}$ )

various levels of a portfolio, a firm or organization. Thus, risk constraints at all levels of a hierarchical organization can be used coherently.

- Finally, VaR can be used as a mechanism to control the over-eagerness of some traders who may assume unwarranted risks as this was the case with Barings (see also Chapter 7).

#### **Example: VaR in banking practice**

Generally, the VaR used by banks converges to the Basle Committee specifications as shown in Table 10.1. Most banks are only beginning to apply VaR methodologies, however. At the Société Générale, for example, market risk measurement consists in measuring the potential loss due to an accident that may occur once in 10 years (with a confidence interval of 99.96 %). Correlations are ignored, leading thereby to an underestimation of this risk. This too, led to the introduction of the VaR approach including correlations, with a probability of 99 %. The first technique was maintained, however, in order to perform stress tests required by regulators. In most banks the time period covered was one day, while the Basle Committee requires 10 days. This means that VaR measures must be aggregated. If risks are not correlated over time, then aggregation is simple, summarized by their sum (and thus leading to a linear growth of variance). In this case, to move from a one day to a 10-day VaR we calculate:

$$VaR_{10\text{ days}} = VaR_{1\text{ day}} * \sqrt{10}$$

When there is an inter-temporal correlation, or long-run memory, the calculation of the VaR is far more difficult (since the variance is essentially extremely large and the measurement of VaR not realistic). Regulatory institutions, operating at international, EU and national levels have pressured banks to standardize their measurements of risks and the application of controls, however. In particular, the CRB operating by a regulation law of 21 July 1995 defines four market risks that require VaR specification:

- (1) Interest-rate risk: relating to obligations, titles, negotiable debts and related instruments.

- (2) Price risk of titles and related instruments.
- (3) Regulated credit risk
- (4) Exchange risk.

Financial institutions have to determine market positions commensurate with Basle (VaR) regulation for these risks.

### Numerical examples

(1) Say that a firm buys in a position in DM that produces a loss if the dollar appreciates (and a profit if it depreciates). We seek to determine the VaR corresponding to the maximum loss that can be sustained in 24 hours with a 5 % probability. Assume that the returns distributions are stable over time. Based on data covering the period of 2000 for the DM/USD exchange rate, a histogram can be constructed and used to calculate the probability of an appreciation larger than 5 %. The critical value for such dollar appreciation turns out to be 1.44 %; which is applied to a market value of \$1 million, and thereby leading to a VaR of \$14 400.

(2) A US investor assumes a long position of DM 140 millions. Volatility in DM/USD FX is: 0.932 while the exchange rate is 1.40 DM/USD. As a result, the VaR for the US investor is calculated by:

$$VaR_{USD} = 140 \text{ millions DM} * 0.932 \% / 1.40 \text{ DM/USD} = 932\,000 \text{ USD}$$

(3) We define next two positions, each with its own VaR calculation,  $VaR_1$ ,  $VaR_2$  respectively. The VaR of both positions is thus

$$VaR = \sqrt{VaR_1^2 + VaR_2^2 + 2\rho_{12}VaR_1VaR_2}$$

where  $\rho_{1,2}$  is the correlation of the positions. The US investor assumes then a long position of DM 140 million in a German bond for 10 years (thereby maintaining a long position in DM). The bond volatility is calculated to be: 0.999 % while the DM/USD volatility is: 0.932 % with correlation  $-0.27$ . The interest-rate and exchange-rate risks are thus:

$$\text{Interest-rate risk: } 140 \text{ million DM} * 0.999 \% / 1.40 = 999\,000 \text{ USD}$$

$$\text{Exchange-rate risk: } 140 \text{ million DM} * 0.932 \% / 1.40 = 932\,000 \text{ USD}$$

As a result, the VaR in both positions is:

$$\begin{aligned} VaR_{USD} &= \sqrt{(999\,000)^2 + (932\,000)^2 + 2 * (-0.27) * 999\,000 * 932\,000} \\ &= 1.17 \text{ million USD} \end{aligned}$$

Non-diversifiable risks that have perfect correlation have a risk which is simply their sum. A correlation 1 would thus lead to a risk of  $999\,000 \text{ USD} + 932\,000 \text{ USD} = 1.93 \text{ million USD}$ . When risk is diversifiable the risk can be reduced with

the difference benefiting the investor. In our case, this is equal to: (\$1.93 million – \$1.17 million = \$760 000).

### 10.3 VaR STATISTICS

VaR statistics can be calculated in a number of ways. These include the analytic variance–covariance approach, the Monte Carlo simulation and other approaches based on extreme distributions and statistical models of various sorts. These approaches are outlined next.

#### 10.3.1 The historical VaR approach

The historical VaR generates scenarios directly from historical data. For each day, returns and associated factors are assessed and characterized probabilistically. This approach generates a number of errors, however. First, markets may be nonstationary and therefore the scenarios difficult to characterize. This has the effect of overestimating the leptokurtic effect (due to heteroscedasticity) while application of the VaR using simulation tends to underestimate these effects because it uses theoretical distributions (such as the normal, the lognormal etc.)

#### 10.3.2 The analytic variance–covariance approach

This approach assumes the normality of returns given by a mean and a standard deviation. Thus, if  $X$  has a standard Normal distribution,  $N(0, 1)$ , then  $P(X > 1.65) = 5\%$ . If we consider again the preceding example, the standard deviation relative to the DM/USD exchange can be easily calculated. For 1995, it has a value of 0.7495%. Assuming a zero mean we can calculate for a risk of 5% the value 1.645 as seen above. As a result, we obtain a critical value of  $0.7495 * 1.645 = 1.233\%$ ; and therefore a VaR of 12 330 USD.

In trading rooms, VaRs are the outcomes of multiple interacting risk factors (exchange-rate risk, interest-rate risk, stock price risks etc. that may be correlated as well). Their joint effects are, thus, assessed by the variance–covariance matrix. For example, a portfolio consisting of a long position in DM and a short in yen with a bivariate normal distribution for risk factors has a VaR that can be calculated with great ease. Let  $\sigma_1$  and  $\sigma_2$  be the standard deviations of two risk factors and let  $\rho_{1,2}$  be their correlation, the resultant standard deviation of the position is thus:

$$\sigma_p = 0.5(\sigma_1^2 + \sigma_2^2 + 2 * \rho_{1,2}\sigma_1\sigma_2)^{1/2}$$

This is generalized to more than two factors as well. Problems arise regarding the definition of multivariate risk distributions, however. Mostly, only marginal distributions are given while the joint multivariate distribution is not easily observable – theoretically and statistically. Recent studies have supported the use of Copula distributions. These distributions are based on the development of a multivariate distribution of risks based on the specification of their marginal

distributions. Such an approach is attracting currently much interest due to the interaction effects of multiple risk sources where only marginal distributions are directly observable. There are other techniques for deriving such distributions, however. For example, the principle of maximum entropy (which we shall elaborate below) also provides a mechanism for generating distributions based on the partial information regarding these distributions.

### 10.3.3 VaR and extreme statistics

The usefulness of extreme statistics arises when the normal probability distribution is no longer applicable, underestimating the real risk to be dealt with, or when investors may be sensitive to extreme losses. These distributions are defined over the Max or the Min of a sample of random returns and have ‘fatter’ tails than say the Normal distribution. For example, say that a portfolio has a loss density function  $F_R(x)$  while we seek the distribution of the largest loss of  $n$  periods:

$$Y_n = \text{Max}\{R_1, R_2, \dots, R_n\}$$

Assuming all losses independent (which is usually a strong assumption), let  $R_{(j)}$  be the  $j$ th order statistic, with  $R_{(1)} < R_{(2)} < R_{(3)} < \dots < R_{(n)}$  over  $n$  periods. Let  $F_j(R)$ ,  $j = 1, 2, \dots, n$  denote the density function of  $R_{(j)}$ . Then, the distribution function of the maximum  $Y_n = R_{(n)}$  is given by:

$$Y_n = R_{(n)} = \text{Max}\{R_1, R_2, \dots, R_n\}$$

If we set a target maximum loss VaR with risk specification  $P_{VaR}$ , then we ought to construct a portfolio whose returns distribution yields:

$$P_{VaR} = P(Y_n < -\text{VaR}) = \int_{-\infty}^{-\text{VaR}} dF_n(\xi)$$

In probability terms:

$$F_n(R) = P\{R_{(n)} < R\} = P\{\text{all } R_i < R\} = [F(R)]^n$$

Similarly for the minimum statistic:  $Y_1 = R_{(1)} = \text{Min}[R_1, R_2, \dots, R_n]$ , we have,

$$\begin{aligned} F_1(R) &= P\{R_{(1)} < R\} = 1 - \Pr\{R_{(1)} > R\} = 1 - \Pr\{\text{all } R_i > R\} \\ &= 1 - [1 - F(R)]^n \end{aligned}$$

and generally, for the  $j$ th statistic,

$$F_{(j)}(R) = \sum_{i=j}^n \binom{n}{i} [F(R)]^i [1 - F(R)]^{n-i}$$

By deriving with respect to  $R$ , we have the distribution:

$$f_j(R) = \frac{n!}{(j-1)!(n-j)!} f(R) [F(R)]^{j-1} [1 - F(R)]^{n-j}$$

If we concentrate our attention on the largest statistic, then in a sample of size  $n$ , the cumulative distribution of the Max statistic is:  $F_{Y_n}(x) = [F_R(x)]^n$ . At the limit, when  $n$  is large, this distribution converges to Gumbel, Weibull or a Frechet distributions, depending on the parameters of the distribution. This fact is an essential motivation for their use in VaR analyses. These distributions are given by:

Gumbel:  $F_Y(y) = \exp(-e^{-y}), y \in IR$

Weibull:  $F_Y(y) = \begin{cases} \exp[-(-y)^{-k}] & \text{for } y < 0 (k < 0) \\ 0 & \text{for } y \geq 0 \end{cases}$

Frechet:  $F_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ \exp(-y^{-k}) & \text{for } y > 0 (k > 0) \end{cases}$

The VaR is then calculated simply since these distributions have analytical cumulative probability distributions as specified above and as the example will demonstrate.

**Example: Weibull and Burr VaR**

Assume a Weibull distribution for the Max loss, given by:

$$f(x/\beta) = a\beta x^{a-1} e^{-\beta x^a}, x \geq 0, \zeta, a > 0; F(x) = 1 - e^{-\beta x^a}$$

where the mean and the variance are:

$$E(x) = \beta^a \Gamma\left(\frac{1}{a} + 1\right), \text{var}(x) = \beta^{2a} \left[ \Gamma\left(\frac{2}{a} + 1\right) - \Gamma^2\left(\frac{1}{a} + 1\right) \right]$$

Say that the parameter  $\beta$  has a Gamma probability distribution given by:  $g(\beta) = \delta\beta^{\nu-1} e^{-\delta\beta}$ , a mixture (Burr) distribution results, given by:

$$f(x) = \frac{\alpha \nu x^{\alpha-1}}{(x^\alpha + \delta)^{\delta+1}}, \quad F(x \geq \Lambda) = \left[ 1 + \frac{\Lambda^\alpha}{\delta} \right]^{-\nu}$$

Thus, if the probability of a loss over the  $\Lambda = \text{VaR}$  is  $P_{\text{VaR}}$ , we have:

$$P_{\text{VaR}} = F(x \geq \text{VaR}) = \left[ 1 + \frac{\text{VaR}^\alpha}{\delta} \right]^{-\nu} \text{ and therefore,}$$

$$\text{VaR} = [\delta \{ [P_{\text{VaR}}]^{-1/\nu} - 1 \}]^{1/\alpha}$$

The essential problem with the use of the extreme VaR approach, however, is that for cumulative risks or stochastic risk processes, these distributions are not applicable since the individual risk events are statistically dependent. In this sense, extreme distributions when they are applied to general time-varying stochastic portfolios have limited usefulness.

**Example: The option VaR by Delta–Gamma**

Consider an option value and assume that its underlying asset has a Normal distribution. For simplicity, say that the option value is a function of an asset price and time to maturity. A Taylor series approximation for the option price is

then:

$$V_{t+h} - V_t \approx \delta(P_{t+h} - P_t) + \frac{1}{2}\Gamma(P_{t+h} - P_t)^2 + \theta h$$

$$\delta = \frac{\partial V_t}{\partial P_t}; \Gamma = \frac{\partial^2 V_t}{\partial P_t^2}, \theta = \frac{\partial V_t}{\partial \tau}, \tau = T - t$$

The option’s rate of return is therefore:

$$R_V = \frac{V_{t+h} - V_t}{V_t}, R_p = \frac{P_{t+h} - P_t}{P_t}, \text{ or}$$

$$R_V \approx \delta R_p \eta + \frac{1}{2}\Gamma R_p^2 \eta P^2 + \theta h/V_t; \eta = P_t/V_t$$

Explicitly, we have then:

$$R_V = \eta \delta R_p + 0.5\eta\Gamma P_t R_p^2 + n \left( \frac{\theta}{V_t} \right)$$

which can be used to establish the first four moments of the rates of return  $R_V$  of the option:

	Option ROR	Underlying
Mean:	$R_V$ $0.5\tilde{\Gamma}\sigma_p^2 + \tilde{\theta}n$	$R_p$ 0
Variance:	$\tilde{\delta}^2\sigma_p^2 + 0.5\tilde{\Gamma}^2\sigma_p^4$	$\sigma_p^2$
Skewness:	$3\tilde{\delta}^2\tilde{\Gamma}\sigma_p^4 + \tilde{\Gamma}^3\sigma_p^6$	0
Kurtosis:	$12\tilde{\delta}^2\tilde{\Gamma}^2\sigma_p^6 + 3\tilde{\Gamma}^4\sigma_p^8 + 3\sigma_p^4$	$3\sigma_p^4$

In practice, after we calculate the Greek coefficients delta, gamma, theta and the variance of the underlying furnished by Risk Metrics, the four moments allow an estimation of the distribution. Such an approach can be generalized further to a portfolio of derivatives and stocks, albeit calculations might be difficult.

**10.3.4 Copulae and portfolio VaR Measurement**

In practice, a portfolio consisting of many assets may involve ‘many sources of risks’, which are in general difficult to assess singly and collectively. Empirical research is then carried out to specify the risk properties of the underlying portfolio. These topics are motivating extensive research efforts spanning both traditional statistical approaches and other techniques. Assume that the wealth level of a portfolio is defined as a nonlinear and differentiable function of a number of assets written as follows:

$$W_t = F(P_{1t}, P_{2t}, P_{3t}, \dots, P_{nt}, B)$$

with  $P_{1t}, P_{2t}, P_{3t}, \dots, P_{nt}$ , the current prices of these assets, while  $B$  is a riskless asset. The function  $F$  is assumed for the moment to be differentiable. Then, in a

small interval of time, a first approximation yields:

$$\Delta W_t = \frac{\partial F}{\partial P_{1t}} \Delta P_{1t} + \frac{\partial F}{\partial P_{2t}} \Delta P_{2t} + \dots + \frac{\partial F}{\partial P_{nt}} \Delta P_{nt} + \frac{\partial F}{\partial B} \Delta B$$

If we set  $N_{it} = \partial F / \partial P_{it}$  to be the number of shares bought at  $t$ , we have a portfolio whose process is:

$$\Delta W_t = N_{1t} \Delta P_{1t} + N_{2t} \Delta P_{2t} + \dots + N_{nt} \Delta P_{nt} + C \Delta B$$

where  $C$  is the investment in a riskless asset. Assume for simplification that the rate of return for each risky asset is  $\alpha_{it} \Delta t = E(\Delta P_{it} / P_{it})$  while volatility is  $\sigma_{it}^2 \Delta t = \text{var}(\Delta P_{it} / P_{it})$ . A normal approximation,  $\Delta W_t \sim \mathbb{N}(\mu_t \Delta t, \Lambda_t \Delta t)$  of the portfolio with mean  $\mu_t$  and variance–covariance matrix  $\Lambda_t$  is:

$$\mu_t = C \Delta B + \sum_{i=1}^n N_{it} \alpha_{it} P_{it}; \Lambda_t = \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 F}{\partial P_{it}^2} \sigma_{it}^2 + \sum_{i=j}^n \sum_{j=1}^n \frac{\partial^2 F}{\partial P_{it} \partial P_{jt}} [\rho_{ij} \sigma_{it} \sigma_{jt}]$$

and, the VaR can be measured simply, as we saw earlier. This might not always be possible, however. If the marginal distributions only are given (namely the mean and the variances of each return distribution) it would be difficult to specify the appropriate distribution of the multiple risk returns to adopt.

Copulae are multivariate distributions constructed by assuming that their marginal distributions are known. For example (see Embrecht *et al.*, 2001), given two sources of risks  $(X_1, X_2)$ , each with their own VaR appropriately calculated, what is the VaR of the sum of the two, potentially interacting sources of risks  $(X_1 + X_2)$ ? Intuitively, the worst case VaR for a portfolio  $(X_1 + X_2)$  occurs when the linear correlation is maximal. However, it is wrong in general. The problem at hand is thus how to construct bounds for the VaR of the joint position and determine how such bounds change when some of the assets may be statistically dependent and when information is revealed, as the process unfolds over time. The Copula provides such an approach which we briefly refer to. For example, this approach can be summarized simply by a theorem of Sklar (see Embrechts *et al.*, 2001), stating that a joint distribution  $F$  has marginals  $(F_1, F_2)$  only if there is a Copula such that:

$$F(x_1, x_2) = C [F_1(x_1), F_2(x_2)]$$

Inversely, given the marginals  $(F_1, F_2)$ , the joint distribution can be constructed by choosing an appropriate Copula  $C$ . This procedure is not simple to apply, but recent studies have contributed to the theoretical foundations of this evolving area of study. For example, if the marginal distributions are continuous, then there is a unique Copula. In addition, Copulae structures can contribute to the estimation of multiple risks VaR. Embrecht, Hoing and Juri (2001) and Embrecht, Kluppelberg and Mikosch (1997) have studied these problems in great detail and the motivated reader ought to consult these references first. Current research is ongoing in specifying such functions and their properties.

### 10.3.5 Multivariate risk functions and the principle of maximum entropy

When some characteristics, data or other information regarding the risk distribution are available, it is possible to define its underlying distribution by selecting that distribution which assumes the ‘least’, that is the distribution with the greatest variability, given the available information. One approach that allows the definition of such distributions is defined by the ‘maximum entropy principle’.

Entropy (or its negative, the negentropy) can be simply defined as a measure for ‘departure from randomness’. Explicitly, the larger the entropy the more the distribution departs from randomness and vice versa, the larger the negentropy, the greater a distribution’s randomness. The origins of entropy arose in statistical physics. Boltzmann observed that entropy relates to ‘missing information’ inasmuch as it pertains to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded. In this sense, information can be interpreted as that which changes a system’s state of randomness (or, equivalently, as that quantity which reduces entropy). For example, for a word, which has  $k$  letters, assuming zeros and ones, and one two, define a sequence of  $k$  letters,  $(a_0, a_1, a_2, \dots, a_k)$ ,

$$a_i = \begin{cases} 1 \\ 0 \end{cases}$$

for all  $i \neq j$ ,  $a_j \neq a_i = 2$  and one  $j$ . The total number of configurations (or strings of  $k + 1$  letters) that can be created is  $N$  where,  $N = 2^k(k + 1)$ . The logarithm to the base 2 of this number of configurations is the information  $I$ , or  $I = \log_2 N$  and, in our case,  $I = k + \log_2(k + 1)$ . The larger this number  $I$ , the larger the number of possible configurations and therefore the larger the ‘randomness’ of the word. As a further example, assume an alphabet of  $G$  symbols and consider messages consisting of  $N$  symbols. Say that the frequency of occurrence of a letter is  $f_i$ , i.e., in  $N$  symbols the letter  $G$  occurs on the average  $N_i = f_i N$  times. There may then be  $W$  different possible messages, where:

$$W = \frac{N!}{\prod_{i=1}^N N_i!}$$

The uncertainty of an  $N$  symbols message is simply the ability to discern which message is about to be received. Thus, if,

$$W = e^{HN} \quad \text{and} \quad H = \lim_{N \rightarrow \infty} \frac{1}{N} \log(W) = \sum p_i \log(1/p_i)$$

which is also known as Shannon’s entropy. If the number of configurations (i.e.  $W$ ) is reduced, then the ‘information’ increases. To see how the mathematical and statistical properties of entropy may be used, we outline a number of problems.

#### (1) Discrimination and divergence

Consider, for example, two probability distributions given by  $[F, G]$ , one theoretical and the other empirical. We want to construct a ‘measure’ that makes

it possible to discriminate between these distributions. An attempt may be reached by using the following function we call the ‘discrimination information’ (Kullback, 1959):

$$I(F, G) = \int F(x) \log \frac{F(x)}{G(x)} dx$$

In this case,  $I(F, G)$  is a measure of ‘distance’ between the distributions  $F$  and  $G$ . The larger the measure, the more we can discriminate between these distributions. For example, if  $G(\cdot)$  is a uniform distribution, then divergence is Shannon’s measure of information. In this case, it also provides a measure of departure from the random distribution. Selecting a distribution which has a maximum entropy (given a set of assumptions which are made explicit) is thus equivalent to the ‘principle of insufficient reason’ proposed by Laplace (see Chapter 3). In this sense, selecting a distribution with the largest entropy will imply a most conservative (riskwise) distribution. By the same token, we have:

$$I(G, F) = \int G(x) \log \frac{G(x)}{F(x)} dx$$

and the *divergence* between these two distributions is defined by:

$$J(F, G) = I(F, G) + I(G, F) = \int [F(x) - G(x)] \log \frac{F(x)}{G(x)} dx$$

which provides a ‘unidirectional and symmetric measure’ of distribution ‘distance’, since  $J(F, G) = J(G, F)$ . For a discrete time distribution  $(p, q)$ , we have similarly:

$$I(p, q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}; I(q, p) = \sum_{i=1}^n q_i \log \frac{q_i}{p_i}$$

$$J(p, q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} + \sum_{i=1}^n q_i \log \frac{q_i}{p_i} = \sum_{i=1}^n (p_i - q_i) \log \frac{p_i}{q_i}$$

For example, say that  $q_i, i = 1, 2, 3, \dots, n$  is an empirical distribution and say that  $p_i, i = 1, 2, 3, \dots, n$  is a theoretical distribution given by the distribution:

$$p_i = \binom{n}{i} p^i (1-p)^{n-i}$$

then:

$$J(p, q) = \sum_{i=1}^n \left[ \binom{n}{i} p^i (1-p)^{n-i} - q_i \right] \log \frac{\binom{n}{i} p^i (1-p)^{n-i}}{q_i}$$

which may be minimized with respect to parameter  $p$ . This will be, therefore, a binomial distribution with a parameter which is ‘least distant’ from the empirical

distribution. For a bivariate state distribution, we have similarly:

$$I(p, q) = \sum_{j=1}^m \sum_{i=1}^n p_{ij} \log \left( \frac{p_{ij}}{q_{ij}} \right); \quad J(p, q) = \sum_{j=1}^m \sum_{i=1}^n (p_{ij} - q_{ij}) \log \left( \frac{p_{ij}}{q_{ij}} \right)$$

while, for continuous distributions, we have also:

$$I(F, G) = \iint F(x, y) \log \frac{F(x, y)}{G(x, y)} dx dy$$

as well as:

$$J(F, G) = \iint [F(x, y) - G(x, y)] \log \frac{F(x, y)}{G(x, y)} dx dy$$

This distribution may then be used to provide 'divergence-distance' measures between empirically observed and theoretical distributions.

The maximization of divergence or entropy can be applied to constructing a multivariate risk model. For example, assume that a nonnegative random variable  $\{\theta\}$  has a known mean given by  $\hat{\theta}$ , the maximum entropy distribution for a continuous-state distribution is given by solution of the following optimization problem:

$$\text{Max } H = - \int_0^{\infty} f(\theta) \log [f(\theta)] d\theta$$

subject to:

$$\int_0^{\infty} f(\theta) d\theta = 1, \hat{\theta} = \int_0^{\infty} \theta f(\theta) d\theta$$

The solution of this problem, based on the calculus of variations, yields an exponential distribution. In other words,

$$f(\theta) = \frac{1}{\hat{\theta}} e^{-\theta/\hat{\theta}}, \theta \geq 0$$

When the variance of a distribution is specified as well, it can be shown that the resulting distribution is the Normal distribution with specified mean and specified variance. The resulting optimization problem is:

$$\text{Max } H = \int_{-\infty}^{\infty} f(\theta) \log \frac{1}{[f(\theta)]} d\theta$$

subject to:

$$\int_{-\infty}^{\infty} f(\theta) d\theta = 1, \hat{\theta} = \int_{-\infty}^{\infty} \theta f(\theta) d\theta, \sigma^2 = \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f(\theta) d\theta$$

which yields (as stated above) the normal probability distribution with mean and variance  $(\hat{\theta}, \sigma^2)$ . If an empirical distribution is available, then, of course, we can use the divergence to minimize the distance between these distributions by the appropriate selection of the theoretical parameters. This approach can be applied equally when the probability distribution is discrete, bounded, multivariate distributions with specified marginal distributions etc. In particular, it is interesting to point out that the maximum entropy of a multivariate distribution with specified mean and known variance–covariance matrix turns out also to be a multivariate normal, implying that the Normal is the most random distribution that has a specified mean and a specified variance. Evidently, if we also specify leptokurtic parameters the distribution will not be Normal.

**Example: A maximum entropy price process:**

Consider a bivariate probability distribution (or a price stochastic process)  $h(x, t)$ ,  $x \in [0, \infty)$ ,  $t \in [a, b]$ . The maximum entropy criterion can be written as an optimization problem, maximizing:

$$\text{Max } H = \int_0^{\infty} \int_a^b h(x, t) \log \frac{1}{[h(x, t)]} dt dx$$

Subject to partial information regarding the distribution  $h(x, t)$ ,  $x \in [0, \infty)$ ,  $t \in [a, b]$ . Say that at the final time  $b$ , the price of a stock is for sure  $X_b$  while initially it is given by  $X_a$ . Further, let the average price over the relevant time interval be known and given by  $\bar{X}_{(a,b)}$ , this may be translated into the following constraints:

$$h(X_a, a) = 1, h(X_b, b) = 1 \quad \text{and} \quad \frac{1}{b-a} \int_a^b \int_0^{\infty} x h(x, t) dx dt = \bar{X}_{(a,b)}$$

to be accounted for in the entropy optimization problem. Of course, we can include additional constraints when more information is available. Thus, the maximum entropy approach can be used as an ‘alternative rationality’ to constructing probability risk models (and thereby modelling the uncertainty we face and computing its VaR measure) when the burden of explicit hypothesis formulation or the justification of the model at hand is too heavy.

**Problem: The maximum entropy distribution with specified marginals**

Define the maximum entropy of the joint distribution by specifying the marginal distributions as constraints. Or

$$\text{Max } H = - \int_0^{\infty} \int_a^b F(x, y) \log [F(x, y)] dy dx$$

subject to:

$$F_1(x) = \int_a^b F(x, y) dy; \quad F_2(y) = \int_0^\infty F(x, y) dx$$

and apply the multivariate calculus of variations to determine the corresponding distribution.

### 10.3.6 Monte Carlo simulation and VaR

Monte Carlo simulation techniques are both widely practised and easy to apply. However, applications of simulation should be made critically and carefully. In simulation, one generates a large number of market scenarios that follow the same underlying distribution. For each scenario the value of the position is calculated and recorded. The simulated values form a probability distribution for the value of a portfolio which is used in deriving the VaR figures. It is clear that by using Monte Carlo techniques one can overcome approaches based solely on a Normal underlying distribution. While such distributions are easy to implement, there is an emerging need for better models, more precise, replicating closely market moves. But where can one find these models? Underlying distributions are essentially derived from historical data and the stochastic model used to interpret data. The same historical data can lead to very different values-at-risk, under different statistical models, however. This leads to the question: what model and what distribution to select? Monte Carlo simulation does not resolve these problems but provides a broader set of models and distributions we can select from. In this sense, in simulation the GIGO (garbage in, garbage out) principle must also be carefully apprehended. The construction of a good risk model is by no means a simple task, but one that requires careful thinking, theoretical knowledge and preferably an extensive and reliable body of statistical information.

## 10.4 VaR EFFICIENCY

### 10.4.1 VaR and portfolio risk efficiency with normal returns

VaR, either as an objective or as a risk constraint, has become another tool for the design of portfolios. Gouriéroux *et al.* (2000), Basak and Shapiro (2001) have attracted attention to this potential usefulness of the VaR. To see how VaR-efficient portfolios can be formulated we proceed as follows: let  $W_t(\cdot)$  be the value of a portfolio at time  $t$  and let, at a time  $t + h$  later, wealth be  $W_{t+h}(\cdot)$ . The probability of the value at risk in the time interval  $(t, t + h)$  is then given by:

$$\Pr[W_{t+h} - W_t < -Var_W(h)] = P_{VaR}$$

Thus, if the return is  $R_W(h) = W_{t+h} - W_t$  we can write instead:

$$\Pr[R_W(h) < -Var_W(h)] = P_{VaR}$$

Models will thus vary according to the return model used in value at risk estimation. For our purposes, let  $N$  be the number of financial assets,  $P(i, t)$  be the price of asset  $i$  at time  $t$  and  $B(t)$  be the budget at time  $t$ . We also set  $a_1, a_2, a_3, \dots, a_N$  be the number of shares held of each asset at time  $t$ . As a result, the wealth state at time  $t$  is:

$$W_t(a) = \sum_{i=1}^n a_i p_{i,t} = a^T p_t$$

while the return in a time interval is:

$$W_{t+1}(a) - W_t(a) = a^T (p_{t+1} - p_t)$$

The loss probability  $\alpha = P_{VaR}$  implies a VaR calculated in terms of the portfolio holdings and the risk constraint, or  $VaR_t(a, \alpha)$  is defined by:

$$P_t [W_{t+1}(a) - W_t(a) < -VaR_t(a, \alpha)] = \alpha$$

where  $P_t$  is the conditional distribution of future asset prices given the information available at time  $t$ . Set the price change  $y_{t+1} = p_{t+1} - p_t$ . Then we have equivalently,

$$P_t [-a^T y_{t+1} > VaR_t(a, \alpha)] = \alpha$$

In other words, the portfolio VaR is determined in terms of information regarding past prices, the portfolio composition and the specified loss probability  $\alpha$ . When the price changes are normally distributed:

$$y_{t+1} \sim \mathbb{N}(\mu_t, \Omega_t) \text{ then } VaR_t(a, \alpha) = -a^T \mu_t + (a^T \Omega_t a)^{1/2} Z_{1-\alpha}$$

where  $Z_{1-\alpha}$  is the  $1 - \alpha$  quantile of the standard normal distribution. We might consider then a number of portfolio design approaches.

### Problem 1

Maximize expected returns subject to a VaR constraint.

### Problem 2

Minimize a VaR risk subject to returns constraints.

The first problem is stated as follows:

$$\text{Max}_a E(a^T y_{t+1}) \text{ subject to } VaR_t(a, \alpha) \leq VaR_0$$

where  $VaR_0$  is a VaR constraint specified by management. The solution of this problem will turn out to be a portfolio allocation which is necessarily a function of the risk parameters  $\{\alpha, VaR_0\}$ . By the same token, we can state the second problem as follows:

$$\text{Min}_a VaR_t(a, \alpha) \text{ subject to } E(a^T y_{t+1}) \geq \bar{R}$$

where  $\bar{R}$  is the required mean return. A solution to these problems by analytical or numerical means is straightforward.

### 10.4.2 VaR and regret

The Savage (1954) regret criterion has inspired a number of approaches called ‘regret–disappointments models’ (Loomes and Sugden, 1982, 1987). According to Bell (1982), disappointment is a psychological reaction to an outcome that does not meet a decision maker’s expectation. In particular, Bell assumed that the measurement of disappointment is assumed to be proportional to the difference between expectation and the outcome below the expectation. Elation, may occur when the outcome obtained is better than its expectation. Usually (for risk-averse decision makers), the ‘cost of reaching the wrong decision’ (disappointment) may be, proportionately, greater than the payoff of having made the right decision (elation). In other words, managers abhor losses, valuing them more than they value gains for having made the right decision. The inverse may also be true. A trader whose income is derived from trade-ins only may be tempted to assume risks which may be larger than the investment firm may be willing to assume. For example, a stock performance below that expected by analysts can have disproportionate effects on stock values while an uncontrollable trader may lead to disaster. To see how to proceed, consider the utility of an investment given by  $u(x)$  and let  $CE$  be its certainty equivalent. That is to say, as we saw in Chapter 3:

$$u(CE) = Eu(x) \quad \text{and} \quad CE = u^{-1} [Eu(x)]$$

Say that ‘disappointment’ occurs when the outcome is below  $CE$ , resulting in an expected utility that is depreciated due to disappointment. As a result, an adjusted expected utility, written by  $V(b)$  with a factor  $b$  can be specified by:

$$V(b) = Eu(x) - bE [u(CE) - u(x) | CE \geq x]$$

Explicitly, it equals the expected utility less the expected utility loss when the ex-post event is below the certainty equivalent,

$$V(b) = Eu(x) - b [Eu(x) - u(x) | CE \geq x]$$

Since  $Eu(x) = u(CE)$ , we also have:

$$\begin{aligned} V(b) &= Eu(x) - b [u(CE) - u(x) | CE \geq x] \\ &= u(CE) (1 - bF(CE)) + bEu(x | x \leq CE) \end{aligned}$$

This procedure allows us to resolve the problem of a reference point endogenously. For example, consider a two-event process:

$$x = (x_1, x_2) \text{ with prob } (\alpha, 1 - \alpha) \text{ with } x_1 > x_2$$

which corresponds to a binomial model. Then, the expected utility is:

$$Eu(x) = u(CE) = \alpha u(x_1) + (1 - \alpha)u(x_2)$$

while

$$V(b) = [\alpha u(x_1) + (1 - \alpha)u(x_2)] [1 - b(1 - \alpha)] + bu(x_2)(1 - \alpha)$$

and therefore:

$$V(b) = \alpha [1 - b(1 - \alpha)] u(x_1) + (1 - \alpha) (1 + b\alpha) u(x_2)$$

which we can write as follows:

$$V(b) = \alpha\beta u(x_1) + (1 - \alpha)(\beta + b)u(x_2); \beta = [1 - b(1 - \alpha)]$$

Since  $b > 0$ , it provides a greater weight to the lower outcome. Further,

$$\Delta = V(b) - Eu(x) = (1 - \beta)[u(x_1) - \alpha u(x_2)]$$

By applying the same argument to the VaR by assuming a linear truncated utility, we have:

$$V(b, VaR) = E(x) + bE[-VaR + x | -VaR > x]$$

As a result, the expected value is:

$$V(b, VaR) = E(x) + b \int_{-\infty}^{-VaR} (x - VaR) dF(x)$$

For example, say that an option price is  $C(0)$  while the strike price over one period is  $K$  and assume two states  $x_1, x_2, x_1 \leq K \leq x_2$ . Thus, the expected value of the option price is:

$$C(0) = \frac{1}{1+r} \left\{ \frac{\alpha}{1+b(1-\alpha)} u(0) + \frac{(1+b)(1-\alpha)}{1+b(1-\alpha)} u(x_2 - K) \right\}$$

If the option price can be observed together with the utility of  $u(0)$ , then the equivalent parameter  $b$  is found to be equal to:

$$0 < b = \frac{\alpha u(0) + (1 - \alpha)u(x_2 - K) - C(0)(1 + r)}{(1 - \alpha)[C(0)(1 + r) - u(x_2 - K)]}$$

Note that  $\alpha u(0) + (1 - \alpha)u(x_2 - K) < C(0)(1 + r)$ , meaning that we require for a positive disappointment parameter (since we have necessarily,  $C(0)(1 + r) < u(x_2 - K)$ ) that the option price be larger than the expected value of the utility of the option lottery. In other words, the option price also includes the investor's desire for the potential to reduce deception by allowing him to regret and reverse a decision badly taken.

## REFERENCES AND ADDITIONAL READING

- Alexander, C.O. (1998) *The Handbook of Risk Management and Analysis*, John Wiley & Sons, Inc., New York.
- Arrow, K.J. (1971) *Essays in the Theory of Risk Bearing*, Markham, Chicago.
- Artzner, P., F. Delbaen, J.M. Eber and D. Heath (1997) Thinking coherently, *Risk*, **10**, 67–71.
- Artzner, P., F. Delbaen, J. Elber and D. Heath (1999) coherent measures of risk, *Mathematical Finance*, **9**, 203–228.
- Basak, S., and A. Shapiro (2001) Value-at-risk-based risk management: Optimal policies and asset prices, *The Review of Financial Studies*, **14**, 371–405.
- Basle Committee (1996) Amendment to the capital accord to incorporate market risks, on banking supervision, January.

- Bauer, C. (2000) Value at risk using hyperbolic distributions, *Journal of Economics and Business*, **52**, 455–467.
- Bell, D.E. (1982) Regret in decision making under uncertainty, *Operations Research*, **30**, 961–981.
- Bell, D.E. (1983) Risk premiums for decision regrets, *Management Science*, **29**, 1156–1166.
- Bell, D.E. (1985) Disappointment in decision making under uncertainty, *Operation Research*, **33**, 1–27.
- Cvitanic, J., and I. Karatsas (1999) On dynamic measures of risk, *Finance and Stochastics*, **3**, 451–482.
- Dowd, K. (1998) *Beyond Value at Risk, The New Science of Risk Management*, John Wiley & Sons, Ltd, Chichester.
- Duffie, D., and J. Pan (1997) An overview of value at risk, *Journal of Derivatives*, **4**, 7–49.
- Embrecht, P., A. Hoing and A. Juri (2001) Recent advances in the application of copulae to non-linear Value at Risk, ETH Department of Mathematics, CH-8092 Zurich, April.
- Embrecht, P., C. Kluppelberg and T. Mikosch (1997) *Modelling Extremal Events*, Springer-Verlag, Berlin.
- Feller, W. (1957, 1966) *An Introduction to Probability Theory and its Applications*, Vols 1 and 2, John Wiley & Sons, Inc., New York.
- Fishburn, P. (1977) Mean risk analysis with risk associated with below target returns, *American Economic Review*, **67**(2), 116–126.
- Gourieroux, C., J.P. Laurent and O. Scaillet (2000) Sensitivity analysis of values at risk, *Journal of Empirical Finance*, **7**, 225–245.
- Gul, Faruk (1991) A theory of disappointment aversion, *Econometrica*, **59**, 667–686.
- Jia, J., and J.S. Dyer (1994) Risk-value theory, Department of Management Science and Information Systems, The Graduate School of Business, University of Texas, Austin, WP, 94/95-3-4.
- Jorion, P. (1997) *Value at Risk: The New Benchmark for Controlling Market Risk*, McGraw-Hill, Chicago.
- Jorion, P. (1999) Risk management lessons from long term capital management, Working Paper, University of California, Irvine.
- Kullback S. (1959) *Information Theory and Statistics*, Wiley, New York.
- Loomes, G., and R. Sugden (1982) Regret theory: An alternative to rational choice under uncertainty, *Economic Journal*, **92**, 805–824.
- Loomes, G., and R. Sugden (1987) Some implications of a more general form of regret theory, *Journal of Economic Theory*, **41**, 270–287.
- Morgan, J.P. (1995) *Introduction to RiskMetrics<sup>TM</sup>*, 4th Edition, November.
- Robinson, Gary (1995) Directeur, R&D Division, Global Market Risk Management, BZW, *Value-at-Risk Analysis: Its Strengths and Limitations*, October.
- Savage, L.J. (1954) *The Foundations of Statistics*, John Wiley & Sons, Inc., New York.
- Schachter, B. (2002) All about value at risk, [www.GloriaMundi.org/var](http://www.GloriaMundi.org/var).
- Shiu, E.W. (1999) Discussion of Philippe Artzner's, 'Application of coherent risk measures to capital requirements in insurance', *North American Actuarial Journal*, **3**(2).
- Shore, H. (1986) Simple general approximations for a random variable and its inverse distribution function based on linear transformations of nonskewed variate, *SIAM Journal on Scientific and Statistical Computing*, **7**, 1–23.
- Shore, H. (1995) Identifying a two parameter distribution by the first two sample moments (partial and complete), *Journal of Statistical and Computational Simulation*, **52**, 17–32.
- Stone, B. (1973) A general class of three parameter risk measures, *Journal of Finance*, **28**, 675–685.
- Tapiero, C.S. (2003) VaR and inventory control, *European Journal of Operations Research*, forthcoming.
- Telser, L.G. (1956) Safety first and hedging, *Review of Economic Studies*, **23**, 1–16.

# Author Index

- Akerlof, G. 68, 226  
Allais' Paradox 52–3  
Amikam, Meir 199  
Aquila 197  
Arnold, L. 92  
Arrow, K.J. 44, 212  
Artzner, P. 309, 310  
Augustine, St 273
- Bachelier, L. 16, 84  
Barrois, T. 7  
Basak, S. 324  
Bell, D.E. 29, 326  
Bensoussan, A. 92, 139, 154  
Bismut, J.M. 92  
Black, F. 226, 252  
Bloomberg 300  
Bolder, David 253  
Bollerslev, T. 280  
Borch, K.H. 8  
Born, M. 79  
Broadie, M. 154  
Brock, W.A. 92  
Brown, Robert 88
- Capocelli, R.M. 209  
Cardano, Girolamo 81–2  
Carr, P. 154, 157  
Cochrane, John H. 66, 67  
Connolly, K.B. 184  
Cotton, Peter 258  
Cox, D.R. 92, 192, 296  
Cox, J.C. 226  
Cox–Ingersoll–Ross (CIR) model 257  
Cramer, Gabriel 51–2  
Cramer, H. 8  
Cvitanic, J. 309
- Darling, D.A. 209  
Delbaen, F. 253  
Detemple, J. 154  
Dixon, Hugo 68  
Donsker Theorem 99–100  
Doob, J.L. 82  
Duff & Phelps 12  
Duffie, G.R. 230
- Eeckoudt, L. 44  
Einstein, Albert 84  
Elliot, R.J. 92  
Embrecht, P. 319  
Engle, R.F. 280
- Fama, E.F. 88, 112, 274  
Feller, W. 51–2, 194  
Fielitz, B. 301  
Filipovic, D. 253  
Fishburn, P. 309  
Fisher, Irving 212  
Fisher, R.A. 6  
Fokker–Planck equation 88  
Foster, D.P. 280  
Fouque, Jean-Pierre 258  
Friedman, Milton 112
- Ghashghaie, S. 80  
Goldman, M.B. 184  
Gourieroux, C. 324  
Green, M.T. 301
- Hogan, M. and Weintraub, K. 252  
Hoing, A. 319  
Huang, J. 154  
Hull, J. 258  
Hurst, H.E. 300

- Ito, K. 92
- Jarrow, R. 227, 242
- Jorion, P. 310
- Julien, H. 139
- Juri, A. 319
- Kahnemann, D. 53
- Kalman, R.E. 79
- Kappa 183
- Karasinski, P. 252
- Karatzas, I. 309
- Kay, John 9
- Kim, I.J. 156–7
- Kimball, M. 44, 45
- Kluppenberg, C. 319
- Koch 197
- Kolmogorov 79, 84
- Kortanek, K.O. 253
- Kullback, S. 291, 321
- Leeson, Nick 202
- Leland, H.E. 226
- Levy, P. 82, 84
- Lintner, J. 63
- Lo, Andrew W. 272
- Longstaff, F. 226
- Loomes, G. 29, 326
- Lorenz, Edward 79, 80
- Lorimier, S. 253
- Lucas, R.E. 112, 113
- Luce, R.D. 7
- Lundberg, F. 8
- Machina, M.J. 53
- Magill, M. 111, 212
- Malkiel, Burton 13
- Malliaris, A.G. 92
- Mandelbrot, B. 88, 272, 300, 301
- Markowitz, Harry 62–3
- McKean, H.P. 92
- Medvedev, V.G. 253
- Merrill Lynch 68, 200
- Merton, R.C. 153, 226
- Mikosch, T. 319
- Miller, H.D. 92, 192
- Miller, M.H. 112
- Milshtein, G.N. 100
- Morgan, J.P. 197, 200
- Morgenstern Oskar, 41
- Muth, John 111, 112, 293
- Nelson, C.R. and Siegel, A.F., approach to interest rates 253
- Nelson, Daniel 96, 280
- Nelson, Erick 199
- Papanicolaou, George 258
- Pearson, Karl 6
- Pliska, S.R. 139, 146
- Pratt, J.W. 44
- Quiggin, J. 53
- Quinzii, M. 111, 212
- Raiffa, H. 7
- Ramaswamy, K. 96
- Reyniers, D. 73, 75
- Ricciardi, L.M. 209
- Roll, Richard 112
- Rubinstein, Ariel 20
- Samuelson Paul 112
- Sandmann, K. 252
- Sargent, T.J. 112
- Savage, L.J. 29
- Schlaiffer, R. 7
- Schwartz, E. 226
- Shapiro, A. 324
- Sharpe, William 63
- Shiu, E.W. 310
- Shreve S.E. 139
- Siegel, A.F. 253
- Siegert, A.J.F. 209
- Singleton, K.J. 230
- Sircar, K. Ronnie 258
- Sklar 319
- Sonderman, D. 252
- Soros, George 84
- Spence, M. 72
- Stone, B. 309
- Sugden, R. 29, 326
- Svensson 253
- Szekely, Gabor 82
- Tapiero, Charles 92, 257
- Taylor H.M. 280

- Tetens, J.N. 7  
Tobin, James 63  
Turnbull, S. 227  
Tverski, A. 53  
  
Van Ness, J.W. 301  
Vasicek 254–8, 258  
Ville, J. 82  
Von Neumann John 41  
  
Walras, L. 212  
Weatherstone, Dennis 310  
Wets, Roger J.B. 253  
White, A. 258  
Wiener, Norbert 84  
Willasen, Y. 45  
Wilmott, P. 182, 183, 262  
  
Yu, G. 156–7

*Index compiled by Annette Musker*

# Subject Index

- ABN AMRO Asset Management 21, 202
- absolute risk aversion, index of 44
- absorbing state 207
- accounting rate of return (ARR) 54–5
- actuarial science, insurance and 7–10
- adverse selection 68–9, 71
- affine structure 254, 255
- affine term structure (ATS) model 261
- Allais' Paradox 52–3
- Allied Irish Bank (AIB) 24
- AMBAC 231
- American call option 116
- American put option 149, 154–7
- AMEX 118
- Aquila 197
- arbitrage 6, 119, 277
  - arbitrage-based strategies 121–2
  - arbitrage pricing theory (APT) 64
- ARCH 7, 80, 278, 280, 282
- Arrow–Debreu paradigm 277
- Arrow–Pratt index of risk aversion 45, 47
- Artificial Intelligence 15
- Asian financial crisis 70
- autocovariance function 91
  
- Baltic Exchange (London) 116
- Bank of Canada 253
- Bankers Trust 199–200
- barrier option 164–5
- basis risk 115
- Basle Committee 310–11, 313
- Bayes criterion 26–7
- Bayes decision making 22–6
- Bayes theory 7
- Bayesian rationality 22
- bear spread 162, 179
- Bermudan option 118
- Bernoulli, Daniel 51–2
  
- beta factor 60, 61, 63
- binomial random walk 96
- binomial tree, multi-period 140–1
- black sheep syndrome 21
- Black–Scholes option formula 6, 147–57, 168, 196
  - volatility and 281
- bond(s)
  - arbitrageurs 214
  - convertible 114, 261–2
  - corporate 217
  - coupon-bearing 215–17
  - default 216, 224–30
  - definition 211
  - fixed rate 211
  - floating rate 211
  - forward rate and 222–4
  - junk 216
  - markets 212
  - mathematics 218
  - options on 260–4, 268–9
  - ratings 231
  - risk 161
  - strips 211
  - valuation, interest-rate processes, yields and 251–9
    - see also* rated bonds
- bonus-malus 8
- bounded rationality 4, 19, 20–2
- bounds, option 152–3
- Brownian motion 16, 80, 88, 92, 99–100
- bull spread 162, 178–9
- butterfly effect 80
- butterfly spread 162, 181
  
- calendar spread 162
- call–call spread strategy 178
- call option 116, 117, 133–4, 149, 169, 172–4

- call option price 135
- calls on calls 164
- calls on forwards 164
- calls on puts 164
- cap 117, 163, 262
- capital adequacy ratio (CAR) 310
- capital adequacy requirements (CAR) 12, 13, 73
- Capital Asset Pricing Model (CAPM) 54, 56, 187
  - capital markets and 63–4
  - investment and 59–61
  - risk-premium and 295
- capital market equilibrium 56
- caplet 262
- caption 163, 264
- cash or nothing 163
- CBOE 117
- Central Limit Theorem 16
- certain cash equivalent 57
- certainty equivalence 43–4, 45
- certification 271
- CFTC 126, 127
- chaos 79–80
- chaotic analysis 272
- cherry-picking 71
- Chicago Board of Trade (CBOT) 114, 117
- Chicago Mercantile Exchange (CME) 13, 114, 117, 126, 197
- chooser option 164
- claims 8
- classical discounting 65
- classical rationality 21
- climatic option 163
- climatic risks 196–7
- collar 262
- collateral call 124
- Commodity Exchange Act 125
- Commodity Futures Trading Commission 170
- commodity option 163, 169
- commodity pool 125
- commodity pool operators (CPOs) 125
- competition risk 11
- complete market 119, 146
- complexity analysis 272
- compound option 162, 165–70
- Computational Tool 15
- condor 162
- condor spread strategies 81
- consumption, individual investment and 57–9
- contingent claim assets 114
- contract 114
- control 24
- convertible bond 114, 261–2
- convertible preferred stock 114
- convex dominance 272
- convexity 218–22
- copula distributions 315, 318–19
- corporate bond 217
- correlation analysis 122
- correlation coefficient 91, 92
- correlation function 91
- correlation risk 197
- corridor (range note) 163, 262
- counter strategies of hedge funds 124
- coupon-bearing bond 215–17
- coverage 225
- covered call 162, 176–7
- Cox–Ingersoll–Ross (CIR) model 257
- credit ratings, downgrading 231–2
- credit risk 12–13
- crowd psychology 21
- cubic utility function 47
- currency option 163
- currency risk 12
- Daiwa Bank 13
- data mining 3, 15
- dead value 141
- death spiral 200
- decision criteria 26–31
- Decision Making Under Uncertainty (DMUU) 4, 7
- decision table 31–3
- default bond 216
  - risky debt and 224–30
- default, definition 227–8
- default rate 225
- degree of absolute prudence 45
- delivery price of stock 141
- Delta 151, 174, 182, 199
- dependencies 8, 9
- derivative asset 114
- derivative contract 114
- derivative security 6
- derivatives losses 198
- destabilizing strategies of hedge funds 124
- deterministic analysis 34

- Deutsche Bank 202
- Deutsche Terminbrosse (DTB) 118
- diffusion–stochastic differential equation
  - models 96
- diffusion–volatility 89
- disappointment 326–7
- discrete time approximations 100–3
- discrimination information 321
- divergence–distance measures 322
- diversification, principle of 62
- Donsker Theorem 99–100
- downside risk 309
- duration 218–22, 309
- dynamic hedging 124
- Dynegy Inc. 231
  
- efficient markets hypothesis 89
- endowment effect 21
- Enron 69, 197, 232
- entropy, maximum, principle of 320–4
  - discrimination and divergence 320–3
  - distribution with specified marginals 323–4
  - price process 323
- environmental litigation 67
- environmental protection 25
- environmental risk 9–10, 196–7
- EOE (European Options Exchange, Amsterdam) 118
- equally likely (Laplace) criteria 34
- equilibrium 293–5
- error normality 88
- Euler equation 65–7
  - equilibrium and 293–5
- Eurodollar deposits 162
- European call option 116, 148, 149
- European Monetary System 120
- event-related funds 123
- ex-ante risk management 23
- excess discount rate 56
- exchange option 163, 164
- exercise date 116
- exercise value 141
- expectation criterion 43
- expected cost criterion 32
- expected monetary cost 32
- Expected Monetary Value (EMV) 26–7, 33–7, 39, 40
- expected opportunity loss (EOL) 33–7
- expected payoff criterion 32
  
- Expected Profit under Perfect Information (EPPi) 33–7
- expected utility 40, 41, 45
  - bounds 45–6
  - critiques of 51–3
  - finance and 53–67
  - insurance, risk management and 48–50
- expected value (Bayes) criterion 26–7
- expected value of sample information (EVSI) 33–7
- exponential utility function 46–7
- ex-post risk management 23
  
- fair competition 67
- fair game 81–2
- ‘fallen angels’ 225
- fat tails 81, 89, 272
- FGIC 231
- filter rule 191
- financial instruments 113–19
- financial physics 15–16
- finite difference model 103
- firm liquidity 8
- First Boston 200
- first passage time 207–9
- Fitch Investor Service 12, 217
- fixed equivalence rate of return (FIE) 55
- flexibility 29
- flipping 70
- floating rate bond 211
- floor 117, 163, 262
- floorlet 262
- floortion 264
- Fokker–Planck equation 88
- forcing contract 75
- forecasts, optimal, rational expectations and 146–7
- foreign exchange (FX) trading 195
  - price of a European call option on 196–7
- foreign exchange risk 10–11
- forward contract 114–16, 141–5
- forward rate agreement (FRA) 115–16
- forward rate, bonds and 222–4
- forward starts 162
- fractional Brownian motion 272, 301
- franchises 75
- Frechet distribution 312, 317
- Freddie Mac 203
- front-running 70
- FSA 231

- fundamental factor model 64
- fundamental theorem of asset pricing 83
- fundamental theory, definition 3
- funding risk 201
- futures
  - contract 114–16, 141–5
  - market 115
  - occurrences, example 145
  - price 143–4
- gambler's ruin problem 191
- gambling 81–2
- Gambling Act (1774) 8
- Gamma 151, 182
- GARCH 7, 80, 278, 280, 282
- generalized method of moments (stochastic discount factor; SDF) 54, 65–7, 113, 274
- genetic testing, insurance and 71
- Gibson Greeting 199–200
- Girsanov Theorem 104–8
- Granite Partners 203
- 'Greeks' 151–2, 174
  - dynamic strategies and 181–4
- Gumbel distribution 317
- HARA (hyperbolic absolute risk aversion) 47
- hazard rate dominance 272
- hedge fund fees 127
- hedge funds 120–5
- hedging 6, 119, 120–3
- hedging portfolio 6
- herd effect index 301
- herding 21, 124–5
- holding cost 139
- homogeneity of reserves 310
- Hong Kong Futures Exchange 118
- Hurst exponent 300, 301
- IBM 198
- immediacy, financial instruments and 5
- immunization 221–2
- in the money (ITM) 151
- incentive compatibility constraint 75
- incompleteness in pricing 276, 277–8
- independent increments 92
- individual rationality constraint 75
- inflation 14–15
- inflation risk 14, 15
- information asymmetry 67–75, 113
- insider trading 113
- insufficient reason (Laplace), principle of 27, 321
- insurance 24
  - actuarial science and 7–10
  - contract 8
- interest rate modelling 251–3
- interest rate swap 262–4
- interest rates and term structure 259, 267–8
- internal rate of return (IRR) 55
- International Monetary Fund (IMF) 70
- invariance by drift 310
- investor protection rules 125–7
- irrational behaviour 39
- Ito's Lemma 93–4, 147, 148, 170, 191, 251, 252
  - applications of 94–108
  - application to wealth process of portfolio of stocks 95–6
- Jensen's inequality 45, 46
- Joseph effect 300
- junk bonds 216
- Junk–Treasury Spread (JTS) 224, 225
- knock-out option 162
- Kolmogorov 79, 84
- Kolmogorov (Fokker–Plank) equation 208
- Laplace (equally likely) criteria 34
- Laplace principle of insufficient reason 27, 321
- Laplace transform 208
- laps and flex 162
- law of the single price 24, 83, 84, 134, 146
- least divergence, principle of 288
- lemon phenomenon 68–9
- leptokurtic distributions 81, 88, 89
- leverage, hedge fund and 127
- linear growth of uncertainty 272
- linear regression 7
- linear risk-sharing rule 74
- Lipschitz conditions 90, 91
- liquidity 5
- liquidity risk 14
- logarithmic utility function 47
- London International Financial Futures Exchange (LIFFE) 118, 197
- London's Baltic Exchange 116

- long call option 172
- long put option 174
- Long-Term Capital Management (LTCM) 120
- lookback option 163
- loss prevention 24–5
- loss rate 225
- lotteries, utility functions and 40–2
- lower bounds of option 152
- Lyapunov stability analysis 272
  
- macro funds 121
- macro hedged funds 121
- macro specific fund strategies 122
- macroeconomic approach 64
- man-made risks 8, 9
- margin calls 124
- Marine Stewardship Council (MSC) 271
- market efficiency 112–13, 184
  - market memory and 274
- market fluctuations risk 201
- market gladiators 161, 162
- market hedge funds 123
- market inefficiencies 25, 184–5
- market integrity protection rules of hedge funds 126
- market rationality 123
- market risk 14
- Markov chain 96, 224, 230
- Markov stochastic process 89
- martingale 82–3
  - Girsanov Theorem and 104–8
  - market efficiency and 112
  - Wiener process as 106–7
- MATHEMATICA 152
- MATIF (Marché à Terme International de France) 118
- MATLAB 152
- maturity 141
- maximax (minimin) criterion 28, 34
- maximin (minimax) criterion 28, 34
- MBIA 231
- mean variance replication hedging 288–91
- mean reversion model 252
- memory, volatility and 273–4
- Merrill Lynch 68, 200
- Merton model 170, 251, 297–9
- Metalgesellschaft (MG) 201–2
  - MG Refining & Marketing (MGRM) 201–2
- minimax (maximin) criterion 28
- minimax regret 28–31, 334
- minimin (maximax) criterion 28
- min-max objectives 26
- monotonicity 310
- Monte Carlo simulation 163
  - VaR and 324
- Moody's 12, 212, 217, 224, 225, 227
  - bond ratings 231, 232
- moral hazard 25, 69–71
- multifractal time series analysis 272
- multiple-factor model 64
- mutual funds regulation 127
  
- net present value (NPV) 55–6, 217–18
- network and information asymmetries 277
- neural networks 3
- New York Federal Reserve Bank 200
- New York Mercantile Exchange (NYMEX) 201
- New York Stock Exchange (NYSE) 70, 118
- Nikkei Index 202–3
- Nippon Investor Service 12
- no arbitrage assumption 134, 135, 241, 242
- nonlinear (non-Gaussian) distributions 81
- non-Gaussian distributions 81
- non-uniqueness in pricing 276
- Normal distribution 88, 89–90, 312, 322–3
- Normal–Wiener processes 89
- numeraire 137
  
- obligation to sell 116
- offshore funds 162
- one-factor models (market premium in the CAPM) 63–4
- opportunity loss 3
- opportunity loss table 32–3, 35
- option misses 197–204
- option writers 161
- option(s) 13, 114, 116–19
  - asset or nothing 163
  - as you like it 165
  - Asian 118, 162, 164
  - barrier 164–5
  - Bermudan 118
  - on bonds 260–4, 268–9
  - bounds 152–3
  - call 116, 117, 133–4, 148, 149, 169, 172–4
  - chooser 164
  - climatic 163

- option(s) (*Cont.*)  
 commodity 163, 169  
 on corporate bonds 162  
 currency 163  
 on dividend-paying stocks 169  
 exchange 163, 164  
 on foreign currencies 169  
 on futures 163, 170  
 on a futures price 170  
 jump processes and 297–9  
 knockout 162  
 long call 172  
 lookback 163  
 on options 164  
 over-the-counter 118  
 packaged 162, 163–5  
 passport 165  
 path-dependent 162  
 practice and 171–84  
 rational expectations 135–7  
 stock 162, 165–70  
 upper bounds of 153  
 valuation 131–41  
 value at risk,, by Delta–Gamma 317–18  
*see also* put options
- Orange County (California) case 199, 200–1
- Ornstein–Uhlenbeck process 97, 254, 297  
 application of Ito’s Lemma to 94–5
- Osaka Stock Exchange (OSE) 203
- out of the money (OTM) 21, 151
- over-insurance 25
- over-the-counter (OTC) derivatives 114
- over-the-counter option 118
- packaged option 162, 163–5
- Pareto–Levy distributions 88, 89
- Pareto optimal risk sharing 74
- participation constraint 75
- passport option 165
- path-dependent options 162
- payback period 54
- payments based in effort 75
- payoff table 29, 31
- Philadelphia Exchange 195, 196
- PHLX 118
- plain vanilla strategies 172–5
- planning experiments 6–7
- platokurtic distributions 88
- Poisson Jump 92
- positive feedback trading 124
- power utility function 47
- predictability of markets 274
- preferred probability distribution 53
- premium 116
- premium payments 8, 48–50
- premium principles 8
- prepayment risk 122
- present value 12
- price risk 197
- PriceWaterhouseCoopers 203
- pricing and classical contract theory 277
- pricing insurance 8
- principal–agent problem 73–5
- private placements 125
- probabilistic analysis 34–7
- Procter & Gamble 199–200
- production-efficiency function 25
- proper hedge risk 201
- protective put 124, 162, 177
- prudence 44–5
- psychology and rationality 277–8
- put option 116, 117, 169, 172–4, 175  
 American 149, 154–7  
 buying 177  
 calls on 164  
 long 174  
 protective 124, 162, 177
- put–call parity 153, 165
- puts on calls 164
- quadratic utility function 47
- quantile risk *see* value at risk (VaR)
- quantity risk 197
- quantos 163
- R/S index 300, 301
- random cash flows 55–6
- random walk 84–91, 272, 282
- randomness, definition 79
- range 309
- range note 163, 262
- range to scale (R/S) analysis 272, 273
- rare events 8–9
- rated bond 212, 216  
 default and 230–51  
 duration 235–9  
 Markov chain and 233–5  
 pricing for risk-free rates 239–44  
 no-arbitrage condition 241–2  
 valuation of two-rates rated bond 242–4

- valuation of default-prone rated bonds
  - 244–51
  - cash value of rated firm 250–1
  - two-rated default bond 248–9
- rational expectations 21–2, 83–4, 111–13, 131
  - optimal forecasts and 146–7
- rationality, principles of 4, 20–2
- recovery ratio 226
- reduced-form models of default 226, 227–8
  - cf structural models 228–30
- reflexivity 84
- regret (Savage regret) 28–31
  - ex-post optimization and 26
  - valuation of firms and 30
  - VaR and 326–7
- regret–disappointments models 326
- reinsurance 49
- remembering 274
- repos 211
- reverse mortgage 215
- reward 40
- rho 151
- right to exercise 116
- risk aversion 43–5
- risk behaviour, utility and 42–8
- risk, definition 5–6, 23, 309
- risk diversification 24
- risk loving 43
- risk management 22, 23–6
- Risk Metrics 318
- risk-neutral pricing 22, 53, 112, 131, 137–9, 166, 184, 185, 186, 188
- risk-neutral probabilities 6, 119, 136, 137, 145–7, 167
- risk-neutral world 142–3
- risk premium 24, 43–4, 45
- risk sharing 8, 47–8, 49
- risk tolerance 42
- risk transfer 49
- robustness 23, 25–6, 44–5
  
- Savage regret criterion *see* regret
- scenario analysis 31–3, 122
- scenario optimization 26, 62
- screening 72–3, 277
- SEC 126
- sector funds 123
- sector risk 14
  
- Security Exchange Act (SEA) 126
- self-selection constraints 72–3
- semi-variance 309
- sensitivity equations 151–2
- September 11th, 2001 9
- sequentiality 274
- Shannon's entropy 320, 321
- shape risk 14
- short selling 123, 174
- signalling –3
- SIMEX (Singapore International)
  - 118, 203
- smile 6, 151
- smoothing splines 253
- Société Générale 313
- Southern Energy 197
- spatial diversification 5
- speculation 119
- Spence, M. 72
- spillover events 9
- spread strategies 178–9
- square root process 257
- St Petersburg Paradox 51–2
- stabilizing strategies of hedge funds
  - 124
- stacked rollover hedge 201
- Standard & Poors 12, 212, 217, 224, 227, 231
- state discretization 96–9
- stationarity 91–2
- statistical approach, APT and 64
- statistical data analysis 6
- stochastic calculus 92–4
- stochastic differential equations (SDEs) 88, 89, 90, 96
- stochastic discount factor (SDF)
  - (generalized method of moments) 54, 65–7, 113, 274, 276, 293
- stochastic dominance 272
- stochastic ordering 272
- stochastic processes, properties of 91–2
- stochastic volatility binomial models
  - 282–93
- stochastic volatility interest-rate models
  - 258–9
- stochastic volatility models 89, 282–93
- stochastic volatility, process discretization
  - 280–1
- stock, buying 177
- stock option 162, 165–70

- stopping time sell and buy strategies 184–95, 226
  - buying/selling on a random walk 189–90
  - pricing a buy/sell strategy on a random walk 193–5
- straddle 179, 180
- strangle 179, 180
- strap 180–1
- stress test 122
- strict stationarity 91
- strike price 116
- strip 180, 181
- strips bonds (Separate Trading of Registered Interest and Principal of Securities) 211
- structural models of default 226–7
  - cf reduced form models 228–30
- sub-additivity of reserves 310
- subjective probability distribution 111
- sub-martingale 82
- sunk costs 20
- super-martingale 82
- supershare 163
- sure-thing principle 52
- Svensson procedures 253
- swaps 117, 163
  - contracts 196
  - interest rate 262–4
- swaption 163, 264
- switch 163
- systematic risk 199
- systemic risk 125
  - reduction rules of hedge funds 126–7
- Taylor series approximation 100, 101
- Taylor series expansion 93, 151
- Tchebycheff bounds on expected utility 45
- technical analysis 273
- technological innovation 25
- term structure 56
  - interest rates and 267–8
- theta 151, 182
- Thomson Bank Watch 12
- time contraction 284–6
- time (process) discretization of
  - continuous-time finance models 96–103
- time phasing of events 5
- Tokyo International Financial Futures Exchange (TIFFE) 118
- too big to fail syndrome 70
- Toronto Stock Exchange (TSE) 118
- trade on volatility 151
- trading strategy 131
- traditional valuation 54–6
- tragedy of the commons 21
- transaction risk 11
- transition bond rating matrix 244
- transitivity axiom 42
- transparency 70
- Treasury bills 212
- trinomial walk 96, 191–3
- truth-in-lending 67
- two-fund separation theorem 63
- upper bounds of option 153
- utility
  - concept of 19, 39–42
  - lotteries and 40–2
  - maximization 61–3
  - risk behaviour and 42–8
  - see also* expected utility
- value at risk (VaR) (quantile risk) 23, 26, 46, 185, 199, 309, 310–11
  - analytic variance–covariance approach 315–16
  - in banking practice 313–14
  - Burr 317
  - copulae and portfolio measurement 318–19
  - definition and applications 311–15
  - efficiency 324–7
  - extreme statistics and 316–17
  - Frechet distribution and 317
  - Gumbel distribution and 317
  - historical approach 315
  - Monte Carlo simulation and 324
  - negentropy 320
  - numerical examples 314–15
  - option, by Delta–Gamma 317–18
  - portfolio risk efficiency with normal returns and 324–5
  - regret and 326–7
  - statistics 315–24
  - Weibull distribution and 317
- value investing funds 123
- Vasicek interest-rate model 254–8
  - nonstationary 258

- Vega 151, 174  
Vivendi Universal SA 231  
volatility 6, 88  
    definition 271–3  
    equilibrium and 275–8  
    implicit 281–2  
    incomplete markets and 276–8  
    memory and 273–4  
    modelling and estimation 272–3  
    process variance and 278–81  
    range process and 299–301  
    risk 14  
    smile 281–2
- Wald's identity 190  
Walras, L. 212
- warrants 114, 162, 163, 168–9  
wealth, definition 8, 131  
wealth process of portfolio of stocks  
    application of Ito's Lemma to 95–6  
Weibull distribution 312, 317  
Wiener (Levy) process 88, 90, 92, 95–6  
    as a martingale 106–7  
Wiener risk 298  
'wild money' 13  
Willasen inequalities 46  
World Bank 162  
WorldCom Inc. 231
- yield curve estimation 24  
zero-coupon bonds 211  
zero-coupon, default-free bond 213–14

*Index compiled by Annette Musker*