Measuring Market Risk

Second Edition

Kevin Dowd

John Wiley & Sons, Ltd

Measuring Market Risk

For other titles in the Wiley Finance Series
please see www.wiley.com/finance

Measuring Market Risk

Second Edition

Kevin Dowd

John Wiley & Sons, Ltd

 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
Visit our Home Page on www.wiley.com

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

Dowd, Kevin.
Measuring market risk / Kevin Dowd.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 13 978-0-470-01303-8 (cloth : alk. paper)
ISBN 10 0-470-01303-6 (cloth : alk. paper)
1. Financial futures—Mathematical models. 2. Risk management—Mathematical models.
3. Portfolio management—Mathematical models. I. Title.
HG6024.3.D683 2005
332.63′2042—dc22 2005010796

British Library Cataloguing in Publication Data

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

ISBN 13 978-0-470-01303-8 (HB)
ISBN 10 0-470-01303-6 (HB)

Typeset in 10/12pt Times by TechBooks, New Delhi, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
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.

Contents

Preface to the Second Edition

> You are responsible for managing your company's foreign exchange positions. Your boss, or your boss's boss, has been reading about derivatives losses suffered by other companies, and wants to know if the same thing could happen to his company. That is, he wants to know just how much market risk the company is taking. What do you say?
>
> You could start by listing and describing the company's positions, but this isn't likely to be helpful unless there are only a handful. Even then, it helps only if your superiors understand all of the positions and instruments, and the risks inherent in each. Or you could talk about the portfolio's sensitivities, i.e., how much the value of the portfolio changes when various underlying market rates or prices change, or perhaps option deltas and gammas. However, you are unlikely to win favor with your superiors by putting them to sleep. Even if you are confident in your ability to explain these in English, you still have no natural way to net the risk of your short position in Deutsche marks against the long position in Dutch guilders. . . . You could simply assure your superiors that you never speculate but rather use derivatives only to hedge, but they understand that this statement is vacuous. They know that the word 'hedge' is so ill-defined and flexible that virtually any transaction can be characterized as a hedge. So what do you say?[1]

The obvious answer, 'The most we can lose is . . . ' is also clearly unsatisfactory, because the most we can possibly lose is everything, and we would hope that the board already knows that. Consequently, Linsmeier and Pearson continue, 'Perhaps the best answer starts: "The value at risk is . . . ".'

So what is value at risk? Value at risk (VaR) is the maximum likely loss over some target period – the most we expect to lose over that period, at a specified probability level. It says that on 95 days out of 100, say, the most we can expect to lose is $10 million or whatever. The board or other recipients specify their probability level – 95%, 99% and so on – and the risk manager can tell them the maximum they can lose at that probability level. The recipients can also specify the horizon period – the next day, the next week, month, quarter, etc. – and again the risk manager can tell them the maximum amount they stand to lose over that horizon period. Indeed, the recipients can specify any combination of probability and horizon period, and the risk manager can give them the VaR applicable to that probability and horizon period.

We then have to face the problem of how to estimate the VaR. This is a tricky question, and the answer is very involved and takes up much of this book. The short answer is, therefore, to read this book or others like it.

However, before we get too involved with VaR, we also have to face another issue. Is a VaR measure the best we can do? The answer is no. There are alternatives to VaR and some

[1] Linsmeier and Pearson (1996, p. 1).

of these – especially the coherent risk measures – are demonstrably superior. Consequently, I would take issue with Linsmeier and Pearson's answer. 'The VaR is . . . ' is sometimes a reasonable answer, but it is often not the best one and it can sometimes be a very bad one. Risk managers who use VaR as their preferred risk measure should really be using coherent risk measures instead: VaR is already passé.

But if coherent risk measures are superior to VaR, why bother to estimate VaR? This is a good question, and also a controversial one. Part of the answer is that there will be a need to estimate VaR for as long as there is a demand for VaR itself: if someone wants the number, then someone will want to estimate it, and whether anyone should want the number in the first place is another matter. In this respect VaR is a lot like the infamous beta. People still want beta numbers, regardless of the well-documented problems of the Capital Asset Pricing Model on whose validity the beta risk measure depends. A purist might say they shouldn't, but the fact is that they do. So the business of estimating betas goes on, even though the CAPM is now widely discredited. The same goes for VaR: a purist would say that VaR is an inferior risk measure, but people still want VaR numbers and so the business of VaR estimation goes on regardless. A second and better reason to estimate the VaR is that the VaR is a quantile (i.e., a quantity associated with a particular cumulative probability), and there are sometimes good reasons to estimate quantiles. For example, we might want to estimate quantiles when dealing with what insurers call 'probability of ruin' problems, where we are interested in a threshold that will be exceeded with a certain probability. Such problems occur very commonly, most particularly when it comes to the determination of reserves or capital requirements. However, there is also a third and more general reason to estimate VaRs: being able to estimate a VaR (or a quantile) is the key to the estimation of better risk measures, since the coherent and other risk measures are essentially weighted averages of quantiles. So we need to be able to estimate quantiles, even if we don't wish to use VaR as our preferred risk measure.

INTENDED READERSHIP

This book provides an overview of the state of the art in market risk measurement. The measures covered include the VaR, but also include coherent risk measures as well. Given the size and rate of growth of this literature, it is impossible to cover the field comprehensively, and no book in this area can credibly claim to do so, even one like this that focuses on risk measurement and does not try to grapple with the much broader field of market risk management. Within the subfield of market risk measurement, the coverage of the literature provided here can claim to be no more than reasonably extensive.

The book is aimed at three main audiences. The first consists of practitioners in risk measurement and management – those who are developing or already using VaR and related risk systems. The second audience consists of students in MBA, MA, MSc and professional programmes in finance, financial engineering, risk management and related subjects, for whom the book can be used as a reference or textbook. The third audience consists of PhD students and academics working on risk measurement issues in their research. Inevitably, the level at which the material is pitched must vary considerably, from basic to advanced. Beginners will therefore find some of it heavy going, although they should get something out of it by skipping over difficult parts and trying to get an overall feel for the material. For their part, advanced readers will find a lot of familiar material, but even many of them should, I hope, find some material here to interest them.

To get the most out of the book requires a basic knowledge of computing and spreadsheets, statistics (including some familiarity with moments and density/distribution functions), mathematics (including basic matrix algebra), and some prior knowledge of finance, most especially derivatives and fixed-income theory. Most practitioners and academics should therefore have relatively little difficulty with it, but for students this material is best taught after they have already done their quantitative methods, derivatives, fixed-income and other 'building block' courses.

USING THIS BOOK

In teaching market risk material over the last few years, it has also become very clear to me that one cannot teach this material effectively – and students cannot really absorb it – if one teaches only at an abstract level. Of course, it is important to have lectures to convey the conceptual material, but risk measurement is not a purely abstract subject, and in my experience students only really grasp the material when they start playing with it – when they start working out VaR figures for themselves on a spreadsheet, when they have exercises and assignments to do, and so on. When teaching, it is therefore important to balance lecture-style delivery with practical sessions in which the students use computers to solve illustrative risk measurement problems.[2]

If the book is to be read and used practically, readers also need to use appropriate spreadsheet or other software to carry out estimations for themselves. Again, my teaching and supervision experience is that the use of software is critical in learning this material, and we can only ever claim to understand something when we have actually calculated it ourselves. The calculation and risk material are intimately related, and the good risk practitioner knows that the estimation of risk measures always boils down to some spreadsheet or other computer function. In fact, much of the action in this area boils down to software issues – comparing alternative software routines, finding errors, improving accuracy and speed, and so on. A book on risk measurement should therefore come with some indication of how risk measurement routines can be implemented on a computer.

It is better still for such books to come with their own software, and this book comes with a CD that contains a selection of risk measurement and related functions in MATLAB (and some Excel ones too) and a manual explaining their use.[3] My advice to users is to print out the manual and go through the functions on a computer, and then keep the manual to hand for later reference.[4] The examples and figures in the book are produced using this software, and readers should be able to reproduce them for themselves. Readers are very welcome to contact me with any feedback. I will keep the Toolbox and the manual up to date on my website (www.nottingham.ac.uk/∼lizkd), and readers are welcome to download updates from there.

In writing this software, I should explain that I chose MATLAB mainly because it is both powerful and user-friendly, unlike its obvious alternatives (VBA, which is neither powerful

[2] For those who wish to use this book for teaching, I also have a complete set of Powerpoint slides, which I am happy to make available on request.

[3] MATLAB is a registered trademark of The MathWorks, Inc. For more information on MATLAB, please visit their website, www.mathworks.com.

[4] The user should copy the Managing Market Risk folder into his or her MATLAB works folder and activate the path to the Managing Market Risk folder thus created (so MATLAB knows the folder is there). The functions were written in MATLAB 6.0 and most of the MMR functions should work if the user has the Statistics Toolbox as well as the basic MATLAB 6.0 or later software installed on their machine. However, a small number of MMR functions draw on functions in other MATLAB toolboxes (such as the Garch Toolbox), so users with only the Statistics Toolbox will find that the occasional MMR function does not work on their machine.

nor particularly user-friendly, or the C or S languages, which are not so user-friendly). I also chose MATLAB in part because it produces very nice graphics, and a good graph or chart is often an essential tool for risk measurement. Unfortunately, the downside of MATLAB is that many users of the book will not be familiar with it or will not have ready access to it, and I can only advise such readers to think seriously about going through the expense and/or effort to get it.[5]

In explaining risk measurement throughout this book, I have tried to focus on the underlying ideas rather on programming code: understanding the ideas is much more important, and the coding itself is mere implementation. My advice to risk measurers is that they should aim to get to the level where they can easily write their own code once they know what they are trying to do. However, for those who want it, the code I use is easily accessible – one simply opens up MATLAB, goes into the Measuring Market Risk (MMR) Toolbox, and opens the relevant function. The reader who wants the code should therefore refer directly to the program coding rather than search around in the text: I have tried to keep the text itself free of such detail to focus on more important conceptual issues.

The MMR Toolbox also has many other functions besides those used to produce the examples or figures in the text. In fact, I have tried to produce a fairly extensive set of software functions that would cover all the obvious estimation measurement problems, as well as some of the more advanced ones. Users – such as students doing their dissertations, academics doing their research, and practitioners working on practical applications – might find some of these functions useful, and they are welcome to make whatever use of these functions they wish. However, they should recognise that I am not a programmer and anyone who uses these functions must do so at his or her own risk. As always in risk management, we should keep our wits about us and not be too trusting of the software we use or the results we get.

OUTLINE OF THE BOOK

The first chapter provides a brief overview of recent developments in risk measurement, and focuses particularly on the remarkable rise to prominence of the VaR in the 1990s. This puts VaR into context, and also explains the attractions that made it so popular. Chapter 2 then looks at three different risk measurement frameworks, based respectively on portfolio theory, VaR and coherent risk measures. This chapter is in many ways the key chapter in the book, and sets out in considerable detail what is wrong with the VaR and why coherent risk measures are superior to it.

Having established what our basic risk measures actually are, Chapter 3 provides an introduction to and overview of the main issues involved in estimating them. Later chapters then fill in some of the detail:

- Chapter 4 discusses the non-parametric approaches, in which we seek to estimate measures of market risk while making minimal assumptions about the distribution of losses or returns.
- Chapter 5 looks at the forecasting of volatilities, covariances and correlations, which are a preliminary to the parametric approaches that follow.

[5] When I first started working on this book, I initially tried writing the software functions in VBA to take advantage of the fact that almost everyone has access to Excel; unfortunately, I ran into too many problems and eventually had to give up. Had I not done so, I would still be struggling with VBA code even now, and this book would never have seen the light of day. So, while I sympathise with those who might feel pressured to learn MATLAB or some other advanced language and obtain the relevant software, I don't see any practical alternative: if you want software, Excel/VBA is just not up to the job – although it can be useful for many simpler tasks and for teaching at a basic level.

- Chapters 6 and 7 discuss the parametric approaches, which estimate risk measures based on assumptions about loss or return distributions. Chapter 6 looks at general parametric approaches and Chapter 7 looks at extreme-value (EV) approaches.
- Chapters 8 and 9 discuss Monte Carlo simulation (or 'random number') methods, with Chapter 8 providing an introduction to these methods in general, and Chapter 9 examining some of the many ways in which these methods can be used to estimate market risk measures. These are immensely powerful methods that can handle a very large range of problems, including very complicated ones.
- Chapter 10 examines the difficult but important subject of how to estimate risk measures for options positions.
- Chapter 11 discusses risk decomposition: how to 'break down' aggregate risk measures and allocate risk to individual positions in our portfolio.

The remaining chapters look at various other important topics related to the estimation of market risk measures:

- Chapter 12 discusses the subject of mapping, where 'real' positions are 'mapped' to surrogate ones that are much easier to handle. We can also think of mapping as the process of describing our positions in terms of combinations of standard building blocks.
- Chapter 13 examines stress testing (or 'what if' analysis). Stress tests are important complements to probabilistic risk measures, and can also be regarded as bona fide risk measures in their own right.
- Chapter 14 discusses the multifaceted issue of liquidity risk: the nature of market liquidity, how to modify estimates of risk measures to allow for it, how to estimate liquidity at risk, and how to estimate crisis-related liquidity risks.
- Chapter 15 deals with backtesting – the application of quantitative methods to determine whether a model's risk estimates are consistent with the assumptions on which the model is based or to rank models against each other.
- Finally, Chapter 16 considers the important subject of model risk – the risk of error in our risk estimates due to inadequacies in our risk models.

REVISIONS TO THE SECOND EDITION

The second edition represents a very substantial revision to the first. There is some updating and I have made a major effort to take into account not only the mainstream financial risk literature, but also the actuarial/insurance literature, which has many useful contributions to offer. To make space for the new material, I have also cut out material on topics (e.g., on quasi-Monte Carlo and lattice methods) that have yet to make a major impact on the risk measurement literature: the new edition is meant to reflect the state of practice, and I have (in places, at least) tried to resist the temptation to put in material that is yet to be widely accepted by the risk management profession. The only real exception is in Chapter 15 on backtesting, but much of the new material in this chapter is based on my own recent research and has already been through considerable scrutiny through the journal refereeing process.

In terms of the argument 'pushed' by the book, I have to admit that I am very much persuaded by arguments made for the superiority of coherent risk measures over the VaR, and I am increasingly conscious of the limitations of the latter. In fact, I have pretty much persuaded myself by now that the VaR is close to useless as a 'proper' risk measure, although that is not to say that the VaR as such is useless, because it also has its uses as a quantile. However, I also

believe that there are major issues with coherent risk measures too: they are far from 'perfect' themselves, and there are many unanswered questions about them. At a deeper level, I also believe that there are major problems with the application of physical science models to social situations, and I remain extremely sceptical of the financial regulatory system, which I believe does more harm overall than good.

The material itself is radically reorganised in the light of feedback, further teaching experience and the like, and I have put in a large number of worked-out examples which show how many of the calculations can be carried out from scratch. There is therefore a much greater emphasis on worked-out examples and on explaining the principles and mechanics of the calculations.

Acknowledgements

It is a real pleasure to acknowledge all those who have contributed in one way or another to this book. To begin with, I should like to thank all those who contributed to the first edition, and the many people who were good enough to give me feedback on it. I should like to thank the UK Economic and Social Research Council whose financial support for a research fellowship on 'Risk Measurement in Financial Institutions' gave me the time and other resources to complete the second edition. I also thank Barry Schachter for his website, www.gloriamundi.org, which was my primary source of research material. I thank Naomi Fernandes, Courteney Esposito and the The MathWorks, Inc., for making MATLAB available to me through their authors' program, and I thank the Wiley team – especially Caitlin Cornish, Carole Millett, Sam Hartley and Sam Whittaker – for many helpful inputs. I thank Andrew Cairns, John Cotter, Tony Courakis, Jon Danielsson, Jim Finnegan, Kostas Giannopoulos, Chris Humphrey, Imad Moosa, and Dave Rowe for their valuable comments on parts of the draft manuscript and/or for other contributions. In addition, I would like to thank my colleagues at the Centre for Risk and Insurance Studies and in Nottingham University Business School generally for their support and feedback, especially Bob Berry, Chris O'Brien, Steve Diacon, Jennifer Howis, Kerry Lambert, Peter Oliver, and Tim Orton. I thank my BSRA colleagues – Chris Mammarelli, Geoff Ihle, and Tony Bimani – for their inputs, I also owe very special thanks to Carlos Blanco, Mark Billings, David Blake, Dave Campbell, Changguo Liu, Ian Gow, Ling Jin, Duncan Kitchin, Dave and Sheila Morris, Anneliese Osterspey, Dave and Frances Owen, Sheila Richardson, Stan and Dorothy Syznkaruk, Margaret Woods, and Basil and Margaret Zafiriou for their invaluable contributions and kindnesses. Finally, as always, my greatest debts are to my family – to my mother, Maureen, my brothers Brian and Victor, and most of all to my wife Mahjabeen and my daughters Raadhiyah and Safiah – for their love and unfailing support, and for their patience with all the time missed as I disappeared to work away on yet another book. I would therefore like to dedicate this book to Mahjabeen and the girls. But before they say anything: yes, I did promise the last time that my next book would be a novel, and not another riveting tome on statistics. However, on second thoughts, I hope they will agree with me that English literature has enough problems already.

1

The Rise of Value at Risk

We can think of financial risk as the risk associated with financial outcomes of one sort or another, but the term 'risk' itself is very difficult to pin down precisely. It evokes notions of uncertainty, randomness, and probability. The random outcomes to which it alludes might be good (e.g., we might win a lottery) or bad (e.g., we might suffer a financial loss), and we may (or may not) prefer to focus on the risks associated with 'bad' events, presumably with a view to trying to protect ourselves against them. There is also the question of quantifiability – some scholars distinguish between 'risk' as something quantifiable and 'uncertainty' as its non-quantifiable counterpart. The notion of 'risk' in its broadest sense therefore has many facets, and there is no single definition of risk that can be completely satisfactory in every situation. However, for our purposes here, a reasonable definition is to consider financial risk as the prospect of financial loss – or maybe gain – due to unforeseen or random changes in underlying risk factors.

In this book we are concerned with the measurement of one particular form of financial risk – namely, market risk, or the risk of loss (or gain) arising from unexpected changes in market prices (e.g., such as security prices) or market rates (e.g., such as interest or exchange rates). Market risks, in turn, can be classified into interest-rate risks, equity risks, exchange rate risks, commodity price risks, and so on, depending on whether the risk factor is an interest rate, a stock price, or another random variable. Market risks can also be distinguished from other forms of financial risk, particularly credit risk (or the risk of loss arising from the failure of a counterparty to make a promised payment) and operational risk (or the risk of loss arising from the failures of internal systems or the people who operate in them).

The theory and the practice of risk management have developed enormously since the pioneering work of Harry Markowitz in the 1950s. The theory has developed to the point where risk management is now regarded as a distinct subfield of the theory of finance, and one that is increasingly taught as a separate subject in the more advanced master's and MBA programmes in finance. The subject has attracted a huge amount of intellectual energy, not just from finance specialists but also from specialists in other disciplines who are attracted to it – as is illustrated by the large number of ivy league theoretical physics PhDs who now go into finance research, attracted not just by high salaries but also by the challenging intellectual problems it poses. The subject has benefited enormously from contributions made by quantitative disciplines such as statistics, mathematics, and computer science (and others, such as engineering and physics). However, the subject is not purely, or even mainly, a quantitative one. At the heart of the subject is the notion of good risk management practice, and above anything else this requires an awareness of the qualitative and organisational aspects of risk management: a good sense of judgement, an awareness of the 'things that can go wrong', an appreciation of market history, and so on. This also means some of the most important principles of risk management actually come from disciplines outside finance, most especially the disciplines of accounting (which tells us about subjects such as management control, valuation and audit), economics (which tells us about how markets behave and about welfare maximisation, among

other things), organisational theory (which tells us about how organisations behave), and law (which pervades almost everything in risk management). So, while risk management involves quantitative methods, the subject itself rests on a foundation that is qualitative. In many ways, the subject is much like engineering: it uses sophisticated tools, but context and judgement are everything. And this, perhaps, is the most important thing for any budding risk manager to appreciate – especially one from a quants background.

Box 1.1 Why Manage Corporate Financial Risks?

At one level, the benefits of risk management are obvious: we reduce the danger of harmful events occurring. However, this response does not fully explain why *firms* might practice financial risk management. Even if individual investors are risk averse and manage the investment portfolio risks, it still does not follow that firms should manage *their* overall corporate risks. If investors have access to perfect capital markets (with all the economic textbook fictions that that entails), they can achieve the degrees of diversification they want through their own actions, and corporate financial risk management would be irrelevant. This is the message of the famous Modigliani–Miller theorem, which says that in an ideal theoretical world with no informational asymmetries, principal–agent problems, taxes, transactions costs or bankruptcy costs, and with 'perfect' frictionless markets, the financial structure of the firm (and, by implication, any risk management) would be irrelevant. Hence, any *explanation* of the benefits of corporate financial risk management must start by identifying which of the Modigliani–Miller assumptions do *not* apply to the real world; relaxing the relevant assumption then enables us to see why firms might benefit from financial risk management. These benefits arise from the following:

- Risk management helps to increase the value of the firm in the presence of bankruptcy costs, because it makes bankruptcy less likely.
- The presence of informational asymmetries means that external finance is more costly than internal finance, and good investment opportunities can be lost. Risk management helps alleviate these problems by reducing the variability of the corporate cash flow.
- Risk management helps investors achieve a better allocation of risks, because the firm would typically have better access to capital markets.
- In the presence of taxes, risk management can help reduce the firm's tax bill, because the amount of tax paid is a convex function of its profits: this means that the less variable its profits, the lower its average tax bill.

1.1 THE EMERGENCE OF FINANCIAL RISK MANAGEMENT

The emergence of financial risk management as a discipline is due to a number of factors. One factor is the phenomenal growth in trading activity since the late 1960s, illustrated by the facts that the average number of shares traded per day in the New York Stock Exchange has grown from a little over $4 million in 1961 to around $1.6 trillion in early 2005, and that turnover in foreign exchange markets has grown from about a billion dollars a day in 1965 to $1.9 trillion in April 2004.[1] There have also been massive increases in the range of instruments traded over the past two or three decades, and trading volumes in these new instruments have also grown very

[1] The latter figure is from BIS (2004a), p. 1.

rapidly. New instruments have been developed in offshore markets and, more recently, in the newly emerging financial markets of eastern Europe, India, East Asia, Latin America, Russia, and elsewhere. New instruments have also arisen for assets that were previously illiquid, such as consumer loans, commercial and industrial bank loans, mortgages, mortgage-based securities, and similar assets, and these markets have grown very considerably since the early 1980s.

There has also been a huge growth of financial derivatives activity. Until 1972 the only derivatives traded were certain commodity futures and various forwards and some over-the-counter (OTC) options. The Chicago Mercantile Exchange then started trading foreign currency futures contracts in 1972, and in 1973 the Chicago Board Options Exchange started trading equity call options. Interest-rate futures were introduced in 1975, and a large number of other financial derivatives contracts were introduced in the following years: swaps and exotics (e.g., swaptions, futures on interest rate swaps, etc.) then took off in the 1980s, and catastrophe, credit, electricity and weather derivatives in the 1990s and mortality derivatives in the 2000s. From negligible amounts in the early 1970s, the total notional amounts held in outstanding OTC derivatives contracts grew to $220 trillion by the first half of 2004.[2]

This growth in trading activity has taken place against an environment that was often very volatile. A volatile environment exposes firms to greater levels of financial risk, and provides incentives for firms to find new and better ways of managing this risk. The volatility of the economic environment is reflected in various ways:

- *Stock market volatility*: Stock markets have always been volatile, but sometimes extremely so: for example, on October 19, 1987, the Dow Jones fell 23% and in the process knocked off over $1 trillion in equity capital; and from July 21 through August 31, 1998, the Dow Jones lost 18% of its value. Other western stock markets have experienced similar falls, and some Asian ones have experienced much worse ones (e.g., the South Korean stock market lost over half of its value over 1997).
- *Exchange rate volatility*: Exchange rates have been volatile ever since the breakdown of the Bretton Woods system of fixed exchange rates in the early 1970s. Occasional exchange rate crises have also led to sudden and significant exchange rate changes, including – among many others – the ERM devaluations of September 1992, the problems of the peso in 1994, the east Asian currency problems of 1997–98, the rouble crisis of 1998, Brazil in 1999 and Argentina in 2001.
- *Interest-rate volatility*: There have also been major fluctuations in interest rates, with their attendant effects on funding costs, corporate cash flows and asset values. For example, the Fed Funds rate, a good indicator of short-term market rates in the US, approximately doubled over 1994.
- *Commodity market volatility*: Commodity markets are notoriously volatile, and commodity prices often go through long periods of apparent stability and then suddenly jump by enormous amounts. Some commodity prices (e.g., electricity prices) also show extremely pronounced day-to-day and even hour-to-hour volatility.

The development of risk management has also been spurred on by concerns with the dangers of improper derivatives use, and by a sorry catalogue of risk management disasters since the early 1990s. These dangers were sounded loud and clear by E. Gerald Corrigan, the then

[2] BIS (2004b), p. 9. However, this figure is misleading, because notional values give relatively little indication of what derivatives contracts are really worth. The true size of derivatives trading is therefore better represented by the market value of outstanding derivatives contracts. The same survey estimated this to be $6.4 trillion – which is a little under 3% of the notional amount, but still an astronomical number in its own right.

President of the Federal Reserve Bank of New York, in an address to the New York Bankers Association in early 1992 in which he told them very bluntly:

> You had all better take a very, very hard look at off-balance sheet activities. The growth and complexity of [these] activities and the nature of the credit settlement risk they entail should give us cause for concern . . . . I hope this sounds like a warning, because it is.

Corrigan's concerns proved well grounded. Time and again in the next few years, apparently respectable institutions stunned the world by announcing massive losses, often through unauthorised trades, and always because of inadequate risk management: Showa Shell in 1993, Metallgesellschaft, Kashima Oil, Procter and Gamble and Orange County in 1994, Barings and Daiwa in 1995, NatWest in 1997, Yakult Honsha and LTCM in 1998, Enron in 2001, Allied Irish Bank and WorldCom in 2002, and Parmalat in 2003. Most of these disasters involved losses of over a billion US$, and this list ignores most of the smaller disasters. These disasters have made the financial community keenly aware of the importance of good risk management practice – as the saying goes, if you think risk management is expensive, just look at the alternative. The disasters of the early 1990s and associated concerns with the dangers of derivatives also encouraged a lot of soul searching in the financial risk community, and led to a series of reports which articulated best practice principles. The most widely cited of these was the G-30 Report (1993), but other, similar, reports were also issued by the US General Accounting Office (1994 and 1996), and the Derivatives Policy Group (1995), among others. They all preached much the same message – the need for good management controls including the need to separate front and back offices, the use of VaR models to estimate firmwide risks, the need for adequate systems to manage derivatives risks, the need to monitor counterparty relationships, the importance of good disclosure, and the like.

A fourth factor contributing to the development of risk management was the rapid advance in the state of information technology. Improvements in IT have made possible huge increases in both computational power and in the speed with which calculations can be carried out. Improvements in computing power mean that new techniques can be used (such as computer-intensive simulation techniques) to enable us to tackle more difficult calculation problems. Improvements in calculation speed then help make these techniques useful in real time, where it is often essential to get answers quickly. This technological progress has led to IT costs falling by about 25–30% a year over the past 40 years or so, and the costs of transmitting data have fallen even faster. Improvements in computing power, increases in computing speed, and reductions in computing costs have thus come together to transform the technology available for risk management. Managers are no longer tied down to the simple 'back of the envelope' techniques that they had to use earlier when they lacked the means to carry out complex calculations. Instead, they can now use sophisticated algorithms programmed into computers to carry out real-time calculations that were not possible before, and the ability to carry out such calculations in turn creates a whole new range of risk management possibilities – and problems. [3]

1.2 MARKET RISK MEASUREMENT

This book is concerned not so much with the broader field of financial risk management or even with the somewhat narrower field of financial risk measurement, but with the smaller

[3] The contributory factors discussed here are by no means exhaustive. Besides the factors mentioned in the text there are also the issues of globalisation and increasing organisational complexity, which have made firms more opaque and harder to manage, and in some ways more prone to fail.

(but still very extensive!) subfield of market risk measurement. However, the basic principles of market risk measurement apply with suitable extensions or modifications to other types of risk, so much of what the book discusses could be useful to risk managers concerned with other financial (or even non-financial) risks. Moreover, although we focus on measurement issues, risk measurement and risk management are intimately related, and we cannot address one without also saying a lot about the other. In any case, whatever the particular kind of risk we are dealing with, the first thing to know about measuring any risk is to appreciate the broader risk management context within which the measurement itself takes place: context is all important.

Having delineated our subject matter, it is useful to look at how the techniques of market risk measurement have developed over the years. We therefore turn now to consider the earlier – in retrospect, antediluvian – risk measurement tools that were used before risk managers had computers capable of carrying out more complex calculations.

1.3 RISK MEASUREMENT BEFORE VaR

1.3.1 Gap Analysis

One common approach is gap analysis, which was initially developed by financial institutions to give a crude idea of interest-rate risk exposure. Gap analysis starts with the choice of an appropriate horizon period (e.g., 1 year ahead). We then determine how much of our asset or liability portfolio will re-price within this period, and the amounts involved give us our rate-sensitive assets and rate-sensitive liabilities for the chosen horizon. The gap is the difference between these, and our interest-rate exposure is taken to be the change in net interest income that occurs in response to a change in interest rates. This is assumed to be equal to the gap times the interest-rate change:

$$\Delta NII = (GAP)\Delta r \tag{1.1}$$

where ΔNII is the change in net interest income and Δr is the change in interest rates.

Gap analysis is fairly simple to carry out, but has its limitations: it only applies to on-balance sheet interest-rate risk, and even then only crudely; it looks at the impact of interest rates on income, rather than on asset or liability values; and results can be sensitive to the choice of horizon period.

1.3.2 Duration Analysis

A second traditional method used by financial institutions for measuring interest-rate risks is duration analysis. The (Macaulay) duration D of a bond (or any other fixed-income security) can be defined as the weighted average term to maturity of the bond's cash flows, where the weights are the present value of each cash flow relative to the present value of all cash flows:

$$D = \frac{\sum_{i=1}^{n} i \times PVCF_i}{\sum_{i=1}^{n} PVCF_i} \tag{1.2}$$

where $PVCF_i$ is the present value of the period i cash flow, discounted at the appropriate spot period yield. The duration measure is useful because it gives an approximate indication of the

sensitivity of a bond price to a change in yield:

$$\%\ change\ in\ bond\ price \approx -\frac{D\Delta y}{(1+y)} \tag{1.3}$$

where y is the yield and Δy the change in yield. The bigger the duration, the more the bond price changes in response to a change in yield. The duration approach is very convenient because duration measures are easy to calculate and the duration of a bond portfolio is a simple weighted average of the durations of the individual bonds in the portfolio. It is also better than gap analysis insofar as it looks at changes in asset (or liability) values, rather than just changes in net income.

However, duration approaches also have similar limitations: they ignore risks other than interest-rate risk; they are crude,[4] and even with various refinements to improve accuracy,[5] duration-based approaches are still inaccurate relative to more sophisticated approaches to interest-rate term structure analysis (such as Heath–Jarrow–Morton) or market-based models (such as Brace–Gatarak–Musiela). Moreover, the main reason for using duration approaches in the past – their (comparative) ease of calculation – is no longer of much significance, since more sophisticated models can now be programmed into personal computers to give their users more accurate answers very rapidly.

1.3.3 Scenario Analysis

Another approach is scenario analysis (or 'what if' analysis), in which we set out different scenarios and investigate what we stand to gain or lose under them. To carry out scenario analysis, we select a set of scenarios – or paths describing how relevant variables, stock prices, interest rates, exchange rates, etc., might evolve over a horizon period. We then postulate the cash flows and/or accounting values of assets and liabilities as they would develop under each scenario, and use the results to come to a view about our exposure. Scenario analyses can be more or less sophisticated, and early scenario analyses were inevitably crude given the limited computing power available. However, they did at least give some idea of what firms stood to lose under specific circumstances.

Scenario analysis is not easy to carry out. A lot hinges on our ability to identify the 'right' scenarios, and there are relatively few rules to guide us when selecting them. We need to ensure that the scenarios are reasonable and do not involve contradictory or excessively implausible assumptions, and we need to think through the interrelationships between the variables involved. We also want to make sure, as best we can, that we have all the main scenarios covered. Scenario analysis also tells us nothing about the likelihood of different scenarios, so we need to use our judgement when assessing the practical significance of different scenarios. In the final analysis, the results of scenario analyses are highly subjective and depend to a very large extent on the skill (or otherwise) of the analyst.

[4] They are crude because they only take a first-order approximation to the change in the bond price, and because they implicitly presuppose that any changes in the yield curve are parallel ones (i.e., all yields across the maturity spectrum change by the same amount). Duration-based hedges are therefore inaccurate against yield changes that involve shifts in the slope of the yield curve.

[5] There are two standard refinements. (1) We can take a second-order rather than first-order approximation to the bond price change. The second-order term – known as convexity – is related to the change in duration as yield changes, and this duration–convexity approach gives us a better approximation to the bond price change as the yield changes. However, duration–convexity usually gives only modest improvements in accuracy over the basic duration approach. (2) An alternative refinement is to use key rate durations (Ho (1992)): if we are concerned about shifts in the yield curve, we can construct separate duration measures for yields of specified maturities (e.g., short-term and long-term yields); these would give us estimates of our exposure to changes in these specific yields and allow us to accommodate non-parallel shifts in the yield curve.

1.3.4 Portfolio Theory

A somewhat different approach to risk measurement is provided by portfolio theory. Portfolio theory starts from the premise that investors choose between portfolios on the basis of their expected return, on the one hand, and the standard deviation (or variance) of their return, on the other.[6] The standard deviation of the portfolio return can be regarded as a measure of the portfolio's risk. Other things being equal, an investor wants a portfolio whose return has a high expected value and a low standard deviation. These objectives imply that the investor should choose a portfolio that maximises expected return for any given portfolio standard deviation. A portfolio that meets these conditions is efficient, and a rational investor will always choose an efficient portfolio. When faced with an investment decision, the investor must therefore determine the set of efficient portfolios and rule out the rest. Some efficient portfolios will have more risk than others, but the more risky ones will also have higher expected returns. Faced with the set of efficient portfolios, the investor chooses one particular portfolio on the basis of his or her own preferred trade-off between risk and expected return. An investor who is very averse to risk will choose a safe portfolio with a low standard deviation and a low expected return, and an investor who is less risk averse will choose a riskier portfolio with a higher expected return.

One of the key insights of portfolio theory is that the risk of any individual asset is not the standard deviation of the return to that asset, but the extent to which that asset contributes to overall portfolio risk. An asset might be very risky (i.e., have a high standard deviation) when considered on its own, and yet have a return that correlates with the returns to other assets in our portfolio in such a way that acquiring the new asset does not increase the overall portfolio standard deviation. Acquiring the new asset would then be riskless, even though the asset held on its own would still be risky. The moral of the story is that the extent to which a new asset contributes to portfolio risk depends on the correlation or covariance of its return with the returns to the other assets in our portfolio – or, if one prefers, the beta, which is equal to the covariance between the return to asset i and the return to the portfolio divided by the variance of the portfolio return. The lower the correlation, other things being equal, the less the asset contributes to overall risk. Indeed, if the correlation is negative, it will offset existing risks and lower the portfolio standard deviation.

Portfolio theory provides a useful framework for handling multiple risks taking account of how those risks interact with each other. It is therefore of obvious use to – and is widely used by – portfolio managers, mutual fund managers and other investors. However, it tends to run into estimation and data problems. The risk-free return is not too difficult to estimate, but estimating the expected market return and the betas is often problematic. The expected market return is highly subjective and each beta is specific not only to the individual asset to which it refers, but also to our current portfolio. To estimate a beta coefficient accurately, we need data on the returns to the new asset and the returns to all our existing assets, and we need a sufficiently long data set to make our statistical estimation techniques reliable. The beta also depends on our existing portfolio and we should, in theory, re-estimate all our betas every time our portfolio changes.

For some time after it was first advanced in the 1950s, the data and calculation requirements of portfolio theory led many to see it as quite impractical. To get around some of these problems,

[6] This framework is often known as the mean–variance framework, because it presupposes that knowledge of the mean and variance (or standard deviation) of portfolio returns are sufficient to guide investors' decisions. This also implies that the variance/standard deviation of portfolio returns can be regarded as a measure of the portfolio risk. We shall have more to say on these issues in Chapter 2.

William Sharpe and others in the 1960s suggested a short-cut, the famous (or infamous) Capital Asset Pricing Model (CAPM), the essence of which was to work with betas estimated against a hypothetical market portfolio. The CAPM made portfolio theory much more practical given the data and computational power available at the time. However, it *is* a short-cut, and is potentially misleading, not least because the widespread adoption of the CAPM often leads people to talk about *the* beta for an asset, as if the asset had only a single beta. The CAPM gives us good answers if the CAPM beta estimated against the hypothetical market portfolio is close to the 'true' beta evaluated against the portfolio we actually hold, and in practice we seldom know whether it is.[7] If the two portfolios are sufficiently different, the 'true' beta might be very different from the CAPM beta, and the CAPM could be very misleading.[8] However, even in its more general form, portfolio theory can also be unreliable, not least because we might have poor parameter estimates and because the underlying assumptions on which it is based – that risks are normal, or near normal – might not be appropriate. We shall have more to say on these issues presently.

1.3.5 Derivatives Risk Measures

When dealing with derivatives positions, we can also estimate their risks by their Greek parameters: the delta, which gives us the change in the derivatives price in response to a small change in the underlying price; the gamma, which gives us the change in the delta in response to a small change in the underlying price (or, if we prefer, the second derivative of the derivative's price with respect to a change in the underlying price); the rho, which gives us the change in derivatives price for a small change in the interest rate; the vega, which gives us the change in derivatives price with respect to a change in volatility; the theta, which gives us the change in derivatives price with respect to time; and so on. A seasoned derivatives practitioner can make good use of estimates of these parameters to assess and manage the risks of a derivatives position, but doing so requires considerable skill. The practitioner needs to be able to deal with a number of different risk 'signals' at the same time, under real-time constraints, and the Greeks themselves can be volatile: for instance, the gamma of an at-the-money vanilla option becomes increasingly unstable as the option approaches expiry, and the volatility of vega is legendary.

However, the use of these measures for financial risk management makes sense only within the confines of a dynamic hedging strategy: the measures, and resulting hedge positions, only work against small changes in risk factors, and only then if they are revised sufficiently frequently. There is always a worry that these measures and their associated hedging strategies

[7] There are also other problems. (1) If we wish to use this short-cut, we have relatively little firm guidance on what the hypothetical portfolio should be. In practice, investors usually use some 'obvious' portfolio such as the basket of shares behind a stock index, but we never really know whether this is a good proxy for the Capital Asset Pricing Model (CAPM) market portfolio or not. It is probably not. (2) Even if we pick a good proxy for the CAPM market portfolio, it is still doubtful that *any* such portfolio will give us good results. If we wish to use proxy risk estimates, there is a good argument that we should abandon single-factor models in favour of multi-factor models that can mop up more systematic risks. This leads us to the Arbitrage Pricing Theory (APT) of Ross (1976). However, the APT has its own problems: we can't easily identify the risk factors, and even if we did identify them, we still don't know whether the APT will give us a good proxy for the systematic risk we are trying to proxy.

[8] We can also estimate risks using statistical reduced-form approaches. The idea is that we postulate a measurable relationship between the exposure-variable we are interested in (e.g., the loss/gain on our bond or FX portfolio or whatever) and the factors that we think influence that loss or gain. We then estimate the parameters of this relationship by an appropriate econometric technique, and the parameter estimates give us an idea of our risk exposures. This approach is limited by the availability of data (i.e., we need enough data to estimate the relevant parameters) and (usually) by linearity assumptions, and it is also usually limited to market risks, because we normally have data only on the prices of marketable securities. There can also be problems caused by misspecification and instability in estimated statistical relationships.

might fail to cover us against major market moves such as stock market or bond market crashes, or a major devaluation. We may have hedged against a small price change, but a large adverse price move in the wrong direction could still be very damaging: our underlying position might take a large loss that is not adequately compensated for by the gain on our hedge instrument. There is also the danger that we may be dealing with a market whose liquidity dries up just as we most need to sell. When the stock market crashed in October 1987, the wave of sell orders prompted by the stock market fall meant that such orders could take hours to execute, and sellers got even lower prices than they had anticipated. The combination of a large market move and the sudden drying up of market liquidity can mean that positions take large losses even though they are supposedly protected by dynamic hedging strategies. It was this sort of problem that undid portfolio insurance and other dynamic hedging strategies in the stock market crash, when many people suffered large losses on positions that they thought they had hedged.

1.4 VALUE AT RISK

1.4.1 The Origin and Development of VaR

In the late 1970s and 1980s, a number of major financial institutions started work on internal models to measure and aggregate risks across the institution as a whole. They started work on these models for their own risk management purposes – as firms became more complex, it was becoming increasingly difficult, and yet also increasingly important, to be able to aggregate their risks taking account of how they interact with each other, and firms lacked the means to do so. These firms were also running into problems managing risks across increasingly diverse positions. They would impose limits on traders and asset managers, but with the information and management systems available at the time, the limits were enforced on a piecemeal basis, and all sorts of inconsistencies and other undesirable effects would result: 'good' trades or investments would be passed over because they ran into arbitrary limits, risks were being taken with inadequate awareness of their overall effects on the firm, reducing risk in one area seldom allowed greater risk-taking elsewhere, and so on. Good capital allocation was also undermined in much the same way. But perhaps the biggest problem was the absence of integrated risk management. There was little consistency between the limits imposed (and capital allowed) and the risks being taken, and the committees entrusted with setting and managing limits and with capital allocation lacked the tools to do much better.

As firms wrestled with these problems, a consensus gradually evolved that what was needed was some sense of the probability of losses at the firmwide level. This gave rise to the notion of value at risk (or VaR), which enabled firms to get a better sense of their overall risks, and to achieve a more rational allocation of limits and capital down the various business lines.

The best known of these systems is the RiskMetrics system developed by JP Morgan. According to industry legend, this system originated when the chairman of JP Morgan, Dennis Weatherstone, asked his staff to give him a daily one-page report indicating risk and potential losses over the next 24 hours, across the bank's entire trading portfolio. This report – the famous '4:15 report' – was to be given to him at 4:15 each day, after the close of trading. To achieve this objective, the Morgan staff had to develop a system to measure risks across different trading positions, across the whole institution, and also aggregate these risks into a single risk measure. The measure used was value at risk (or VaR), or the maximum likely loss over the next trading day, and the VaR was estimated from a system based on standard

portfolio theory, using estimates of the standard deviations and correlations between the returns to different traded instruments. While the theory was straightforward, making this system operational involved a huge amount of work: measurement conventions had to be chosen, data sets constructed, statistical assumptions agreed, procedures chosen to estimate volatilities and correlations, computing systems established to carry out estimations, and many other practical problems had to be resolved. Developing this methodology took a long time, but the main elements – the data systems, the risk measurement methodology, and the basic mechanics – were all in place and working by around 1990. At that point it was decided to start using the '4.15 report', and it was soon found that the new risk management system had a major positive effect. In particular, it 'sensitised senior management to risk–return trade-offs and led over time to a much more efficient allocation of risks across the trading businesses'.[9] The new risk system was highlighted in JP Morgan's 1993 research conference and aroused a great deal of interest from potential clients who wished to buy or lease it for their own purposes.

The publication of the G30 and other reports around the same time highlighted the potential usefulness of VaR systems in a much more prominent way, and the notion of VaR itself, almost unknown in 1990, rapidly became the most talked about subject in the risk management field: very soon every self-respecting financial institution wanted its own VaR system.

Meanwhile, other financial institutions had been working on their own internal models, and VaR software systems were also being developed by specialist companies that concentrated on software but were not in a position to provide data. The resulting systems differed quite considerably from each other. Even where they were based on broadly similar theoretical ideas, there were still considerable differences in terms of subsidiary assumptions, use of data, parameter estimation procedures, and many other 'details'. Besides, not all VaR systems were based on portfolio theory: some systems were built using historical simulation approaches that estimate VaR from histograms of past profit and loss data, and other systems were developed using Monte Carlo (or random number) simulation techniques.

These firms were keen to develop and strengthen their management consultancy businesses, but at the same time they were conscious of the limitations of their own models and were wary about giving too many secrets away. While most firms kept their models secret, JP Morgan decided to make its data and basic methodology available so that outside parties could use them to write their own risk management software. Early in 1994, Morgan set up the RiskMetrics unit to do this and the RiskMetrics model – a simplified version of the firm's own internal model – was completed in eight months. In October that year, Morgan made its RiskMetrics system and the necessary data freely available on the internet: outside users could now access the RiskMetrics model and plug their own position data into it.

This bold move attracted a lot of attention, and the resulting public debate about the merits of RiskMetrics was useful in further raising awareness of VaR and of the issues involved in building and operating VaR systems.[10] Making the RiskMetrics data available also gave a major boost to the spread of VaR systems by giving software providers and their clients access to data sets that they were often unable to construct themselves. It also encouraged many of

[9] Guldimann (2000), p. 57.

[10] A good example is the exchange between Longerstaey and Zangari (1995), on the one hand, and Lawrence and Robinson (1995a), on the other, on the safety or otherwise of RiskMetrics. The issues covered in this debate – the validity of underlying statistical assumptions, the estimation of volatilities and correlations, and similar issues – go right to the heart of risk measurement, and will be dealt with in more detail in later chapters.

the smaller software providers to adopt the RiskMetrics approach or make their own systems compatible with it.

The subsequent adoption of VaR systems was very rapid, first among securities houses and investment banks, and then among commercial banks, pension funds and other financial institutions, and non-financial corporates. As the models proliferated, the VaR concept also became more familiar, and by the mid-1990s the VaR had already established itself as *the* dominant measure of financial risk – a meteoric rise when one considers that the VaR was almost unknown only a few years earlier. The state of the art also improved rapidly. Developers and users became more experienced; the combination of plummeting IT costs and continuing software development meant that systems became more powerful, much faster, and much more sophisticated; VaR systems were extended to cover more types of instruments; and the VaR methodology itself was extended to deal with other types of risk besides the market risks for which VaR systems were first developed, including credit risks, liquidity (or cash flow) risks and operational risks.

Box 1.2 Portfolio Theory and VaR

In some respects VaR is a natural progression from earlier portfolio theory (PT). Yet there are also important differences between them:

- PT interprets risk in terms of the standard deviation of the return, while VaR approaches interpret it in terms of the maximum likely loss. The VaR notion of risk – the VAR itself – is more intuitive and (arguably) easier for laypeople to grasp.
- PT presupposes that P/L or returns are normally (or near normally) distributed, whereas VaR approaches can accommodate a very wide range of possible distributions. VaR approaches are therefore much more flexible.
- VaR approaches can be plausibly applied to a much broader range of risk problems: PT theory is (for the most part) limited to market risks, while VaR approaches can be applied much more flexibly to credit, liquidity and other risks as well.
- The variance–covariance approach to VaR has the same theoretical basis as PT – in fact, its theoretical basis *is* portfolio theory – but the other two main approaches to VaR (i.e., the historical simulation and simulation approaches) do not. VaR systems can be based on a wider range of estimation methods.

1.4.2 Attractions of VaR

So what is VaR, and why is it important? The basic concept was nicely described by Linsmeier and Pearson:

> Losses greater than the value at risk are suffered only with a specified small probability. Subject to the simplifying assumptions used in its calculation, value at risk aggregates all of the risks in a portfolio into a single number suitable for use in the boardroom, reporting to regulators, or disclosure in an annual report. Once one crosses the hurdle of using a statistical measure, the concept of value at risk is straightforward to understand. It is simply a way to describe the magnitude of the likely losses on the portfolio.[11]

[11] Linsmeier and Pearson (1996), p. 3.

The VaR has a number of significant attractions over traditional risk measures:

- VaR provides a *common* consistent measure of risk across different positions and risk factors. VaR can be applied to any type of portfolio, and it enables us to compare the risks of different portfolios. It enables us to measure the risk associated with a fixed-income position, say, in a way that is comparable to the risk associated with an equity position. In this the VaR is a distinct improvement over traditional methods (e.g., duration and convexity approaches only apply to fixed-income positions, Greek risk measures only apply to derivatives positions, and portfolio theory approaches apply to equity and similar (e.g., commodity) positions).
- VaR enables us to *aggregate* the risks of subpositions into an overall measure of portfolio risk, and in doing so take account of the ways in which different risk factors interact (or correlate) with each other. This is another attraction of the VaR, because most traditional risk measures do not (easily) allow for the 'sensible' aggregation of component risks.
- VaR is *holistic* in that it takes full account of all driving risk factors, whereas many traditional approaches either only look at risk factors one at a time (e.g., Greek measures) or else resort to simplifications to collapse multiple risk factors into one (e.g., duration–convexity approaches collapse the spot-rate curve into a single yield, and CAPM approaches collapse different equity returns into a single 'market' return). VaR is also holistic in another sense: it focuses assessment on a complete portfolio, often at the firmwide level, and not just on individual positions in it.
- VaR is *probabilistic*, and gives a risk manager useful information on the probabilities associated with specified loss amounts. By comparison, many traditional measures (e.g., duration–convexity, Greeks, etc.) only give us the answers to 'what if' questions and don't give an indication of likelihoods.
- VaR is expressed in the simplest and most easily understood *unit of measure*, namely, 'lost money'. Many other measures are expressed in less transparent units of measure (e.g., average period to cash flow, etc.). Hence, the VaR is expressed in terms of a unit that is easier to convey.

These are significant attractions, which do a lot to explain why VaR became so popular.

VaR information can be used in many ways (albeit, not without problems, some of which we will address in due course): (1) Senior management can use it to set their overall risk target, and from that determine risk targets and position limits down the line. If they want the firm to increase its risks, they would increase the overall VaR target, and vice versa. (2) VaR can be used to determine capital requirements, both at the firmwide and business-unit level: the riskier the activity, the higher the VaR and the greater the capital requirement. VaR can also be used to specify the position limits imposed on business units. (3) VaR can be useful for reporting and disclosing purposes, and firms increasingly make a point of reporting VaR information in their annual reports.[12] (4) VaR-based decision rules can guide investment, hedging, trading and portfolio management decisions, and do so taking account of the implications of alternative choices for the portfolio risk as a whole.[13] It can also be used to carry out portfolio-wide (or macro) hedging strategies that are otherwise difficult to implement.[14] (5) VaR information can be used to provide new remuneration rules for traders, managers and other employees that take

[12] For more on the use of VaR for reporting and disclosure purposes, see, e.g., Dowd (2000b), Jorion (2002), Moosa and Knight (2001) or Woods *et al.* (2004).

[13] VaR-based decision rules are covered more fully in, e.g., Dembo (1997), Lucas and Klaassen (1998), Dowd (1999a,c) and Sentana (2001).

[14] Such strategies are explained in more detail in, e.g., Litterman (1996), Kuruc and Lee (1998) and Dowd (1999a).

account of the risks they take, and so discourage the excessive risk-taking that occurs when employees are rewarded on the basis of profits alone, without any reference to the risks they took to get those profits. (6) Systems based on VaR methodologies can be used to measure other risks such as credit, liquidity and operational risks. In short, VaR can help provide for a more consistent and more integrated approach to the management of different financial risks, and so lead to better risk management overall.

1.4.3 Criticisms of VaR

Most risk practitioners embraced VaR with varying degrees of enthusiasm, but there were also those who warned that VaR had deeper problems and could be dangerous.

A key issue was the validity or otherwise of the statistical and other assumptions underlying VaR, and both Nassim Taleb (1997a,b)[15] and Richard Hoppe (1998) were critical of the naïve transfer of mathematical and statistical models from the physical sciences where they are well suited to social systems where they were often invalid. Such applications often ignore important features of social systems – the ways in which intelligent agents learn and react to their environment, the non-stationarity and dynamic interdependence of many market processes, and so forth – features that undermine the plausibility of many models and leave VaR estimates wide open to major errors.

A related argument was that VaR estimates are too imprecise to be of much use, and empirical evidence on this issue is worrying, as it suggests that different VaR models can give vastly different VaR estimates (see, e.g., Beder (1995)). To make matters worse, VaR models are also exposed to considerable implementation risk as well – so even theoretically similar models could give quite different VaR estimates because of differences in the ways in which the models were implemented (Marshall and Siegel (1997)).

The danger here is obvious: if VaR estimates are too inaccurate and users take them seriously, they could take on much bigger risks and lose much more than they had bargained for. As Taleb put it, 'You're worse off relying on misleading information than on not having any information at all. If you give a pilot an altimeter that is sometimes defective he will crash the plane. Give him nothing and he will look out the window.'[16] Such criticism is not easy to counter.

A deeper problem is that risk is endogenous: if VaR is used to control or remunerate risk-taking, those being controlled will respond to VaR constraints in the pursuit of their own interest. For example, traders will have an incentive to seek out and trade positions where risk is over- or underestimated (Ju and Pearson (1999)). They will therefore take on more risk than suggested by VaR estimates that fail to take account of how traders or other affected parties will respond – so our VaR estimates will be biased downwards – and the evidence suggests that the magnitude of these underestimates can be very substantial. VaR limits might also encourage traders to respond by taking more low-probability, high-impact risks, their motivation being that such risks are likely to pay off (because they increase earnings in normal times), and the occasional very high loss is allowable because it meets the VaR constraint: VaR limits can encourage traders to write deep-out-of-the-money options against their employers' assets.

[15] Taleb was also critical of the tendency of some VaR proponents to overstate the usefulness of VaR. He was particularly dismissive of Philippe Jorion's (1997) claim that VaR might have prevented disasters such as Orange County. Taleb's response was that these disasters had other causes – especially, excessive leverage. As he put it, a Wall Street clerk would have picked up these excesses with an abacus, and VaR defenders overlook the point that there are simpler and more reliable risk measures than VaR. Taleb is clearly right: any simple duration analysis should have revealed the rough magnitude of Orange County's interest-rate exposure. The root problem with Orange County was not the absence of VaR, as such, but the absence of risk management.

[16] Taleb (1997a), p. 37.

There are also good reasons to think that the use of VaR as a regulatory constraint might discourage good risk management practices.[17]

Others suggested that the use of VaR might destabilise the financial system. Thus, Taleb (1997a) pointed out that VaR players are dynamic hedgers, and need to revise their positions in the face of changes in market prices. If everyone uses VaR, there is a danger that this hedging behaviour will make uncorrelated risks become very correlated – and again firms will bear much greater risk than their VaR models might suggest. Poorly thought through regulatory VaR constraints can also destabilise the financial system by inducing banks to increase their risk-taking: for example, a VaR cap gives risk managers an incentive to protect themselves against mild losses, but not against larger ones in excess of VaR. VaR regulatory constraints can also exacerbate cyclical effects, and so aggravate financial crises, or even bring them about.[18]

Proponents of VaR could respond that many of these criticisms are not specific to VaR as such, but would also apply (in varying extents) to other risk measures as well. They could also argue that some of these problems were due to the misuse of VaR rather than to the VaR itself; for example, they could claim with considerable justification that some of the problems regarding the regulatory use of VaR were due to the failings of the regulatory system. But not all these criticisms can be answered, and new problems were also coming to light that would, in time, not only undermine the VaR's position as the dominant measure of financial risk, but also destroy the VaR's claim to be regarded as a 'proper' measure of financial risk in the true sense of the term. This takes us to the next major development in financial risk management, the theory of coherent financial risk measures, and this is perhaps best deferred to the next chapter.

[17] See, e.g., Danielsson and Zigrand (2001).
[18] See Danielsson (2002), Danielsson and Zigrand (2001), Basak and Shapiro (2001) and Danielsson *et al.* (2001).

Appendix
Types of Market Risk

As explained in the text, market risk is the risk of loss associated with unexpected movements in market prices or rates, and is to be distinguished from other types of risk such as credit risk (involving possible losses from default events) and operational risk (involving losses from people or systems failures). However, market risk cannot be entirely divorced from these other types of risk, and can sometimes be created by them. For example, credit (i.e., default) events can lead to changes in market prices or rates (e.g., they might affect bond spreads or bond prices, or the prices of credit derivatives) and so trigger market losses. Operational events can also lead to market losses, a good example being the collapse in the value of the shares in Barings Bank to £1.

There are many different categories of market risk, and the most common categories include:

- Equity risks: the market risks associated with positions in equity markets.
- Fixed-income risks: the risks associated with positions in fixed-income instruments (e.g., bonds and structured notes) and interest-sensitive instruments generally (e.g., interest-rate derivatives).
- FX risks: risks associated with foreign and cross-currency positions.
- Commodity risks: risks on agricultural, energy, metals, and similar positions.
- Miscellaneous market risks: risks on weather, temperature, and catastrophe instruments, underwriting risks on insurance portfolios, the risks of property positions, and so on.

Market risks also differ enormously in terms of the types of players who are taking the relevant risks, the reasons they take them, the horizons to which they operate, and the valuation methods used to establish profits or losses. These factors are all interrelated, and are also related to the types of instrument involved and the risk management methods used.

One way to look at these issues is from the perspective of context. To begin with, there is a trading context. Players in a trading context might be securities houses and similar firms who trade equities, futures, options, bonds, and so on. They would operate on shorter horizons, and often on horizons of a trading day, or of one or two weeks at five trading days to a week, and some traders even operate on intra-day horizons. They would operate on both organised exchanges and relatively liquid OTC (over-the-counter) markets. The sophistication of the instruments traded will vary enormously, from very simple vanilla instruments at one end to highly sophisticated instruments at the other. They would also operate with advanced market risk measurement systems, and they would typically value their positions (and so determine gains and losses) using mark-to-market valuation methodologies. In addition, these traders would often aim to beat benchmarks, although the type of benchmark used would depend on the market. Performance evaluation would sometimes involve an adjustment for risks taken, and sometimes not.

There is also an investment context, which would include many of the activities of banks. In this context, firms operate to longer horizons and deal in instruments that often involve significant credit risk (i.e., so market and credit risks are very closely related), and valuation

is difficult because secondary markets are thin or non-existent. These operators would include many banks, who would be concerned with medium-term horizons, and insurance and pension funds, with the latter in particular being concerned with very long-term horizons. Risk assessment is often not as sophisticated as in a trading context, and it is less common for performance evaluation to be risk-related. There are also other contexts. These include a treasury (or payments-related) context, in which financial institutions are involved in large-scale electronic funds transfer (EFT) through payments systems such as SWIFT, CHIPS, etc. The amounts involved are massive, and there are major concerns with intra-day credit exposures and the payments at risk. There is also an insurance context, in which insurance companies trade market risk exposures obtained through their underwriting activities, and a property context, in which operators buy and sell exposures to property prices.

There are related differences in terms of valuation methodologies. Essentially, market positions can be valued in one of three main ways. The first and in many ways the best valuation methodology is mark-to-market: positions are revalued periodically at current market prices, and investors realise their profits/losses with each such revaluation. This methodology works well with liquid markets and fairly clear end-of-day market prices based on real market trades. The classic example is where positions on organised markets are marked-to-market at the end of each trading day, and traders' margins are adjusted accordingly so that all gains or losses are realised immediately. The applicability of mark-to-market therefore depends on having a liquid market for the instruments concerned, and mark-to-market is closely associated with short-term trading horizons.

A second valuation methodology is mark-to-model: positions are revalued against hypothetical prices generated by a model. This is often used as a substitute for mark-to-market in situations where mark-to-market is not feasible, so 'current' market prices either do not exist or are unreliable because trading is too thin. Mark-to-model is not as reliable as mark-to-market and depends crucially on the validity of the models used, including the ways in which they are calibrated. The models used therefore have to be carefully chosen and calibrated, and regularly reviewed. Mark-to-model is also open to abuse if the models are poorly chosen or tampered with, and there have been numerous cases of hidden losses and fraud related to the misuse of valuation models (e.g., a common one being to fiddle the values of volatility parameters in option-pricing models to artificially boost mark-to-model valuations and, of course, to boost the bonuses that go with them).

A third approach to valuation is book valuation using standard accounting methods (i.e., GAAP). In theory, book valuation is highly questionable and has many well-known deficiencies when used to value market positions: historical costs often give poor indications of current values; book valuation treats depreciation in a crude and ad hoc manner, and gives scope for losses to be hidden, and earnings are often excessively smoothed. However, it can be applied in circumstances where other approaches are not feasible or are themselves even less reliable, and one of its advantages is precisely that its deficiencies are well understood. Managers using book methods can therefore make their own discretionary adjustments to take account of any biases or other problems they perceive in the book values.

We can also differentiate market risks in terms of where they stand along a liquidity spectrum that recognises both the common elements among different risks as well as their differences from each other.[19] There are four key points along this spectrum:

[19] This was suggested and explained further by Drzik (1996).

- *Smooth markets*: These are textbook liquid markets with large numbers of participants and high turnovers. Standard examples are currency markets or markets for US Treasuries. Positions are easily valued by marking to market and risk is assessed on the basis of VaR supplemented by stress tests.
- *Choppy markets*: These markets are less deep and less liquid, and have fewer participants and lower turnover. Examples are OTC markets in equities. Participants tend to use similar methods as participants in smooth markets, but these methods are less reliable in choppy markets because of liquidity and valuation problems.
- *Icy markets*: These markets are even thinner and less liquid, and secondary markets exist but are very limited. Trades tend to be negotiated rather than screen based, and prices are often not transferable across deals. Pricing is usually mark-to-model supplemented by some marking to market with adjustments for liquidity and other concerns. Risk calculations are mainly carried out by loss modelling, with some VaR analysis adjusted for liquidity risks.
- *Frozen markets*: These markets are extremely thin and there are few, if any, secondary markets. Assets are usually bought to hold to maturity and products are highly tailored. Pricing is highly judgemental, and often based on reserve-adjusted book values supplemented by marking to model and relatively arbitrary allowances for illiquidity. Risk evaluation is highly problematic.

Each point on this continuum is characterised by certain features:

- *Valuation*: Market prices are good valuation guides in smooth markets, somewhat less useful in choppy markets, even less use in icy markets, and of no use in frozen ones. We therefore rely almost entirely on mark-to-market valuation for smooth markets. When markets become choppy we make more use of mark-to-model valuation and to some extent judgemental methods. With icy markets the balance shifts further towards judgemental and book methods, and with frozen markets we rely on little else.
- *Risk control*: Risk control in smooth markets is straightforward, since profits and losses are easy to ascertain, and relies mainly on limits backed up by *ex post* monitoring and performance evaluation. Risk control in choppy markets is similar except for the need to pay more attention to issues of liquidity and unrealised gains/losses. In icy and frozen markets we have to devote even more attention to these issues, and also to systems to mitigate or stop losses from accumulating.
- *Remuneration*: As markets become less liquid, we should place increasing emphasis on deferred compensation. Deferment enables us to reduce the compensation of individuals whose decisions produce losses that only become apparent later. The prospect of such penalties gives them more incentive to act responsibly.

In short, market risks are complex and multidimensional. In attempting to estimate them, it is very important to keep in mind that any estimates we make are very dependent on the valuation methodologies used, and these are closely related to market liquidity and the assumptions embodied in the ways in which valuation methodologies are chosen and implemented, and these assumptions may not be valid. Thus, estimates of market risk are inevitably subject both to *liquidity risk* and to more general *model risk*.

2

Measures of Financial Risk

This chapter deals with measures of financial risk. As we have already seen, work on financial risk management over the last decade or so has tended to focus on the VaR, but there are many other risk measures to choose from, and it is important to choose the 'right' one. To put our discussion into a specific context, suppose we are working to a daily horizon period. At the end of day $t - 1$, we observe that the value of our portfolio is P_{t-1} but, looking forward, the value of our portfolio at the end of tomorrow, P_t, is uncertain. Ignoring any intra-day returns or intra-day interest, if P_t turns out to exceed P_{t-1}, we will make a profit equal to the difference, $P_t - P_{t-1}$; and if P_t turns out to be less than P_{t-1}, we will make a loss equal to $P_{t-1} - P_t$. Since P_t is uncertain, as viewed from the end of $t - 1$, then so too is the profit or loss (P/L) (or return). Thus, our next-period P/L (or return) is risky, and we want a framework or paradigm to measure this risk.

This chapter examines three such measurement frameworks, the first based on the mean–variance or portfolio-theory approach, the second based on VaR, and the third based on the newer coherent risk measures mentioned at the end of the last chapter. We will discuss these in their chronological order, but before discussing them in any detail, it is worth highlighting the themes underlying the ways in which these frameworks have evolved. Three themes in particular stand out:

- The first is the drive to extend the range of P/L or return distributions that can be handled. The mean–variance framework is quite limited in this regard, as it only applies if we are dealing with normal or near-normal distributions – or, more precisely, if we are dealing with data that are (or can be transformed to become) elliptically distributed. By contrast, the later frameworks are very general and can accommodate any form of distribution (although some distributions are much easier to handle than others). So a key theme is the desire to escape from the confines of a framework that can only handle normal or near-normal distributions, and this is very important because many of the empirical distributions we might encounter are very non-normal.
- A second and related theme is to improve the usefulness of the resulting risk measure. In the mean–variance framework, the measure of risk is the standard deviation (of P/L or returns) or some simple transformation of it. In the normal (or near-normal) circumstances assumed in that framework, this standard deviation is also an ideal risk measure, and we can use it to identify our risk-expected return trade-off and make decisions accordingly. For its part, the VaR framework gives us a risk measure, the VaR, that is more or less equivalent in usefulness to the standard deviation if we are dealing with normal (or near-normal) distributions. The VaR also has the advantage that it can be estimated for any distribution, but it has major problems as a *usable* risk measure in the presence of seriously non-normal distributions. The VaR framework therefore liberates us from the confines of near normality in the sense that it provides a risk measure that can be estimated for any distribution we like, but this turns out to be an empty victory, because the *usefulness* of VaR as a measure of risk is highly questionable outside the confines of near-normal distributions. This problem motivated the

development of the third and latest framework, the coherent framework: this provides risk measures that have the benefits of the VaR (i.e., they apply to any distribution) but, unlike the VaR, can be used more reliably for decision-making in the presence of seriously non-normal distributions. So, in short, the second theme is the drive to produce risk measures that can be useful outside the confines of near-normality.

- There is also a third theme. Each framework allows us to aggregate individual risks in an intellectually respectable way, but the portfolio theory approach is rather limited in its range of application – essentially, it applies to equity and similar types of risks – whereas the VaR and coherent approaches are much more general in their ranges of application. However, this greater range of application comes at a cost: we have to deal with problems of valuation and market illiquidity that do not usually arise in the more limited cases considered by classical portfolio theory, and a considerable amount of effort has gone into dealing with these sorts of problems. The importance of being able to 'generalise' the range of applicability of our risk measures has been further reinforced by the emergence of enterprise-wide risk management (ERM; sometimes also known as integrated risk management) as a major theme of financial risk management since the late 1990s. ERM seeks to measure and manage risks across different categories in a holistic and integrated way across the firm as a whole, and in doing so take account of the ways in which different risk categories interact with each other. ERM is a hugely important theme of modern risk management and we don't have space to say much about it in this book, but it suffices for the moment to note that successful ERM *presupposes* a risk measurement framework that is capable of extension across the major risk categories that a firm has to deal with, and the basic portfolio-theory framework is far from adequate to this task.

We should keep these three themes in mind as we now proceed to examine each of these frameworks in turn.

2.1 THE MEAN–VARIANCE FRAMEWORK FOR MEASURING FINANCIAL RISK

The traditional approach used to measure financial risks is the mean–variance framework: we model financial risk in terms of the mean and variance (or standard deviation, the square root of the variance) of P/L (or returns).[1] As a convenient (although oversimplified) starting point, we can regard this framework as underpinned by the assumption that daily P/L (or returns) obeys a normal distribution.[2] A random variable X is normally distributed with mean μ and variance σ^2 (or standard deviation σ) if the probability that X takes the value x, $f(x)$, obeys the following probability density function (pdf):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{2.1}$$

where x is defined over $-\infty < x < \infty$. A normal pdf with mean 0 and standard deviation 1, known as a standard normal, is illustrated in Figure 2.1.

A pdf gives a complete representation of possible random outcomes: it tells us what outcomes are possible, and how likely these outcomes are. This particular pdf is the classic bell curve.

[1] For a good account of portfolio theory and how it is used, see, e.g., Elton and Gruber (1995).

[2] To simplify the text, we shall sometimes talk 'as if' the mean–variance framework requires normality. However, the mean–variance approach in fact only requires that we assume ellipticality and elliptical distributions are more general. (We have more to say on ellipticality in Appendix 2 to this chapter.) Nonetheless, the mean–variance framework is most easily understood in terms of an underlying normality assumption, and non-normal elliptical distributions are harder to understand, less tractable and in any case share many of the same features as normality.

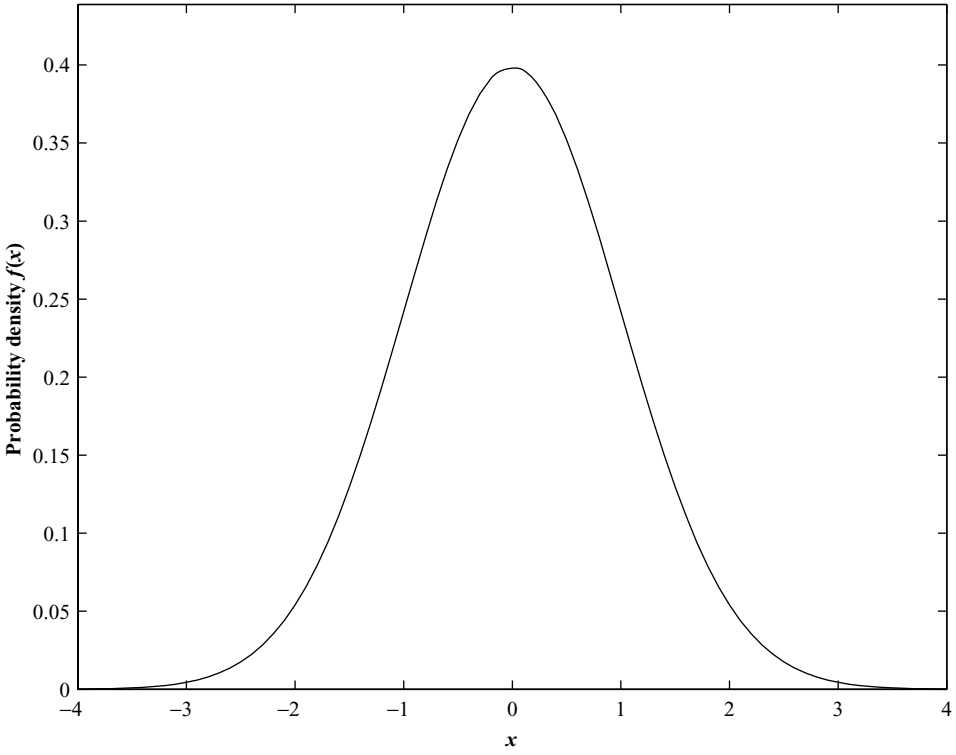

Figure 2.1 The normal probability density function

It tells us that outcomes (or x-values) are more likely to occur close to the mean μ; it also tells us that the spread of the probability mass around the mean depends on the standard deviation σ: the greater the standard deviation, the more dispersed the probability mass. The pdf is also symmetric around the mean: X is as likely to take a particular value $\mu + x$ as to take the corresponding negative value $\mu - x$. The pdf falls as we move further away from the mean, and outcomes well away from the mean are very unlikely, because the tail probabilities diminish exponentially as we go further out into the tail. In risk management, we are particularly concerned about outcomes in the left-hand tail, which corresponds to high negative returns – or big losses, in plain English.

The assumption of normality is attractive for various reasons. One reason is that it often has some, albeit limited, plausibility in circumstances where we can appeal to the central limit theorem. Another attraction is that it provides us with straightforward formulas for both cumulative probabilities and quantiles, namely:

$$\Pr[X \leq x] = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dX \tag{2.2a}$$

$$q_\alpha = \mu + \sigma z_\alpha \tag{2.2b}$$

where α is the chosen confidence level (e.g., 95%), and z_α is the standard normal variate for that confidence level (e.g., so $z_{0.95} = 1.645$). z_α can be obtained from standard statistical tables or from spreadsheet functions (e.g., the 'normsinv' function in Excel or the 'norminv'

function in MATLAB). Equation (2.2a) is the normal distribution (or cumulative density) function: it gives the normal probability of X being less than or equal to x, and enables us to answer probability questions. Equation (2.2b) is the normal quantile corresponding to the confidence level α (i.e., the lowest value we can expect at the stated confidence level) and enables us to answer quantity questions.

A related attraction of particular importance in the present context is that the normal distribution requires only two parameters – the mean and the standard deviation (or variance), and these parameters have ready financial interpretations: the mean is the expected return on a position, and the standard deviation can be interpreted as the risk associated with that position. This latter point is perhaps the key characteristic of the mean–variance framework: it tells us that we can use the standard deviation (or some function of it, such as the variance) as our measure of risk. And conversely, the use of the standard deviation as our risk measure indicates that we are buying into the assumptions – normality or, more generally, ellipticality – on which that framework is built.

To illustrate how the mean–variance approach works, suppose we wish to construct a portfolio from a particular universe of financial assets. We are concerned about the expected return on the portfolio, and about the variance or standard deviation of its returns. The expected return and standard deviation of return depend on the composition of the portfolio, and assuming that there is no risk-free asset for the moment, the various possibilities are shown by the curve in Figure 2.2: any point inside this region (i.e., below or on the curve) is attainable by a suitable asset combination. Points outside this region are not attainable. Since the investor regards a higher

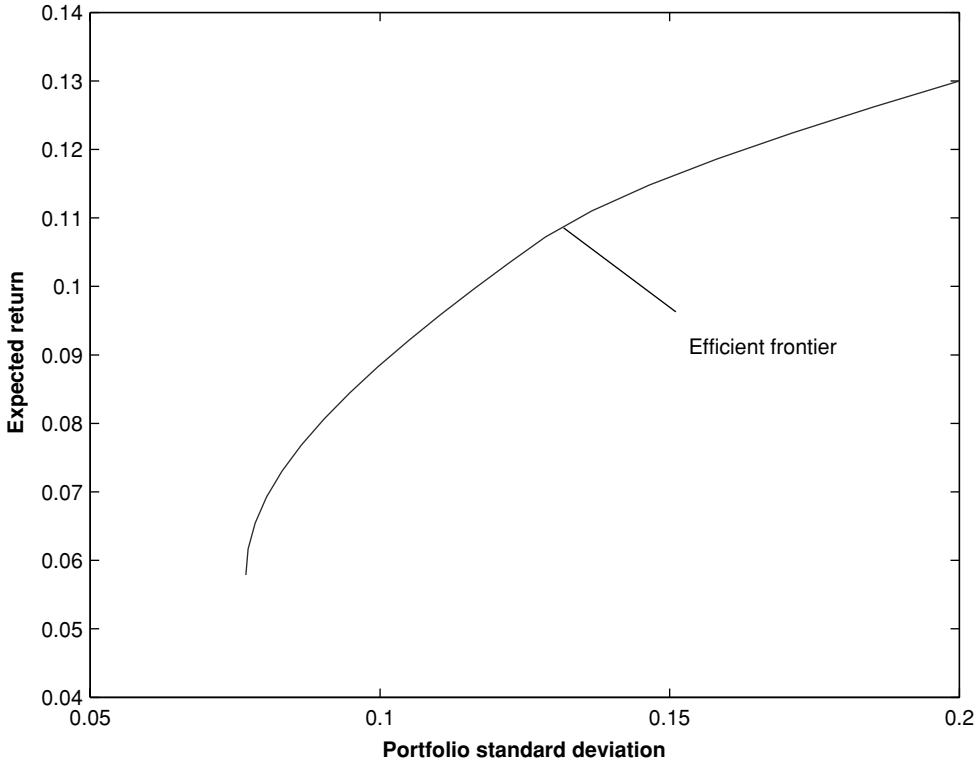

Figure 2.2 The mean–variance efficient frontier without a risk-free asset

expected return as 'good' and a higher standard deviation of returns (i.e., in this context, higher risk) as 'bad', the investor wants to achieve the highest possible expected return for any given level of risk; or equivalently, wishes to minimise the level of risk associated with any given expected return. This implies that the investor will choose some point along the upper edge of the feasible region, known as the efficient frontier. The point chosen will depend on their risk-expected return preferences (or utility or preference function): an investor who is more risk-averse will choose a point on the efficient frontier with a low risk and a low expected return, and an investor who is less risk-averse will choose a point on the efficient frontier with a higher risk and a higher expected return.

Figure 2.2 is one example of the mean–variance approach. However, the mean–variance approach is often presented in a slightly different form. If we assume a risk-free asset and (for simplicity) assume there are no short-selling constraints of any kind, then the attainable set of outcomes can be expanded considerably – and this means a considerable improvement in the efficient frontier. Given a risk-free rate equal to 4.5% in Figure 2.3, the investor can now achieve any point along a straight line running from the risk-free rate through to, and beyond, a point or portfolio *m* just touching the top of the earlier attainable set. *m* is also shown in the figure, and is often identified with the 'market portfolio' of the CAPM. As the figure also shows, the investor now faces an expanded choice set (and can typically achieve a higher expected return for any given level of risk).

So the mean–variance framework gives us a nice approach to the twin problems of how to measure risks *and* how to choose between risky alternatives. On the former question, our primary concern for the moment, it tells us that we can measure risk by the standard deviation of

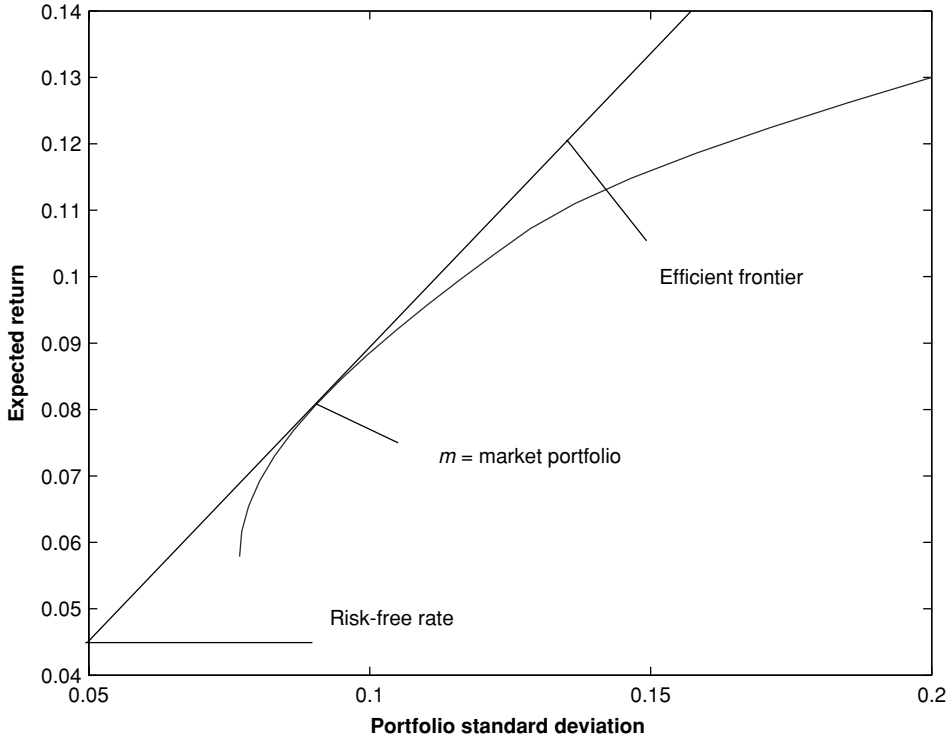

Figure 2.3 The mean–variance efficient frontier with a risk-free asset

returns. Indeed, it goes further and tells us that the standard deviation of returns is in many ways an *ideal* risk measure in circumstances where risks are normally (or elliptically) distributed.

However, the standard deviation can be a very unsatisfactory risk measure when we are dealing with seriously non-normal distributions. Any risk measure at its most basic level involves an attempt to capture or summarise the shape of an underlying density function, and although the standard deviation does that very well for a normal (and up to a point, more general elliptical) distribution, it does not do so for others. Recall that any statistical distribution can be described in terms of its moments or moment-based parameters such as mean, standard deviation, skewness and kurtosis. In the case of the normal distribution, the mean and standard deviation can be anything (subject only to the constraint that the standard deviation can never be negative), and the skewness and kurtosis are 0 and 3. However, other distributions can have quite different skewness and/or kurtosis, and therefore have quite different shapes than the normal distribution, and this is true even if they have the same mean and standard deviation.

To illustrate this for the skewness, Figure 2.4 compares a normal distribution with a skewed one (which is in fact a Gumbel distribution). The parameters of these are chosen so that both distributions have the same mean and standard deviation. As we can see, the skew alters the whole distribution, and tends to pull one tail in while pushing the other tail out. A portfolio theory approach would suggest that these distributions have equal risks, because they have equal standard deviations, and yet we can see clearly that the distributions (and hence the 'true'

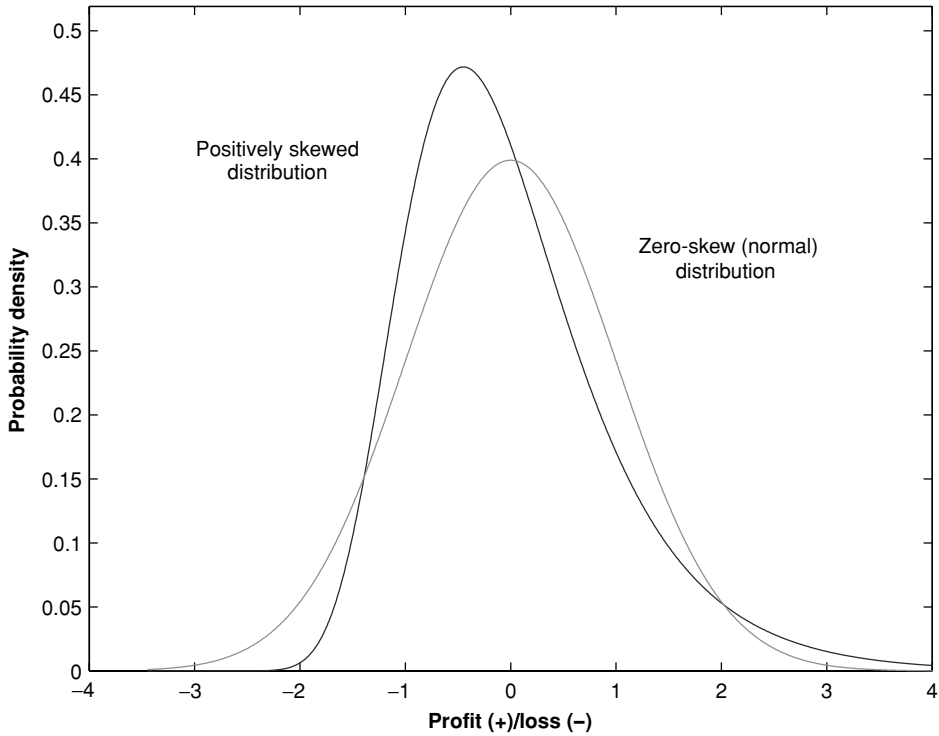

Figure 2.4 A skewed distribution
Note: The symmetric distribution is standard normal, and the skewed distribution is a Gumbel with location and scale equal to $-0.57722\sqrt{6}/\pi$ and $\sqrt{6}/\pi$.

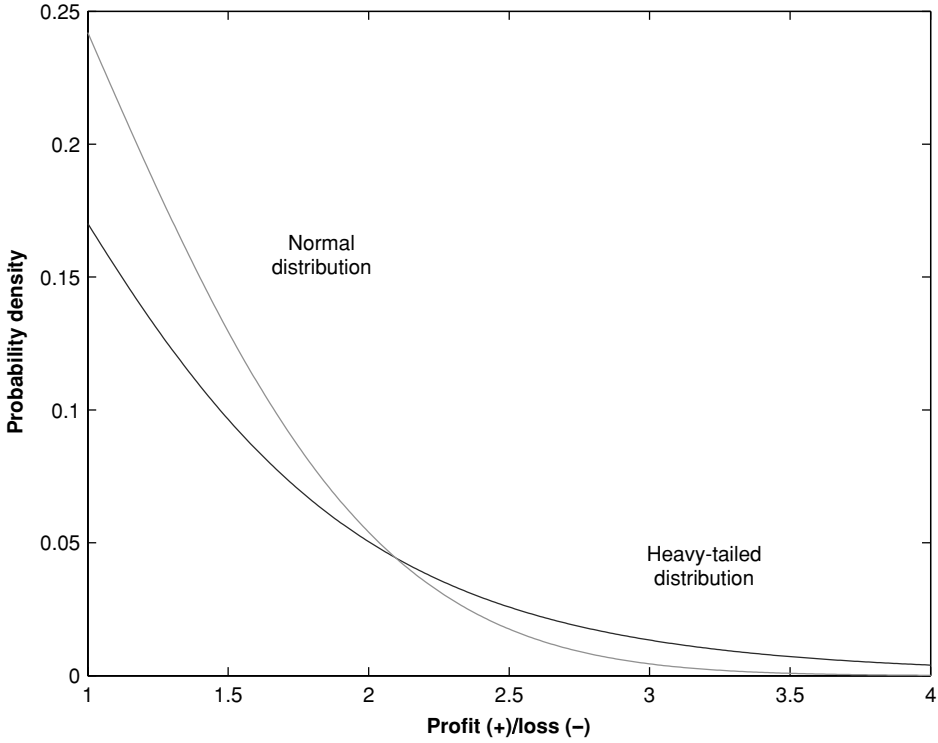

Figure 2.5 A heavy-tailed distribution
Note: The symmetric distribution is standard normal, and the heavy-tailed distribution is a t with mean 0, std 1 and 5 degrees of freedom.

risks, whatever they might be) must be quite different. The implication is that the presence of skewness makes portfolio theory unreliable, because it undermines the normality assumption on which it is (archetypically) based.

To illustrate this point for excess kurtosis, Figure 2.5 compares a normal distribution with a heavy-tailed one (i.e., a t distribution with 5 degrees of freedom). Again, the parameters are chosen so that both distributions have the same mean and standard deviation. As the name suggests, the heavy-tailed distribution has a longer tail, with much more mass in the extreme tail region. Tail heaviness – kurtosis in excess of 3 – means that we are more likely to lose (or gain) a lot, and these losses (or gains) will tend to be larger, relative to normality. A portfolio theory approach would again suggest that these distributions have equal risks, so the presence of excess kurtosis can also make portfolio theory unreliable.

Thus, the normality assumption is only strictly appropriate if we are dealing with a symmetric (i.e., zero-skew) distribution with a kurtosis of 3. If our distribution is skewed or has heavier tails – as is typically the case with financial returns – then the normality assumption is inappropriate and the mean–variance framework can produce misleading estimates of risk. This said, more general elliptical distributions share many of the features of normality and with suitable reparameterisations can be tweaked into a mean–variance framework. The mean–variance framework can also be (and often is) applied conditionally, rather than unconditionally, meaning that it might be applied conditionally on sets of parameters that might themselves

be random. Actual returns would then typically be quite non-normal (and often skewed and heavy tailed) because they are affected by the randomness of the parameters as well as by the randomness of the conditional elliptical distribution. But even with their greater flexibility, it is still doubtful whether conditionally elliptical distributions can give sufficiently good 'fits' to many empirical return processes. And, there again, we can use the mean–variance framework more or less regardless of the underlying distribution if the user's utility (or preference) function is a quadratic function that depends only on the mean and variance of return (i.e., so the user only *cares* about mean and standard deviation). However, such a utility function has undesirable properties and would itself be difficult to justify.

So the bottom line is that the mean–variance framework tells us to use the standard deviation (or some function of it) as our risk measure, but even with refinements such as conditionality, this is justified only in limited cases (discussed elsewhere), which are often too restrictive for many of the empirical distributions we are likely to meet.

Box 2.1 Traditional Dispersion Risk Measures

There are a number of traditional measures of risk based on alternative measures of dispersion. The most widely used is the standard deviation (or its square, the variance), but the standard deviation has been criticised for the arbitrary way in which deviations from the mean are squared and for giving equal treatment to upside and downside outcomes. If we are concerned about these, we can use the mean absolute deviation or the downside semi-variance instead: the former replaces the squared deviations in the standard deviation formula with absolute deviations and gets rid of the square root operation; the latter can be obtained from the variance formula by replacing upside values (i.e., observations above the mean) with zeros. We can also replace the standard deviation with other simple dispersion measures such as the entropy measure or the Gini coefficient.

A more general approach to dispersion is provided by a Fishburn (or lower partial moment) measure, defined as $\int_{-\infty}^{t}(t-x)^{\alpha} f(x)dx$. This measure is defined on two parameters: α, which describes our attitude to risk (and which is not to be confused with the confidence level!), and t, which specifies the cut-off between the downside that we worry about and the upside that we don't worry about. Many risk measures are special cases of the Fishburn measure or are closely related to it. These include: the downside semi-variance, which is very closely related to the Fishburn measure with $\alpha = 2$ and t equal to the mean; Roy's safety-first criterion, which corresponds to the Fishburn measure where $\alpha \to 0$; and the expected shortfall (ES), which is a multiple of the Fishburn measure with $\alpha = 1$. In addition, the Fishburn measure encompasses the stochastic dominance rules that are sometimes used for ranking risky alternatives:[3] the Fishburn measure with $\alpha = n + 1$ is proportional to the nth order distribution function, so ranking risks by this Fishburn measure is equivalent to ranking by nth order stochastic dominance.

[3] An nth order distribution function is defined as $F^{(n)}(x) = 1/(n-1)! \int_{-\infty}^{x} (x-u)^{n-1} f(u)du$, and X_1 is said to be nth order stochastically dominant over X_2 if $F_1^{(n)}(x) \leq F_2^{(n)}(x)$, where $F_1^{(n)}(x)$ and $F_2^{(n)}(x)$ are the nth degree distribution functions of X_1 and X_2(see Yoshiba and Yamai (2001, p. 8)). First-order stochastic dominance implies that the distribution function for X_1 is never above the distribution function for X_2, second-order stochastic dominance implies that their second-degree distribution functions do not cross, and so on. Since a risk measure with nth degree stochastic dominance is also consistent with lower degrees of stochastic dominance, first-order stochastic dominance implies second and higher orders of stochastic dominance, but not the reverse. First-order stochastic dominance is a very implausible condition that will hardly ever hold (as it implies that one distribution always gives higher values than the other, in which case choosing between the two is trivial), second-order stochastic dominance is less unreasonable, but will often not hold; third-order stochastic dominance is more plausible, and so on: higher orders of stochastic dominance are more plausible than lower orders of stochastic dominance.

2.2 VALUE AT RISK

2.2.1 Basics of VaR[4]

We turn now to our second framework. As we have seen already, the mean–variance framework works well with elliptical distributions, but is not reliable where we have serious non-normality. We therefore seek an alternative framework that will give us risk measures that are valid in the face of more general distributions. We now allow the P/L or return distribution to be less restricted, but focus on the tail of that distribution – the worst $p\%$ of outcomes, and this brings us back to the VaR. More formally, if we have a confidence level α and set $p = 1 - \alpha$, and if q_p is the p-quantile of a portfolio's prospective profit/loss (P/L) over some holding period, then the VaR of the portfolio at that confidence level and holding period is equal to:

$$VaR = -q_p \tag{2.3}$$

The VaR is simply the negative of the q_p quantile of the P/L distribution.[5] Thus, the VaR is defined contingent on two arbitrarily chosen parameters – a confidence level α, which indicates the likelihood that we will get an outcome no worse than our VaR, and which might be any value between 0 and 1; and a holding or horizon period, which is the period of time until we measure our portfolio profit or loss, and which might be a day, a week, a month, or whatever.

Some VaRs are illustrated in Figure 2.6, which shows a common probability density function (pdf) of profit/loss over a chosen holding period.[6] Positive values correspond to profits, and negative observations to losses, and positive values will typically be more common than negative ones. If $\alpha = 0.95$, the VaR is given by the negative of the point on the x-axis that cuts off the top 95% of P/L observations from the bottom 5% of tail observations. In this case, the relevant x-axis value (or quantile) is –1.645, so the VaR is 1.645. The negative P/L value corresponds to a positive VaR, indicating that the worst outcome at this level of confidence is a loss of 1.645.[7] Let us refer to this VaR as the 95% VaR for convenience. Alternatively, we could set $\alpha = 0.99$ and in this case the VaR would be the negative of the cut-off between the bottom 1% tail and everything else. The 99% VaR here is 2.326.

Since the VaR is contingent on the choice of confidence level, Figure 2.6 suggests that it will usually increase when the confidence level changes.[8] This point is further illustrated in the next

[4] The roots of the VaR risk measure go back to Baumol (1963, p. 174), who suggested a risk measure equal to $\mu + k\sigma$, where μ and σ are the mean and standard deviation of the distribution concerned, and k is a subjective confidence-level parameter that reflects the user's attitude to risk. As we shall see, this risk measure is comparable to the VaR under the assumption that P/L is normal or t distributed.

[5] It is obvious from the figure that the VaR is unambiguously defined when dealing with a continuous P/L distribution. However, the VaR can be ambiguous when the P/L distribution is discontinuous (e.g., as it might be if the P/L distribution is based on historical experience). To see this, suppose there is a gap between the lowest 5% of the probability mass on the left of a figure otherwise similar to Figure 2.4, and the remaining 95% on the right. In this case, the VaR could be the negative of any value between the left-hand side of the 95% mass and the right-hand side of the 5% mass: discontinuities can make the VaR ambiguous. However, in practice, this issue boils down to one of approximation, and won't make much difference to our results given any reasonable sample size.

[6] The figure is constructed on the assumption that P/L is normally distributed with mean 0 and standard deviation 1 over a holding period of 1 day.

[7] In practice, the point on the x-axis corresponding to our VaR will usually be negative and, where it is, will correspond to a (positive) loss and a positive VaR. However, this x-point can sometimes be positive, in which case it indicates a profit rather than a loss and, hence, a negative VaR. This also makes sense: if the worst outcome at this confidence level is a particular profit rather than a loss, then the VaR, the likely loss, must be negative.

[8] Strictly speaking, the VaR is non-decreasing with the confidence level, which means that the VaR can sometimes remain the same as the confidence level rises. However, the VaR cannot fall as the confidence level rises.

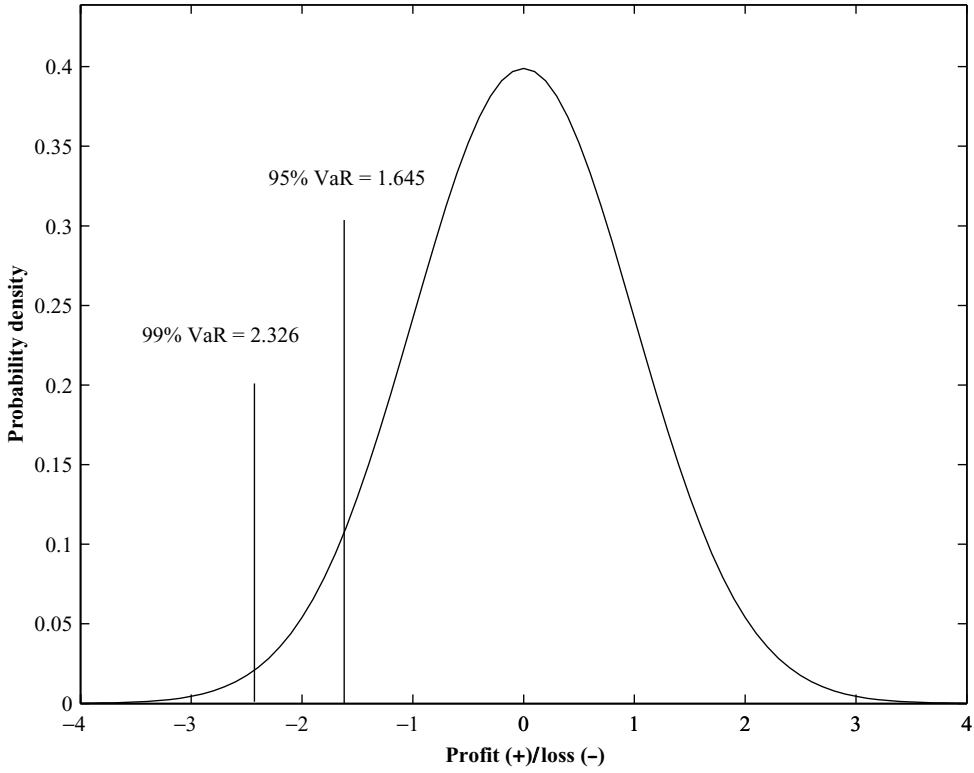

Figure 2.6 Value at risk
Note: Produced using the 'normalvarfigure' function.

figure (Figure 2.7), which shows how the VaR varies as we change the confidence level. In this particular case, which is also quite common in practice, the VaR not only rises with the confidence level, but also rises at an increasing rate – a point that risk managers might care to note.

As the VaR is also contingent on the holding period, we should consider how the VaR varies with the holding period as well. This behaviour is illustrated in Figure 2.8, which plots 95% VaRs based on two alternative μ values against a holding period that varies from 1 day to 100 days. With $\mu = 0$, the VaR rises with the square root of the holding period, but with $\mu > 0$, the VaR rises at a lower rate and would in fact eventually turn down. Thus, the VaR varies with the holding period, and the way it varies with the holding period depends significantly on the μ parameter.

Of course, each of the last two figures only gives a partial view of the relationship between the VaR and the parameters on which it depends: the first takes the holding period as given and varies the confidence level, and the second varies the holding period while taking the confidence level as given. To form a more complete picture, we need to see how VaR changes as we allow both parameters to change. The result is a VaR surface – as shown in Figure 2.9, based here on a hypothetical assumption that $\mu = 0$ – that enables us to read off the VaR for any given combination of these two parameters. The shape of the VaR surface

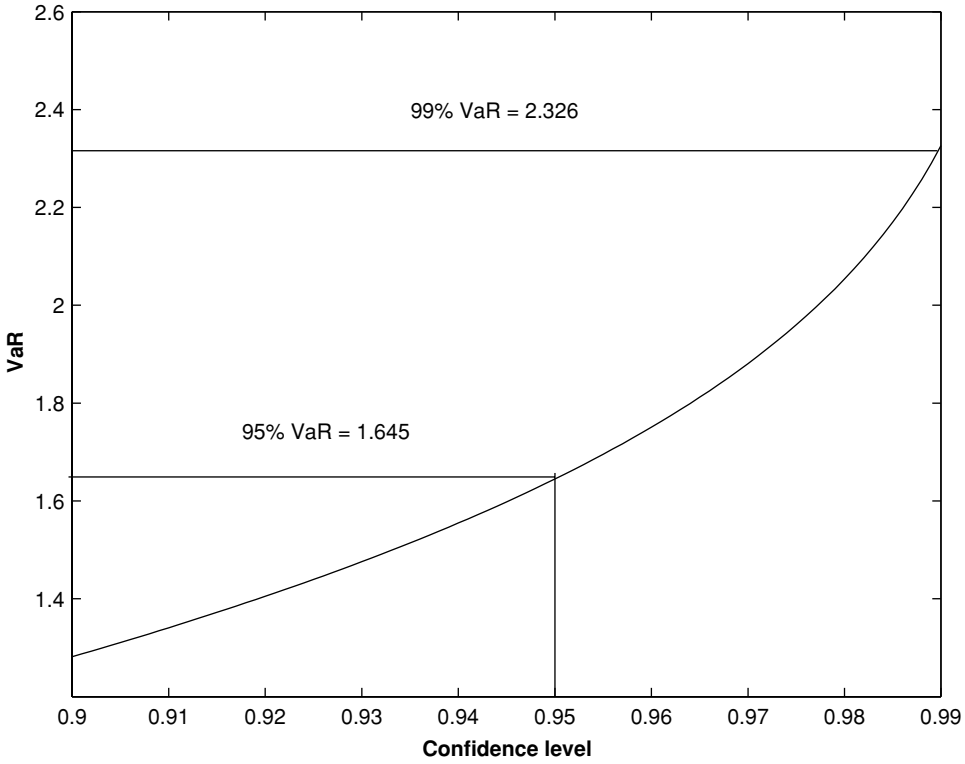

Figure 2.7 VaR and confidence level
Note: Produced using the 'normalvarplot2D_cl' function.

shows how VaR changes as underlying parameters change, and conveys a great deal of risk information. In this case, which is typical of many, the surface rises with both confidence level and holding period to culminate in a spike – indicating where our portfolio is most vulnerable – as both parameters approach their maximum values.

2.2.2 Determination of the VaR Parameters

The use of VaR involves two arbitrarily chosen parameters – the confidence level and the holding period – but how do we choose the values of these parameters?

The choice of confidence level depends on the purposes to which our risk measures are put. For example, we would want a high confidence level if we were using our risk measures to set firmwide capital requirements, but for backtesting, we often want lower confidence levels to get a reasonable proportion of excess-loss observations. The same goes if we were using VaR to set risk limits: many institutions prefer to use confidence levels in the region of 95% to 99%, as this is likely to produce a small number of excess losses and so force the people concerned to take the limit seriously. And when using VaRs for reporting or comparison purposes, we would probably wish to use confidence levels that are comparable to those used

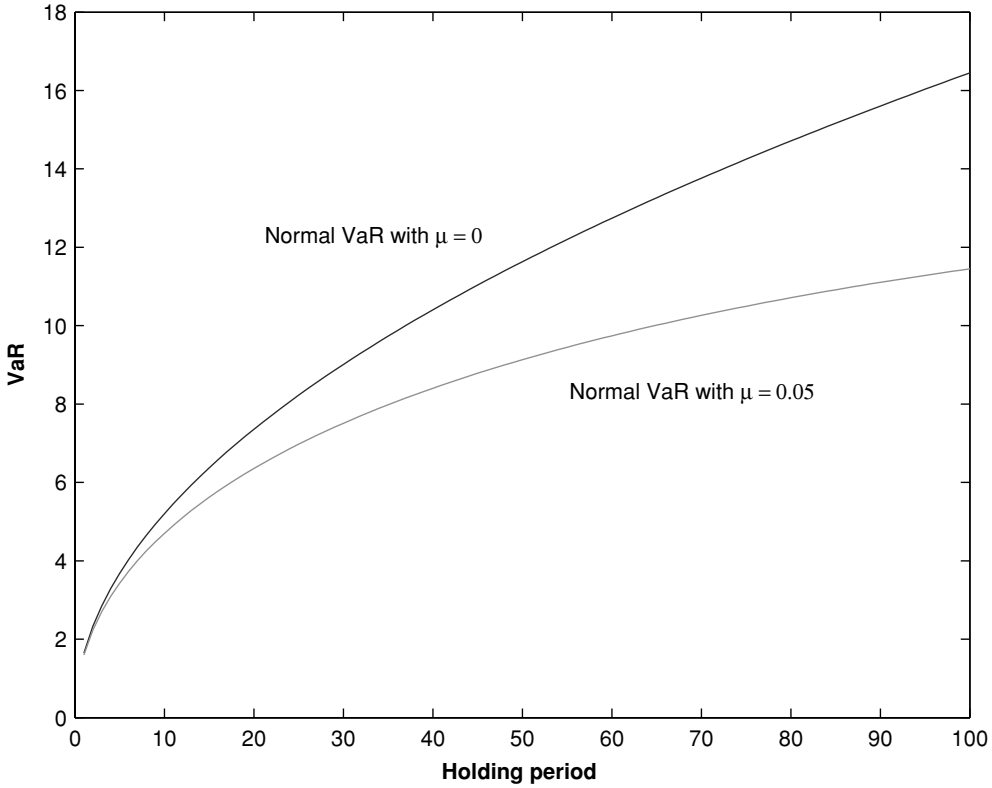

Figure 2.8 VaR and holding period

for similar purposes by other institutions, which are again typically in the range from 95% to 99%.

The usual holding periods are one day or one month, but institutions can also operate on other holding periods (e.g., one quarter or more), depending on their investment and/or reporting horizons. The holding period can also depend on the liquidity of the markets in which an institution operates: other things being equal, the ideal holding period appropriate in any given market is the length of time it takes to ensure orderly liquidation of positions in that market. The holding period might also be specified by regulation: for example, BIS capital adequacy rules stipulate that banks should operate with a holding period of two weeks (or 10 business days). The choice of holding period can also depend on other factors:

- The assumption that the portfolio does not change over the holding period is more easily defended with a shorter holding period.
- A short holding period is preferable for model validation or backtesting purposes: reliable validation requires a large data set, and a large data set requires a short holding period.

Thus, the 'best' choice of these parameters depends on the context. However, it is a good idea to work with ranges of parameter values rather than particular point values: a VaR surface is much more informative than a single VaR number.

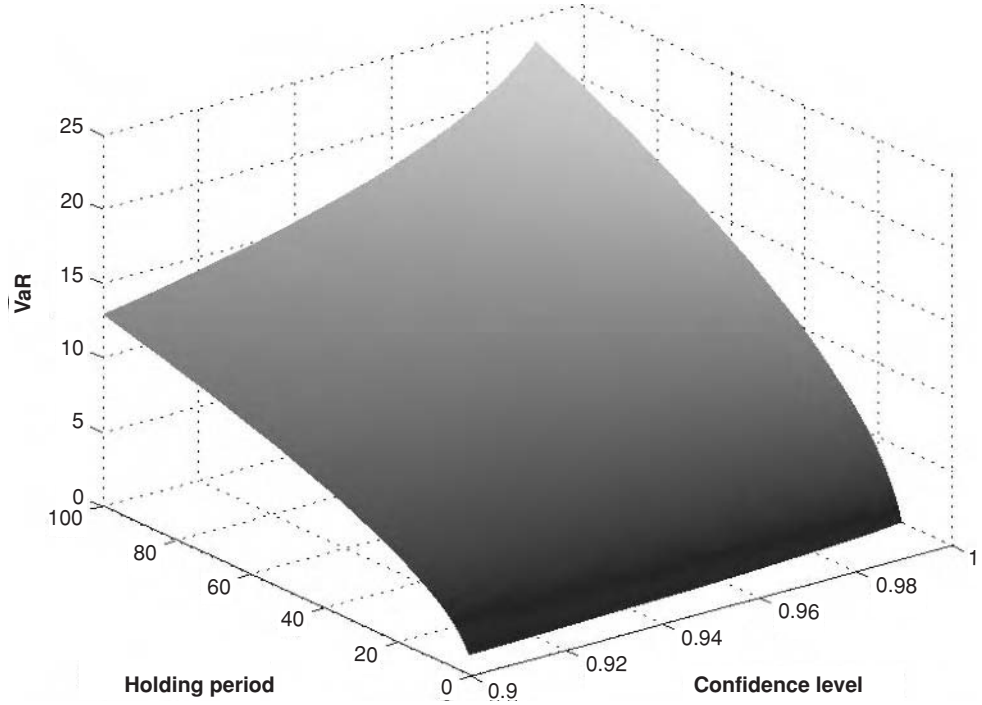

Figure 2.9 A VaR surface
Note: Produced using the 'normalvarplot3D' function. This plot is based on illustrative assumptions that $\mu = 0$ and $\sigma = 1$.

2.2.3 Limitations of VaR as a Risk Measure

We discussed some of the advantages of VaR – the fact that it is a common, holistic, probabilistic risk measure, etc. – in the last chapter. However, the VaR also has its drawbacks. Some of these we have met before – that VaR estimates can be subject to error, that VaR systems can be subject to model risk (i.e., the risk of errors arising from models being based on incorrect assumptions) or implementation risk (i.e., the risk of errors arising from the way in which systems are implemented). On the other hand, such problems are common to many if not all risk measurement systems, and are not unique to VaR ones.

Yet the VaR also has its own distinctive limitations as a risk measure. One important limitation is that the VaR only tells us the most we can lose if a tail event does *not* occur (e.g., it tells us the most we can lose 95% of the time); if a tail event *does* occur, we can expect to lose more than the VaR, but the VaR itself gives us no indication of how much that might be. The failure of VaR to take account of the magnitude of losses in excess of itself implies that two positions can have the same VaR – and therefore appear to have the same risk if we use the VaR to measure risk – and yet have very different risk exposures.

This can lead to some very undesirable outcomes. For instance, if a prospective investment has a high expected return but also involves the possibility of a very high loss, a VaR-based decision calculus might suggest that the investor should go ahead with the investment if the higher loss does not affect the VaR (i.e. because it exceeds the VaR), regardless of the size of

the higher expected return and regardless of the size of the possible loss. Such a categorical acceptance of any investment that increases expected return – regardless of the possible loss, provided only that it is insufficiently probable – undermines sensible risk–return analysis, and can leave the investor exposed to very high losses.[9]

If the VaR can lead an investor working on his/her own behalf to make perverse decisions, it creates even more scope for problems when there are principal–agent (or delegation) issues. This would be the case where decision-making is decentralised and traders or asset managers work to VaR-defined risk targets or remuneration packages. The classic example is where traders who face a VaR-defined risk target have an incentive to sell out-of-the-money options that lead to higher income in most states of the world and the occasional large hit when the firm is unlucky. If the options are suitably chosen, the bad outcomes will have probabilities low enough to ensure that there is no effect on the VaR, and the trader benefits from the higher income (and hence higher bonuses) earned in 'normal' times when the options expire out of the money. Thus the fact that VaR does not take account of what happens in 'bad' states can distort incentives and encourage traders to 'game' a VaR target (and/or a VaR-defined remuneration package) and promote their own interests at the expense of the interests of the institutions that employ them.[10]

So the VaR has a number of serious limitations as a risk measure, and we will have more to say on some of these presently. There are some nice ironies here. We have seen that there is one important class of distributions where VaR is in many ways a very good measure of risk: these distributions are of course the elliptical distributions. In such circumstances the VaR works well, but in such circumstances we do not really need it: the VaR is then merely a simple transformation of the standard deviation, and a VaR framework tells us nothing that we could not have found out from a basic mean–variance framework. Thus, in the face of elliptical distributions, the mean–variance framework works well and the value of upgrading to a VaR framework is negligible. Yet the whole point of upgrading from the mean–variance framework to something more general is to be able to measure the risks associated with seriously non-normal distributions. The VaR enables us to do this, but it is in exactly these circumstances that the VaR is not a reliable (and perhaps not even useful) risk measure. The bottom line is a delightful irony: *where the VaR is reliable, we don't need it; and where we do need it, it isn't reliable*. We therefore need an alternative framework that can give us useful risk measures in a seriously non-normal environment.

2.3 COHERENT RISK MEASURES

2.3.1 The Coherence Axioms and Their Implications

We therefore turn to our third risk measurement paradigm: the theory of coherent risk measures proposed by Artzner *et al.* (1997, 1999). This approach provides the first formal (i.e.,

[9] To elaborate further: a VaR-based risk–return analysis only makes intuitive sense if returns are elliptically distributed. If returns are non-elliptical, then a VaR-based risk–return analysis is inconsistent with classical (von Neumann–Morgenstern) expected utility theory. Indeed, it appears that unless we assume ellipticality (which we usually cannot) then a VaR-based risk–return analysis can only be justified if preferences are quadratic (i.e., more specifically, if agents don't care about higher moments, which is weird) or lexicographic, and lexicographic preferences are highly implausible because they allow no substitutability in utility between risk and expected return. (For more on some of these issues, see Grootveld and Hallerbach (2004).) A VaR-based risk–return analysis can only be justified under conditions that are empirically usually too restrictive and/or a priori implausible.

[10] We can sometimes ameliorate these problems by using more VaR information. For example, the trader who spikes his firm might be detected if the VaR of his position is estimated at a higher confidence level as well. A (partial) solution to our earlier problems is, therefore, to look at more than one point on the VaR–confidence level curve and not just to look at a single VaR figure. However, such 'solutions' are often not practically feasible and, in any case, fail to address the root problem.

mathematically grounded) theory of financial risk. Their starting point is a simple but profound one: that although we all have an intuitive sense of what financial risk entails, it is difficult to give a quantitative assessment of financial risk unless we specify what we actually mean by a measure of risk. For example, we all have a vague notion of temperature, but it is hard to conceptualise it clearly without the notion of a thermometer, which tells us how temperature should be measured. In much the same way, the notion of risk itself is hard to conceptualise without a clear idea of what we mean by a measure of risk. To clarify these issues, Artzner *et al.* postulated a set of axioms – the axioms of coherency – and began to work out their implications.

Let X and Y represent any two portfolios' P/L (or future values, or more loosely, the two portfolios themselves), and let $\rho(.)$ be a measure of risk over a chosen horizon.[11] The risk measure $\rho(.)$ is said to be coherent if it satisfies the following properties:

(i) *Monotonicity*: $Y \geq X \Rightarrow \rho(Y) \leq \rho(X)$.
(ii) *Subadditivity*: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.
(iii) *Positive homogeneity*: $\rho(hX) = h\rho(X)$ for $h > 0$.
(iv) *Translational invariance*: $\rho(X + n) = \rho(X) - n$ for some certain amount n.

Properties (i), (iii) and (iv) are essentially 'well-behavedness' conditions intended to rule out awkward outcomes.[12]

The most important property is (ii), subadditivity. This tells us that a portfolio made up of subportfolios will risk an amount which is no more than, and in some cases less than, the sum of the risks of the constituent subportfolios. Subadditivity is the most important criterion we would expect a 'reasonable' risk measure to satisfy. It reflects an expectation that when we aggregate individual risks, they diversify or, at worst, do not increase: the risk of the sum is always less than or equal to the sum of the risks. Subadditivity means that aggregating risks does not increase overall risk.[13]

Subadditivity is more than just a matter of theoretical 'tidiness' and has important practical implications. For example, non-subadditivity is treacherous because it suggests that diversification might be a bad thing, which would suggest the laughable conclusion that putting all your eggs into one basket might be good risk management practice! It also means that in adding risks together we might create an extra 'residual' risk that someone has to bear, and that didn't exist before. This would have some awkward consequences:

- Non-subadditive risk measures can tempt agents trading on an organised exchange to break up their accounts, with separate accounts for separate risks, in order to reduce their margin requirements. This would concern the exchange because the margin requirements on the

[11] At a deeper level, we can also start with the notion of an acceptance set, the set of all positions acceptable to some stakeholder (e.g., a financial regulator). We can then interpret the risk measure $\rho(.)$ as the minimum extra cash that has to be added to the risky position and invested prudently in some reference asset to make the risky position acceptable. If $\rho(.)$ is negative, its negativity can be interpreted as the maximum amount that can be safely withdrawn, and still leave the position acceptable.

[12] The other conditions can be understood from the last footnote. Monotonicity means that a random cash flow or future value Y that is always greater than X should have a lower risk: this makes sense, because it means that less has to be added to Y than to X to make it acceptable, and the amount to be added is the risk measure. Positive homogeneity implies that the risk of a position is proportional to its scale or size, and makes sense if we are dealing with liquid positions in marketable instruments. Translational invariance requires that the addition of a sure amount reduces *pari passu* the cash needed to make our position acceptable, and is obviously valid when one appreciates that the cash needed is our risk measure.

[13] However, the coherence axioms can run into a problem relating to liquidity risk. If a position is 'large' relative to the market, then doubling the size of this position can more than double the risk of the position, because bid prices will depend on the position size. This raises the possibility of liquidity-driven violations of homogeneity and subadditivity. Perhaps the best way to resolve this difficulty, suggested by Acerbi (2004, p. 150), is to add a liquidity charge to a coherent risk measure. This charge would take account of relative size effects, but also have the property of going to zero as size/illiquidity effects become negligible.

separate accounts would no longer cover the combined risks, and so leave the exchange itself exposed to possible loss.

- If regulators use non-subadditive risk measures to set capital requirements, then a financial firm might be tempted to break itself up to reduce its regulatory capital requirements, because the sum of the capital requirements of the smaller units would be less than the capital requirement of the firm as a whole.
- If risks are subadditive, adding risks together would give us an overestimate of combined risk, and this means that we can use the sum of risks as a conservative estimate of combined risk. This facilitates decentralised decision-making within a firm, because a supervisor can always use the sum of the risks of the units reporting to him or her as a conservative back-of-the-envelope risk measure. But if risks are not subadditive, adding them together gives us an underestimate of combined risks, which makes the sum of risks treacherous and therefore effectively useless as a back-of-the-envelope measure.

The bottom line is that subadditivity matters.

This spells trouble for the VaR, because VaR is not subadditive. Recall that for a risk measure to be subadditive, the subadditivity condition $\rho(X + Y) \leq \rho(X) + \rho(Y)$ must apply for *all* possible X and Y. We can therefore prove that VaR is not subadditive if we can find a *single* counter-example where VaR violates this condition. Now consider the following:

> We have two identical bonds, A and B. Each defaults with probability 4%, and we get a loss of 100 if default occurs, and a loss of 0 if no default occurs. The 95% VaR of each bond is therefore 0, so $\text{VaR}(A) = \text{VaR}(B) = \text{VaR}(A) + \text{VaR}(B) = 0$. Now suppose that defaults are independent. Elementary calculations then establish that we get a loss of 0 with probability $0.96^2 = 0.9216$, a loss of 200 with probability $0.04^2 = 0.0016$, and a loss of 100 with probability $1 - 0.9216 - 0.0016 = 0.0768$. Hence $\text{VaR}(A + B) = 100$. Thus, $\text{VaR}(A + B) = 100 > 0 = \text{VaR}(A) + \text{VaR}(B)$, and the VaR violates subadditivity. Hence, the VaR is not subadditive. QED

We can only 'make' the VaR subadditive if we impose restrictions on the form of the P/L distribution. It turns out, in fact, that we can only 'make' the VaR subadditive by imposing the severe restriction that the P/L distribution is elliptically distributed,[14] and this is of limited consolation because in the real world non-elliptical distributions are the norm rather than the exception.

The failure of VaR to be subadditive is a fundamental problem because it means that VaR has no claim to be regarded as a 'proper' risk measure at all. A VaR is merely a quantile. It has its uses as a quantile, but it is very unsatisfactory as a risk measure. There is also a deeper problem:

> from an epistemologic point of view the main problem with VaR is not its lack of subadditivity, but rather the very fact that no set of axioms for a risk measure and therefore no unambiguous definition of financial risk has ever been associated with this statistic. So, despite the fact that some VaR supporters still claim that subadditivity is not a necessary axiom, none of them, to the best of our knowledge, has ever tried to write an alternative meaningful and consistent set of axioms for a risk measure which are fulfilled also by VaR.[15]

Given these problems with the VaR, we seek alternative, coherent, risk measures that retain the benefits of the VaR – in terms of providing a common, aggregative, holistic, etc. measure of risk – while avoiding its drawbacks. If they are to retain the benefits of the VaR, we might

[14] Artzner *et al.* (1999), p. 217.
[15] Acerbi (2004), p. 150.

also expect that any such risk measures will be 'VaR-like' in the sense that they will reflect the quantiles of the P/L or loss distribution, but will be non-trivial functions of those quantiles rather than a single 'raw' quantile taken on its own.

2.3.2 The Expected Shortfall

A good candidate is the expected shortfall (ES). The ES is the average of the worst $100(1-\alpha)\%$ of losses:[16]

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_p dp. \tag{2.4}$$

If the loss distribution is discrete, then the ES is the discrete equivalent of Equation (2.4):

$$ES_\alpha = \frac{1}{1-\alpha}\sum_{p=0}^{\alpha}[p\text{th highest loss}] \times [\text{probability of } p\text{th highest loss}] \tag{2.5}$$

The subadditivity of ES follows naturally. If we have N equal-probability quantiles in a discrete P/L distribution, then:

$$ES_\alpha(X) + ES_\alpha(Y) \tag{2.6}$$

$= [\text{mean of } N\alpha \text{ highest losses of } X] + [\text{mean of } N\alpha \text{ highest losses of } Y]$
$\geq [\text{mean of } N\alpha \text{ highest losses of } (X+Y)]$
$= ES_\alpha(X+Y)$

A continuous loss distribution can be regarded as the limiting case as N gets large. In general, the mean of the $N\alpha$ worst cases of X and the mean of the $N\alpha$ worst cases of Y will be bigger than the mean of the $N\alpha$ worst cases of $(X+Y)$, except in the special case where the worst X and Y occur in the same $N\alpha$ events, and in this case the sum of the means will equal the mean of the sum. It is easy to show that the ES also satisfies the other properties of coherence, and is therefore coherent (Acerbi (2004, proposition 2.16)).

An illustrative ES is shown in Figure 2.10. If we express our data in loss terms, the VaR and ES are shown on the right-hand side of the figure: the VaR is 1.645 and the ES is 2.063. Both VaR and ES depend on the underlying parameters and distributional assumptions, and these particular figures are based on a 95% confidence level and 1-day holding period, and on the assumption that daily P/L is distributed as standard normal (i.e., with mean 0 and standard deviation 1).

[16] The ES is one of a family of closely related risk measures, members of which have been variously called the expected tail loss, tail conditional expectation (TCE), tail VaR, conditional VaR, tail conditional VaR and worst conditional expectation, as well as expected shortfall. Different writers have used these terms in inconsistent ways, and there is an urgent need to cut through the confusion created by all this inconsistent terminology and agree on some consensus nomenclature. This said, the substantive point is that this family of risk measures has two significant substantially distinct members. The first is the measure we have labelled the ES, as defined in Equation (2.4); this is defined in terms of a probability threshold. The other is its quantile-delimited cousin, most often labelled as the TCE, which is the average of losses exceeding VaR, i.e., $TCE_\alpha = -E[X|X > q_\alpha(X)]$. The ES and TCE will always coincide when the loss distribution is continuous, but the TCE can be ambiguous when the distribution is discrete, whereas the ES is always uniquely defined (see Acerbi (2004, p. 158)). We therefore ignore the TCE in what follows, because it is not an interesting statistic except where it coincides with the ES.

It is also interesting to note that the ES risk measure has been familiar to insurance practitioners for a long time: it is very similar to the measures of conditional average claim size that have long been used by casualty insurers. Insurers are also very familiar with the notion of the conditional coverage of a loss in excess of a threshold (e.g., in the context of reinsurance treaties). For more on ES and its precursors, see Artzner *et al.* (1999, pp. 223–224).

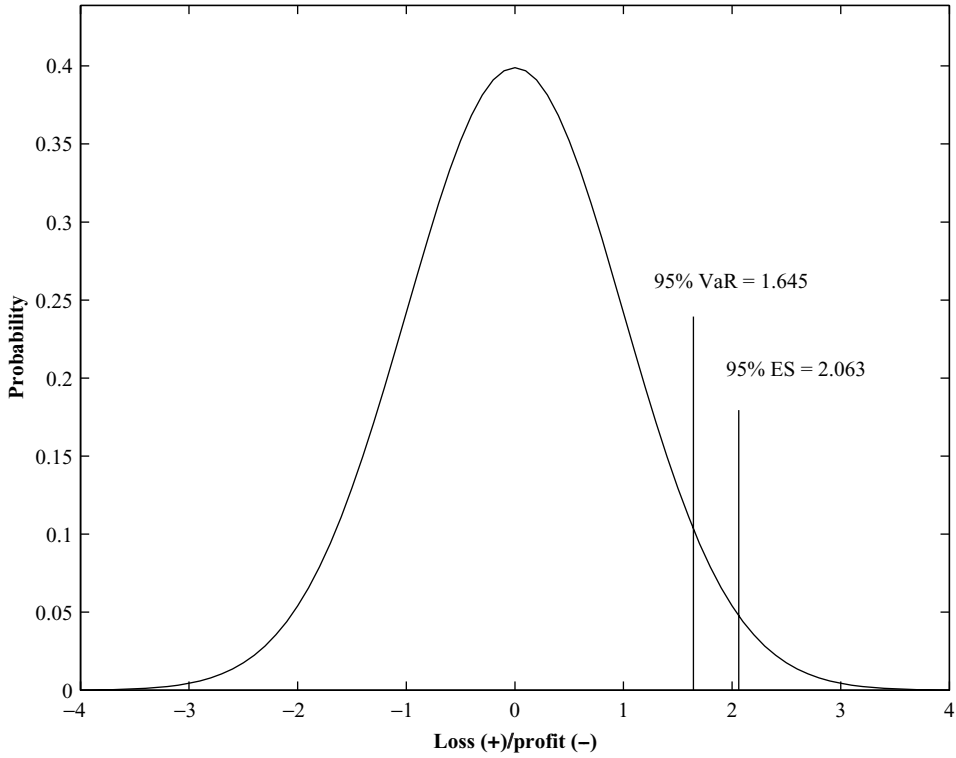

Figure 2.10 Expected shortfall
Note: Produced using the 'normalvaresfigure' function.

Since the ES is conditional on the same parameters as the VaR itself, it is immediately obvious that any given ES figure is only a point on an ES curve or ES surface. The ES–confidence level curve is shown in Figure 2.11. This curve is similar to the earlier VaR curve shown in Figure 2.7 and, like it, tends to rise with the confidence level. There is also an ES–holding period curve corresponding to the VaR–holding period curve shown in Figure 2.8.

There is also an ES surface, illustrated for the $\mu = 0$ case in Figure 2.12, which shows how ES changes as both confidence level and holding period change. In this case, as with its VaR equivalent in Figure 2.9, the surface rises with both confidence level and holding period, and spikes as both parameters approach their maximum values.

Like the VaR, the ES provides a common consistent risk measure across different positions, it takes account of correlations in a correct way, and so on. It also has many of the same uses as the VaR. However, the ES is also a better risk measure than the VaR for a number of reasons:

- The ES tells us what to expect in bad states – it gives an idea of how bad bad might be – while the VaR tells us nothing other than to expect a loss higher than the VaR itself.
- An ES-based risk-expected return decision rule is valid under more general conditions than a VaR-based risk-expected return decision rule: in particular, the ES-based rule is consistent with expected utility maximisation if risks are rankable by a second-order stochastic dominance rule, while a VaR-based rule is only consistent with expected utility

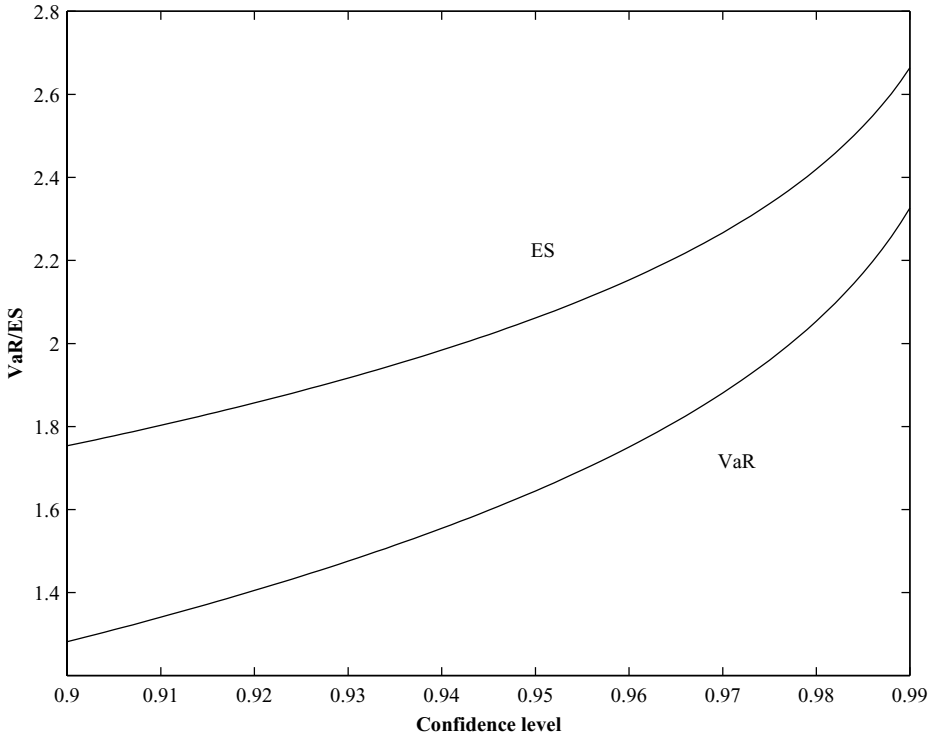

Figure 2.11 ES and the confidence level
Note: Produced using the 'normalvaresplot2D_cl' function.

maximisation if risks are rankable by a (much) more stringent first-order stochastic dominance rule.[17]

- Because it is coherent, the ES always satisfies subadditivity, while the VaR does not. The ES therefore has the various attractions of subadditivity, and the VaR does not.
- Finally, the subadditivity of ES implies that the portfolio risk surface will be convex, and convexity ensures that portfolio optimisation problems using ES measures, unlike ones that use VaR measures, will always have a unique well-behaved optimum.[18] In addition, this convexity ensures that portfolio optimisation problems with ES risk measures can be handled very efficiently using linear programming techniques.[19]

In short, the ES easily dominates the VaR as a risk measure.

2.3.3 Spectral Risk Measures

However, the ES is also rarely, if ever, the 'best' coherent risk measure. Going back to first principles, suppose we define more general risk measures M_ϕ that are weighted averages of

[17] See Yoshiba and Yamai (2001), pp. 21–22.
[18] See, e.g., Uryasev (2000) and Acerbi and Tasche (2002).
[19] See Rockafellar and Uryasev (2002) and Uryasev (2000).

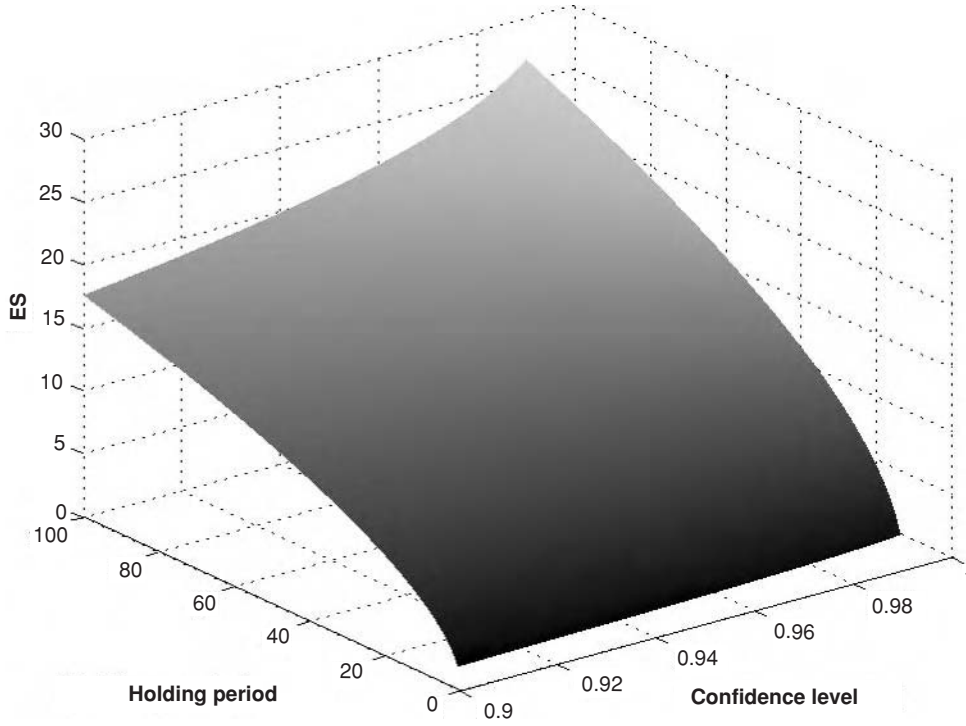

Figure 2.12 The ES surface
Note: Produced using the 'normalesplot3D' function. This plot is based on illustrative assumptions that $\mu = 0$ and $\sigma = 1$.

the quantiles of the loss distribution:

$$M_\phi = \int_0^1 \phi(p) q_p dp \tag{2.7}$$

where the weighting function $\phi(p)$ remains to be determined. This function is also known as the risk spectrum or risk-aversion function.

It is interesting to note that both the ES and the VaR are special cases of Equation (2.7). The ES is a special case of M_ϕ obtained by setting $\phi(p)$ to the following:

$$\phi(p) = \begin{cases} 0 & \text{if} \quad p < \alpha \\ 1/(1-\alpha) & \quad\quad p \geq \alpha \end{cases} \tag{2.8}$$

The ES gives tail losses an equal weight of $1/(1-\alpha)$, and other quantiles a weight of 0. The VaR is also a special case – albeit a highly degenerate one – of M_ϕ. Because the VaR is just a single quantile, the spectral risk measure is the VaR if $\phi(p)$ takes the form of a Dirac delta function, which assigns a probability 1 to the event $p = \alpha$, and a probability of 0 to $p \neq \alpha$. This is degenerate because it gives an infinite value to the pdf at $p = \alpha$ and a zero value to the pdf everywhere else. So one measure places equal weight on tail losses, and the other places no weight at all on them.

However, we are concerned for the moment with the broader class of coherent risk measures. In particular, we want to know the conditions that $\phi(p)$ must satisfy in order to make M_ϕ coherent. The answer is the class of (non-singular) spectral risk measures, in which $\phi(p)$ takes the following properties (Acerbi (2004, proposition 3.4)):[20]

- *Non-negativity:* $\phi(p) \geq 0$ for all p in the range [0,1].
- *Normalization*: $\int_0^1 \phi(p)dp = 1$.
- *Weakly increasing*: If some probability p_2 exceeds another probability p_1, then p_2 must have a weight bigger than or equal to that of p_1.

The first two conditions are fairly obvious as they require that weights should be positive and sum to 1. The critical condition is the third one. This condition reflects the risk-aversion, requiring that the weights attached to higher losses should be bigger than, or certainly no less than, the weights attached to lower losses. Given that it ensures coherence, this condition suggests that *the key to coherence is that a risk measure must give higher losses at least the same weight as lower losses*.

The weights attached to higher losses in spectral risk measures are thus a direct reflection of the user's risk-aversion. If a user has a 'well-behaved' risk-aversion function, then the weights will rise smoothly, and the rate at which the weights rise will be related to the degree of risk aversion: the more risk-averse the user, the more rapidly the weights will rise. This is exactly as it should be.

The connection between the $\phi(p)$ weights and risk-aversion sheds further light on the inadequacies of the ES and the VaR. We saw earlier that the ES is characterised by all losses in the tail region having the same weight. If we interpret the weights as reflecting the user's attitude toward risk, these weights imply that the user is risk-*neutral* between tail-region outcomes. Since we usually assume that agents are risk-*averse*, this would suggest that the ES is not, in general, a good risk measure to use, notwithstanding its coherence. If a user is risk-averse, it should have a weighting function that gives *higher* losses a *higher* weight.[21]

The implications for the VaR are much worse, and we can see that the VaR's inadequacies are related to its failure to satisfy the increasing-weight property. With the VaR, we give a large weight to the loss associated with a p-value equal to α, and we give a lower (indeed, zero) weight to any *greater* loss. The implication is that the user is actually risk-*loving* (i.e., has *negative* risk-aversion) in the tail loss region.[22] To make matters worse, since the weight

[20] Strictly speaking, the set of spectral risk measures is the convex hull (or set of all convex combinations) of all ESs for all α belonging to [0,1]. There is also an 'if and only if' connection here: a risk measure M_ϕ is coherent if and only if M_ϕ is spectral and $\phi(p)$ satisfies the conditions indicated in the text. Moreover, there is also a good argument that the spectral measures so defined are the only really interesting coherent risk measures. Acerbi (2004, pp. 180–182) goes on to show that all coherent risk measures that satisfy the two additional properties of comonotonic additivity and law invariance are also spectral measures. The former condition is that if two random variables X and Y are comonotonic (i.e., always move in the same direction), then $\rho(X + Y) = \rho(X) + \rho(Y)$; comonotonic additivity is an important aspect of subadditivity, and represents the limiting case where diversification has no effect. Law invariance boils down to the (for practical purposes essential) requirement that a measure is estimable from empirical data. Both conditions are very important, and coherent risk measures that do not satisfy them – that is to say, non-spectral coherent risk measures – are seriously questionable.

[21] The claim that the selection of the ES as the preferred risk measure indicates risk-neutrality is confirmed from the perspective of the downside risk or lower partial moment literature (see, e.g., Fishburn (1977)). The parameter α reflects the degree of risk aversion, and the user is risk-averse if $\alpha > 1$, risk-neutral if $\alpha = 1$, and risk-loving if $0 < \alpha < 1$. However, we would only choose the ES as our preferred risk measure if $\alpha = 1$ (Grootveld and Hallerbach (2004, p. 36)). Hence, the use of the ES implies that we are risk-neutral.

[22] Following on from the last footnote, the expected utility-downside risk literature also indicates that the VaR is the preferred risk measure if $\alpha = 0$. From the perspective of this framework, $\alpha = 0$ indicates an extreme form of risk-loving (Grootveld and Hallerbach (2004, p. 35)). Thus, two very different approaches both give the same conclusion that VaR is only an appropriate risk measure if preferences exhibit extreme degrees of risk-loving.

drops to zero, we are also talking about risk-loving of a very aggressive sort. The blunt fact is that with the VaR weighting function, we give a large weight to a loss equal to the VaR itself, and we don't *care at all* about any losses exceeding the VaR! It is therefore hardly surprising that the VaR has its problems.

To obtain a spectral risk measure, the user must specify a particular form for their risk-aversion function. This decision is subjective, but can be guided by the economic literature on utility-function theory. An example is an exponential risk-aversion function:

$$\phi_\gamma(p) = \frac{e^{-(1-p)/\gamma}}{\gamma(1 - e^{-1/\gamma})} \tag{2.9}$$

where $\gamma \in (0, \infty)$ reflects the user's degree of risk-aversion: a smaller γ reflecting a greater degree of risk-aversion. This function satisfies the conditions required of a spectral risk measure, but is also attractive because it is a simple function that depends on a single parameter γ, which gets smaller as the user becomes more risk-averse.

A spectral risk-aversion function is illustrated in Figure 2.13. This shows how the weights rise with the cumulative probability p, and the rate of increase depends on γ. The more risk-averse the user, the more rapidly the weights rise as losses increase.

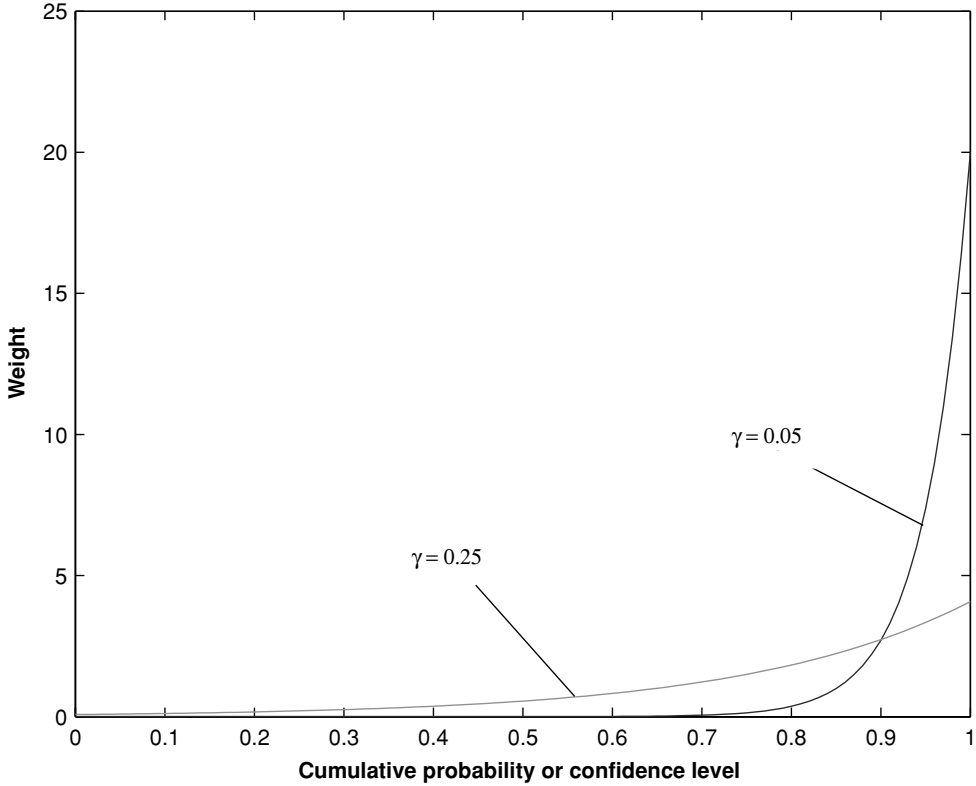

Figure 2.13 Exponential–spectral weights

To obtain our spectral measure M_ϕ using the exponential weighting function, we choose a value of γ and substitute $\phi(p)$ (or Equation (2.9)) into Equation (2.7) to get:

$$M_\phi = \int_0^1 \phi(p) q_p dp = \int_0^1 \frac{e^{-(1-p)/\gamma}}{\gamma(1 - e^{-1/\gamma})} q_p dp \tag{2.10}$$

The spectral–exponential measure is therefore a weighted average of quantiles, with the weights given by the exponential risk-aversion function (Equation (2.9)). It can be estimated using a suitable numerical integration method (discussed further in Chapter 3).

We can also see how the risk measure itself varies with the degree of risk-aversion from the plot of M_ϕ against γ given in Figure 2.14. As we can see, M_ϕ rises as γ gets smaller. The risk measure rises as the user becomes more risk-averse. It is also curious to note that the shape of this curve is reminiscent of the curves describing the way the VaR changes with the confidence level (see Figure 2.7), and this suggests that the spectral parameter γ plays a similar role in spectral measures as the confidence level plays in the VaR.

All this indicates that there is an optimal risk measure for each user, and the optimal measure depends on the user's risk-aversion function. Two users might have identical portfolios, but

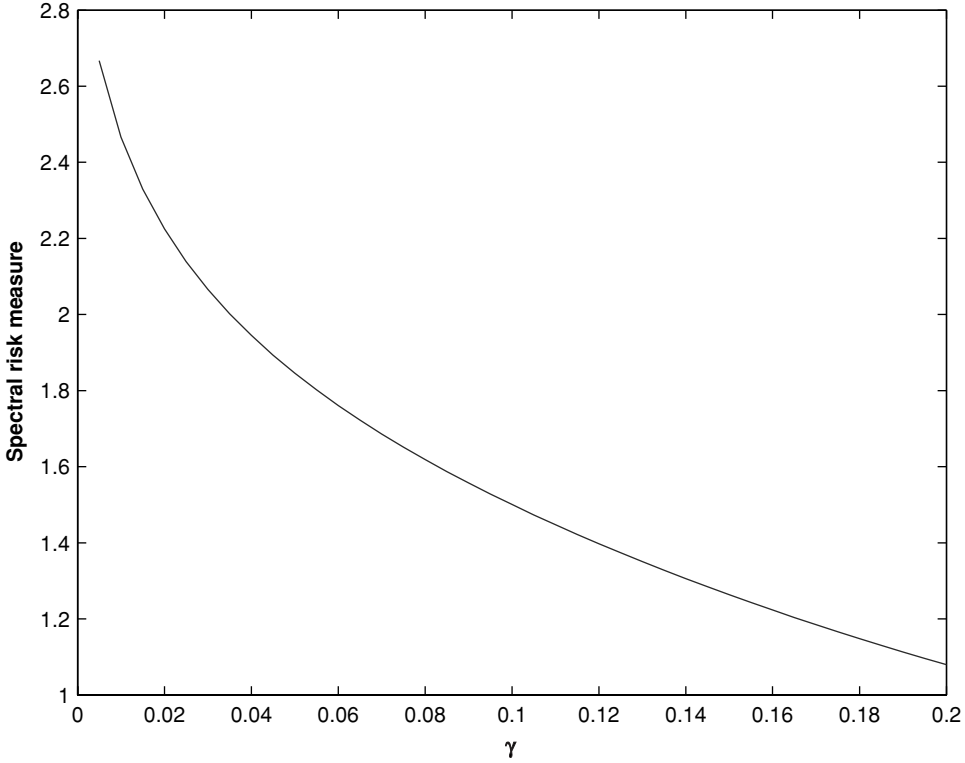

Figure 2.14 Plot of exponential spectral risk measure against γ
Note: Obtained using the 'normal_spectral_risk_measure_plot' function in the MMR Toolbox.

their risks – in the spectral-coherent sense of the term – will only be guaranteed to be the same if they also have exactly the same risk-aversion. From a methodological or philosophical perspective, this means that 'risk' necessarily has a subjective element, even if one subscribes to a frequentist view of probability that maintains that 'probability is objective'. When it comes to risk measures, there is no 'one size that fits all'. This also implies that (true) risk would be very difficult to regulate effectively, if only because regulators could not anticipate the impact of such regulations without subjective information that they are hardly likely to have.[23]

2.3.4 Scenarios as Coherent Risk Measures

The theory of coherent risk measures also sheds some interesting light on usefulness of scenario analyses, as it turns out that the results of scenario analyses can be interpreted as coherent risk measures. Suppose we consider a set of loss outcomes combined with a set of associated probabilities. The losses can be regarded as tail drawings from the relevant distribution function, and their expected (or average) value is the ES associated with this distribution function. Since the ES is a coherent risk measure, this means that the outcomes of scenario analyses are also coherent risk measures. The outcomes of scenario analyses are therefore 'respectable' risk measures, and this means that the theory of coherent risk measures provides a solid risk-theoretical justification for stress testing!

This argument can be extended in some interesting ways. Consider a set of 'generalised scenarios' – a set of n loss outcomes and a *family* of distribution functions from which the losses are drawn. Take any one of these distributions and obtain the associated ES. Now do the same again with another distribution function, leading to an alternative ES. Now do the same again and again. It turns out that the maximum of these ESs is itself a coherent risk measure: if we have a set of m comparable ESs, each of which corresponds to a different loss distribution function, then the maximum of these ESs is a coherent risk measure.[24] Furthermore, if we set $n = 1$, then there is only one tail loss in each scenario and each ES is the same as the probable maximum loss or likely worst-case scenario outcome. If we also set $m = 1$, then it immediately follows that the highest expected loss from a single scenario analysis is a coherent risk measure; and if $m > 1$, then the highest expected of m worst case outcomes is also a coherent risk measure. In short, the ES, the highest expected loss from a set of possible outcomes (or loss estimates from scenario analyses), the highest ES from a set of comparable ESs based on different distribution functions, and the highest expected loss from a set of highest losses, are all coherent risk measures.

[23] There are also other important implications. Any convex combination of two coherent risk measures is also a coherent risk measure, so a manager presiding over two different business units might take the overall risk measure to be some convex combination of the risks of the two subsidiary units. Furthermore, there is no requirement that the risks of the business units will be predicated on the same confidence level or risk-aversion parameters.

[24] An example of a scenario-based coherent risk measure is given by the outcomes of worst-case scenario analyses (WCSA) suggested by Boudoukh *et al.* (1995) and Bahar *et al.* (1997): in essence, these take a large number of sample drawings from a chosen distribution, and the risk measure is the mean of the sample highest losses. Another example of a standard stress testing framework whose outcomes qualify as coherent risk measures is the Standard Portfolio Analysis of Risk (SPAN) system used by the Chicago Mercantile Exchange to calculate margin requirements. This system considers 16 specific scenarios, consisting of standardised movements in underlying risk factors. Fourteen of these are fairly moderate scenarios, and two are extreme. The measure of risk is the maximum loss incurred across all scenarios, using the full loss from the first 14 scenarios and 35% of the loss from the two extreme ones. (Taking 35% of the losses on the extreme scenarios can be regarded as allowing for the extreme losses to be less probable than the others.) The calculations involved can be interpreted as producing the maximum expected loss under 16 distributions. The SPAN risk measures are coherent because the margin requirement is equal to the shortfall from this maximum expected loss.

Thus, the outcomes of (simple or generalised) scenarios can be interpreted as coherent risk measures. However, the reverse is also true as well: coherent risk measures can be interpreted as the outcomes of scenarios associated with particular density functions. This is very useful, because it means that we can always estimate coherent risk measures by specifying the relevant scenarios and then taking (as relevant) their (perhaps probability-weighted) averages or maxima: all we need to know are the loss outcomes (which are quantiles from the loss distribution), the density functions to be used (which give us our probabilities), and the type of coherent risk measure we are seeking.[25,26]

Box 2.2 Distortion Risk Measures

Distortion risk measures are closely related to coherent measures, but emerged from the actuarial/insurance literature rather than the mainstream financial risk literature. They were proposed by Wang (1996) and have been applied to a wide variety of insurance problems, most particularly to the determination of insurance premiums.

A distortion risk measure is the expected loss under a transformation of the cumulative density function known as a distortion function, and the choice of distortion function determines the particular risk measure. More formally, if $F(x)$ is some cdf, the transformation $F^*(x) = g(F(x))$ is a distortion function if $g{:}[0,1] \rightarrow [0,1]$ is an increasing function with $g(0) = 0$ and $g(1) = 1$. The distortion risk measure is then the expectation of the random loss X using probabilities obtained from $F^*(x)$ rather than $F(x)$. Like coherent risk measures, distortion risk measures have the properties of homogeneity, positive homogeneity, and translational invariance; they also share with spectral risk measures the property of comonotonic additivity. To make good use of distortion measures, we would choose a 'good' distortion function, and there are many distortion functions to choose from. The properties we might look for in a 'good' distortion function include continuity, concavity, and differentiability; of these, continuity is necessary and sufficient for the distortion risk measure to be coherent, and concavity is sufficient (Wang *et al*. (1997)).

Of the various distortion functions the best-known is the renowned Wang transform (Wang (2000)):

$$g(u) = \Phi[\Phi^{-1}(u) - \lambda]$$

where λ can be taken to be equal to $\Phi^{-1}(\alpha)$ or to the market price of risk. This distortion function is everywhere continuous and differentiable. The continuity of this distortion function also means that it produces coherent risk measures, and these measures are superior to the ES because they take account of the losses below VaR, and also take better account of extreme losses.

[25] Coherent risk measures produce other surprises too. There is an intimate link between coherent risk measures and the generalised arbitrage bounds or 'good deal bounds' of Cerny and Hodges (1999). This leads to some interesting and profound interrelationships between coherent risk measures, arbitrage valuation, valuation bounds, portfolio optimisation and utility maximisation. For more on these, see Jaschke and Küchler (2000).

[26] Another important related family of risk measures are the dynamic or multi-period risk measures. Multi-period measures take account of interim cash flows, and allow us to look at risk measures *over* a period rather than just at the *end* of it. Dynamic risk measures are also more satisfactory in dynamic situations where, for example, 10-day risk measures are rolled forward from one day to the next. When used in the context of larger optimisation problems, dynamic risk measures are less prone to consistency issues over time. For more on these measures, see, e.g., Wang (1999) and Cvitanic and Karatzas (1999).

2.4 CONCLUSIONS

This chapter has reviewed three alternative risk measurement frameworks. The first, the mean–variance framework, is adequate in the face of a limited set of situations (i.e., if we assume returns or losses are elliptical or if we impose unreasonable restrictions on the utility function). This leaves us with the problem of finding a 'good' risk measure that can be used in less restrictive conditions. The answer often proposed is to use the VaR. Because the VaR is simply a quantile, we can estimate it for any distribution we like. However, the VaR has serious flaws as a measure of risk, and there are good grounds to say that it should not be regarded as a 'proper' risk measure at all. A better answer is to use coherent risk measures. These give us 'respectable' measures of risk that are valid for all possible return or loss distributions, and they are manifestly superior to the VaR as a risk measure. The solution is therefore to upgrade further from VaR to coherent (or distortion) risk measures. These better risk measures are straightforward to estimate if one already has a VaR calculation engine in place, as the costs of upgrading from a VaR calculation engine to a coherent (or distortion) risk measure engine are very small. Perhaps the key lesson in all of this is that it is much less important *how* we estimate risk measures; it is much more important that we estimate the *right* risk measure.

Appendix 1
Probability Functions

A2.1 PROBABILITY DENSITY FUNCTIONS, CUMULATIVE DENSITY FUNCTIONS AND QUANTILES

A probability density function (pdf) gives the probability that a random variable will take a particular value, or take a value in a particular range. More formally, the probability that a random variable X falls between a and b can be written as:

$$\Pr[a < X < b] = \int_a^b f(x)dx \tag{A2.1}$$

where $f(x)$ is the probability density function. The pdf must be non-negative for all values of x (i.e., $f(x) \geq 0$ for all x) and the integral of Equation (A2.1) over all possible x-values must be 1:

$$\int_{x_{\min}}^{x_{\max}} f(x)dx = 1 \tag{A2.2}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of x, and will in many cases be $-\infty$ and ∞.

Corresponding to the pdf $f(x)$ is the cumulative density function $F(x)$ (cdf, sometimes known as the distribution function). This gives the probability that X will be less than or equal to a particular value x:

$$F(x) = \Pr[X \leq x] = \int_{x_{\min}}^{x} f(u)du \tag{A2.3}$$

$F(x)$ varies from 0 to 1. It takes the value 0 when X is at its minimum value (i.e., $F(x_{\min}) = 0$), and it takes the value 1 when X is at its maximum value (i.e., $F(x_{\max}) = 1$). The probability that X exceeds x is 1 minus the probability that $X \leq x$ (i.e., $\Pr[X > x] = 1 - F(x)$).

Note also that in risk measurement we are often very interested in the x-value that corresponds to a given cumulative probability or confidence level. This x-value is the inverse of a cdf, given by $x = F^{-1}(p)$, where p is a specified cumulative probability. x is often known as a quantile or percentile point, and the VaR itself is just a quantile.

A2.2 CONDITIONAL VS UNCONDITIONAL DISTRIBUTIONS

In applying distributions to data, there is also an important distinction between conditional and unconditional distributions. A conditional distribution gives us the distribution conditional on the values of one or more parameters, and might say, for example, that a certain return is normal conditional on a given volatility value. A conditional distribution is often used in

short-term financial risk measurement problems (e.g., where we might forecast 1-day VaRs using a conditional distribution with a volatility that changes randomly from one day to the next). By comparison, an unconditional distribution gives us the distribution of the relevant random variable without specifying the values of the conditioning factors. For example, the unconditional distribution might specify that a return is normal. The unconditional distribution therefore gives us the actual distribution of the random variable, and is often used for longer-term risk measurement problems.

A2.3 FEATURES OF PROBABILITY DENSITY FUNCTIONS

A pdf has a number of interesting features. One of these is its central tendency, which is variously described by any of the following:

- The *mean* (often represented by μ): the expected value $E(X) = \mu = \int_{x_{\min}}^{x_{\max}} xf(x)dx$.
- The *median*: that value, x_{median}, such that there is a 50% probability that X will exceed x_{median}, and a 50% probability that X will be less than or equal to x_{median}. Thus, x_{median} is implicitly defined by $F(x_{median}) = 0.5$, so $x_{median} = F^{-1}(0.5)$.
- The *mode*: the most likely value, or that value of x at which $f(x)$ is at a maximum.

The mean, median and mode will be the same for a symmetric pdf, and will generally differ for asymmetric distributions.

A second feature is the pdf's dispersion, which can also be described in different ways, including:

- The *variance* (often represented as σ^2): the average squared distance between possible X values and the mean (i.e., $\sigma^2 = \int_{x_{\min}}^{x_{\max}} (x - \mu)^2 f(x)dx = E(X^2) - [E(X)]^2$.
- The *standard deviation* σ: the square root of the variance.
- The *mean absolute deviation* (MAD): the average value of the difference between possible values of X and some measure of central tendency.

The variance/standard deviation is the most popular, in part because of its convenient analytical properties. However, the MAD has the advantage of being more robust to outliers or errors in data.

A third feature is the *skewness*, or asymmetry of the pdf. We can measure this by the coefficient of skewness:

$$skewness = \frac{E[(X - \mu)^3]}{\sigma^3} \tag{A2.4}$$

or by other measures such as differences between the various measures of central tendency.

A fourth feature is the *heaviness of the tails* of the distribution – how quickly the tail probabilities diminish as we go further out into the tail. There are various measures of tail heaviness, but the most common is the coefficient of kurtosis:

$$kurtosis = \frac{E[(X - \mu)^4]}{\sigma^4} \tag{A2.5}$$

However, some of these measures do not always exist: for example, the variance does not exist (or does not have a finite value) if the tails of the pdf are too heavy.

Box A2.1 Location, Scale and Shape Parameters

The statistical literature often talks of the 'location' and 'scale' parameters of a distribution. Loosely speaking, the location parameter refers to a natural measure of the distribution's central tendency, and the scale parameter refers to a natural measure of its dispersion. These parameters are often useful when transforming series or deriving results. More particularly, it is often useful to derive results using standardised variates, where the latter are the original variates minus the location parameter, both divided by the scale parameter; we then unravel our results to produce answers specified in terms of the original variates. For example, if we are dealing with a variate X which is distributed as $N(\mu, \sigma^2)$, where μ and σ are the mean and standard deviation of X, we would often standardise X using the transform $Z = (X - \mu)/\sigma$; we then derive our results for Z and use the reverse transform $X = \mu + \sigma Z$ to infer our results for X. In this case, we are using the mean as the location parameter and the standard deviation as the scale parameter. We would use the same location and scale parameters when working with other familiar distributions such as the Student-t.

However, the natural measures of central tendency and dispersion depend on the distribution, and for some distributions these are not the mean and standard deviation. For instance, with a Gumbel distribution the natural measure of central tendency is the mode, and the natural measure of dispersion is not the standard deviation σ, but $\sigma\sqrt{6/\pi}$. So for a Gumbel we would take our location and scale parameters to be the mode and $\sigma\sqrt{6/\pi}$, rather than the mean and the standard deviation. When speaking generally of central tendency and dispersion, we should therefore talk of location and scale parameters, and when dealing with any specific distribution, we should identify the location and scale parameters appropriate to the distribution at hand.

In addition to the location and scale parameters, statisticians sometimes also refer to the shape parameter of a distribution. This parameter refers loosely to a natural measure of the tail-heaviness of the distribution, and again varies with the distribution. For example, with a Student-t distribution the shape would be the degrees of freedom υ, whereas with a Lévy distribution the shape would be the index α.

A2.4 MOMENTS, MOMENT-GENERATING FUNCTIONS AND CHARACTERISTIC FUNCTIONS

A pdf can also be described in terms of its moments. The rth moment of a pdf is defined by

$$r\text{th } \textit{moment} = \int_{-\infty}^{\infty} x^r f(x)dx \tag{A2.6}$$

We often work with central moments, which are the moments around the mean:

$$r\text{th } \textit{central moment} = \int_{-\infty}^{\infty} (x - \mu)^r f(x)dx \tag{A2.7}$$

The moments give us a way of describing the pdf, and are closely related to the various parameters described earlier:

- The first moment is the mean.
- The second central moment is the variance.
- The third central moment is the skewness times σ^3.
- The fourth central moment is the kurtosis times σ^4.

Where moments exist, they tell us a lot about the pdf, and in practice knowledge of the moments can often enable us to recover the whole pdf.

For the rth moment to exist (or be finite-valued), the moment-generating function (Equation (A2.6)) tells us that we need $f(x)$ to decline faster than $|x^r|$ rises, as x moves out into the tail: this means that $f(x)$ should decay faster than $1/|x|^r$ as $|x|$ gets big.

When carrying out computations, it is often helpful to work with the characteristic function $\hat{f}(q)$ of the pdf $f(x)$, which is the latter's Fourier transform:

$$\hat{f}(q) = \int e^{iqx}(x) f(x) dx \tag{A2.8}$$

This is useful because the moments can be obtained by successively differentiating the characteristic function and setting q to 0:

$$r\text{th } \textit{moment} = \left((-i)^r \frac{d^r}{dq^r} \hat{f}(q) \right)_{q=0} \tag{A2.9}$$

This enables us to estimate the moments provided we can carry out the relevant differentiations.

A2.5 SKEWNESS

The skewness and kurtosis both play important roles in risk measurement. As noted already, the third moment gives an indication of the asymmetry or skewness of the distribution. The skewness coefficient (see Equation (A2.4)) will be zero for a symmetric distribution, and non-zero for an asymmetric one. The sign of the coefficient also indicates the direction of the skew: a positive skew indicates a short tail on the left and a long tail on the right, and a negative skew indicates the opposite.

An example of a positively skewed distribution was shown in Figure 2.4. The skew alters the whole distribution, and tends to pull one tail in while pushing the other tail out. If a distribution is skewed, we must take account of its skew if we are to be able to estimate its probabilities and quantiles – and hence its risks – properly.

A2.6 KURTOSIS AND TAIL BEHAVIOUR

In risk measurement, we are also frequently concerned with how quickly the tails diminish as our x-values move further away from the mean. This 'tail heaviness' is reflected in the kurtosis parameter (see Equation (A2.5)). A good benchmark here is the normal distribution, which has a kurtosis of 3 and a pdf that is proportional to $\exp(-(x-\mu)^2/2\sigma^2)$. This pdf decays very quickly as $|(x-\mu)/\sigma|$ gets big, so x-values that are more than a small number of standard deviations from the mean are very rare, and rapidly become more so as we move further out into the tail. For example, the probability of an observation more than 2σ from the mean is 4.55%,

the probability of an observation more than 3σ from the mean is 0.27%, and the probability of an observation more than 4σ from the mean is 0.00006%.

If the kurtosis is greater than 3, our tail is heavier than under normality. Such heavy tails are common in financial returns, and indicate that extreme events are more likely, and more likely to be large, than under normality. If the kurtosis is less than 3, our tail is thinner than under normality. Thin tails indicate that extreme events are less likely, and less likely to be large, than under normality. However, empirical return processes often have heavier than normal tails, and therefore have kurtoses bigger than 3, whereas lighter than normal tails are very rare in such processes.

The effect of excess kurtosis was illustrated in Figure 2.5, which shows in close-up how a symmetric heavy-tailed distribution compares to a standard normal one with the same mean and standard deviation. Because the total area under the density curve must always be 1, the distribution with the heavier tails has less probability mass in its more central region to offset its greater mass in the tail regions. However, the greater tail mass means that for higher confidence levels, heavy-tailed distributions will also have greater than normal VaRs. The tails of heavy-tailed distributions can also be described by an asymptotic power law:

$$tail\ f(x) \sim \frac{1}{|x|^{1+\alpha}} \tag{A2.10}$$

A good example is a Lévy distribution, which obeys a power law with index $\alpha \leq 2$. Under a power law, all rth-order moments are infinite for $r \geq \alpha$, so the third, fourth and higher moments of a Lévy distribution are infinite. Indeed, a Lévy distribution does not even have a finite variance except in the special case of $\alpha = 2$, and in this case the Lévy is equivalent to a normal.

A2.7 SOME IMPORTANT DISTRIBUTIONS

There are many different distributions used in risk management. A good starting point is the normal or Gaussian distribution, which we have already mentioned, whose pdf is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{A2.11}$$

This is based on two parameters, the mean μ and standard deviation σ, and its tail probabilities decline exponentially (i.e., very rapidly) as we go further out into the tail.

The normal can be generalised in various ways. One important generalisation is the family of elliptical distributions. These distributions are so called because their log densities lie on ellipsoids (or, in two dimensions, ellipses) determined by equations of the form $(x-\mu)^2/\sigma^2 = c$ for some constant c (or, in multivariate form, $(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\mathbf{T}} = c$) where $\boldsymbol{\mu}$ is a vector of location terms and $\boldsymbol{\Sigma}$ is a matrix of scale terms). Loosely speaking their pdfs can be expressed in 'normal-like' terms but with different location and/or scale parameters. Hence, if we take the general multivariate case (see section A2.9 below), the pdf of an elliptical distribution can be expressed in the form:

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \varphi([\mathbf{x}-\boldsymbol{\mu}]^T \boldsymbol{\Sigma}^{-1/2}[\mathbf{x}-\boldsymbol{\mu}]) \tag{A2.12}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as defined above. The family of elliptical distributions includes the normal as a special case (i.e., when $\boldsymbol{\mu}$ is a vector of means, and $\boldsymbol{\Sigma}$ is a covariance matrix), but also includes

other useful distributions such as the t (provided that the variance is finite) and the normal inverse Gaussian (NIG), which often provide better fits to empirical return data. However, with suitable reparameterisations, they can also be expressed in terms of the same location and scale parameters as the normal (i.e., mean and variance or covariance matrix). Elliptical distributions share many of the convenient features of the normal, but have the advantage that they are able to accommodate moderate degrees of skewness and excess kurtosis, although non-symmetrical elliptical distributions are rarely used for financial problems. The downside of using these distributions is that parameter estimation can sometimes be difficult (although there are exceptions, such as the t).

The elliptical distributions are closely related to various other families of distributions. These include the hyperbolic distributions (so called because their log densities lie on hyperbolas), generalised hyperbolic distributions, normal mixture distributions, and Lévy distributions. Each of these families is some generalisation of the normal, but the interrelationships between them are extremely complex, and many distributions besides the normal are members of more than one of these families. We shall have more to say on some of these in Chapter 6. Like the elliptical distributions, these distributions share some of the convenient features of the normal, but they can also accommodate moderate degrees of excess kurtosis and (sometimes) skewness. However, parameter estimation can again be difficult.

The other major family of distributions we sometimes meet in financial risk measurement are the extreme-value (EV) distributions, which are suitable for handling problems involving extreme (or low-probability, high-impact) events. We shall have more to say on these in Chapter 7.

A2.8 STABILITY, SELF-SIMILARITY, AND DOMAINS OF ATTRACTION

There are also other properties that we sometimes look for in the distributions we choose. One of these is the property of stability: a pdf is stable if the sum of two or more random variables drawn from that distribution is distributed according to the same distribution. It is clearly helpful in practice to have stability so we can avoid situations where adding random variables leads to a sum with a different distribution. However, the only distributions that are stable are Lévy ones, for which reason the Lévy distributions are sometimes also known as stable Lévy distributions. Thus, the normal is stable because it is a Lévy with $\alpha = 2$, and a Cauchy (or Lorentzian) distribution is stable because it is a Lévy with $\alpha = 1$, but a distribution such as the t is not stable except in the special case where it has 1 degree of freedom and is then equivalent to a Cauchy.

A related property, also unique to Lévy distributions, is scale invariance or self-similarity: this means that a distribution retains its shape over alternative scales. So imagine $Z(1)$ is a process defined over a period of length 1. If $Z(t)$ is self-similar, then it is effectively a scaled version of $Z(1)$, i.e.:

$$Z(t) = t^{1/\alpha} Z(1) \tag{A2.13}$$

Self-similar processes scale (or grow with t) at a rate $1/\alpha$. The famous square root of time rule (i.e., where $Z(t)$ grows with the square root of t) applies as a special case where $\alpha = 2$ and $Z(t)$ is a geometric Brownian motion corresponding to a normal distribution. This also tells us that the square root of time rule applies only if we have normal risks.

A third helpful notion is that of domains of attraction. To say that a distribution has a domain of attraction means that any distribution 'close' to it will have similar properties. This is important because it indicates that small departures from an assumed process (e.g., because of errors in our data) should not produce large errors in any inferences we make. More formally, suppose that S_n is the sum of n random variables from a chosen distribution. The distribution of S_n will typically converge as n gets large, so if certain conditions are met (independence, etc.), the distribution of S_n will converge to a normal distribution. This is what the central limit theorem means. But put another way, the distribution of S_n is in the domain of attraction of the normal, and the domain of attraction itself consists of the set of all distributions of S_n that converge to the normal as n gets large. This particular domain of attraction is therefore a counterpart of the standard central limit theorem. There are also other domains of attraction: for example, the set of all distributions of S_n that converge in the limit to the Lévy is the domain of attraction of the Lévy, and is associated with the central limit theorem as it applies to the Lévy. There are also domains of attraction associated with highest and lowest observations in a sample of size n, and these are the domains of attraction encountered in extreme-value theory. These latter domains of attraction are associated with extreme-value theorems.

A2.9 MULTIVARIATE DISTRIBUTIONS

Finally, there are multivariate distributions: distributions involving more than one random variable. If we consider the two-variable case for simplicity – although generalisation to the n-variable case is straightforward – the joint distribution function for the random variables X and Y is given by:

$$F(x, y) = \Pr[X \leq x, Y \leq y] \tag{A2.14}$$

With a joint probability distribution, we also have to face the issue of how to model the dependence, if any, between the two random variables. We discuss this issue further in the appendix to Chapter 5. Suffice it for the moment to note that dependence is often modelled in terms of linear correlation coefficients, but these only give a valid representation of dependence under restrictive (i.e., elliptical) conditions. For more general distributions dependence needs to be modelled using copulas.

There is also a related issue: in a univariate context, the value of the distribution function is the confidence level, but in a multivariate context it is not. Instead, in the multivariate context the confidence level is $\Pr[X + Y \leq q]$, where q is the VaR or some simple function of it (e.g., its negative, if the random variables are denominated as P/L). We need to infer the probability of the sum from the distribution function or, put another way, we need to obtain the confidence level that goes with some particular VaR, given the distribution function $F(x, y)$. If we can use correlations as our dependence measure, then we can usually sidestep this problem by using correlation-based parametric formulas for the VaR, but if we are using copulas then this problem must be addressed. We shall have more to say on this in Chapter 6.

Multivariate distribution functions are also to be distinguished from marginal distributions, which are the distribution functions for each variable considered separately, i.e.,

$$F(x) = \Pr[X \leq x] \quad \text{and} \quad F(y) = \Pr[Y \leq y] \tag{A2.15}$$

The key difference between the multivariate and the marginal distributions is that the former do, and the latter do not, take account of any dependence structure between the two random variables.

Appendix 2
Regulatory Uses of VaR

Value at risk is also used by financial regulators to determine financial institutions' regulatory capital requirements against market risk. Under the 1996 Amendment to the Basel Accord, banks judged to have sound risk management practices were allowed the option of having their market risk regulatory capital requirements determined by their own VaR estimates. This is known as the 'internal model' approach to regulatory capital requirements. If VaR_t is a VaR forecast relevant to day t, the non-specific daily capital requirement is:

$$\max\left\{VaR_t, \frac{k}{60}\sum_{i=1}^{60} VaR_{t-i+1}\right\} \tag{A2.16}$$

where k, sometimes known as the hysteria factor, is a multiplier in the range [3,4] whose value is set by the bank's supervisor conditional on the results of backtest exercises (see Box 15.2: Regulatory Backtesting Requirements): the better the backtest results, the lower the k. The application of this multiplier is sometimes justified as providing insurance against model risk, non-normal market moves, and similar factors. However, it is effectively pulled out of thin air. The Amendment also requires that VaR be derived at the 99% confidence level using a 10-day holding period. (As an aside, it is interesting to note that these parameters have no real relevance for bank solvency, even though the whole point of capital regulation in the first place is to promote bank solvency.) However, banks are allowed to proxy the 10-day VaR by multiplying the 1-day VaR by the square root of 10. Banks are allowed to calculate the VaR using their own preferred models, subject to certain conditions: models had to be approved by the supervisor, the model used for setting regulatory capital requirements had to be the same one used for internal risk management (the so-called use test), and they had to meet certain technical conditions (e.g., they had to cover non-linear Greek factors, and so on). The actual capital requirement is equal to Equation (A2.16) plus an additional 'specific risk' charge, which covers factors such as credit-related risks on market instruments (e.g., counterparty default risk on OTC positions). This charge is specific to the institution and imposed at the discretion of the supervisor.

The new Basel Accord, Basel II, retains these broad features, but includes some major new provisions relating to credit risk models. Under Basel II, institutions will be allowed to tie their credit risk capital charges to their own credit risk models. Basel II also introduces a new capital charge for operational risk and permits institutions to model operational risk for regulatory capital purposes.

There are also parallel developments in the insurance world, motivated in part by the convergence of insurance and banking, and the desire to achieve a more uniform regulatory treatment of all financial institutions. The EU's proposed new insurance regulation framework, Solvency II, also has some interesting features of its own, including provisions for the use of ES to determine regulatory capital.

3
Estimating Market Risk Measures: An Introduction and Overview

This chapter provides a brief introduction and overview of the main issues in market risk measurement. Our main concerns are:

- *Preliminary data issues*: How to deal with data in profit/loss form, rate-of-return form, and so on.
- *Basic methods of VaR estimation*: How to estimate simple VaRs, and how VaR estimation depends on assumptions about data distributions.
- How to estimate coherent risk measures.
- How to gauge the precision of our risk measure estimators by estimating their standard errors.
- *Overview*: An overview of the different approaches to market risk measurement, and of how they fit together.

We begin with the data issues.

3.1 DATA

3.1.1 Profit/Loss Data

Our data can come in various forms. Perhaps the simplest is in terms of profit/loss (or P/L). The P/L generated by an asset (or portfolio) over the period t, P/L_t, can be defined as the value of the asset (or portfolio) at the end of t plus any interim payments D_t minus the asset value at the end of $t-1$:

$$P/L_t = P_t + D_t - P_{t-1} \tag{3.1}$$

If data are in P/L form, positive values indicate profits and negative values indicate losses.

If we wish to be strictly correct, we should evaluate all payments from the same point of time (i.e., we should take account of the time value of money). We can do so in one of two ways. The first way is to take the present value of P/L_t evaluated at the end of the previous period, $t-1$:

$$\textit{present value}\,(P/L_t) = \frac{(P_t + D_t)}{(1+d)} - P_{t-1} \tag{3.2}$$

where d is the discount rate and we assume for convenience that D_t is paid at the end of t. The alternative is to take the forward value of P/L_t evaluated at the end of period t:

$$\textit{forward value}\,(P/L_t) = P_t + D_t - (1+d)P_{t-1} \tag{3.3}$$

which involves compounding P_{t-1} by d. The differences between these values depend on the discount rate d, and will be small if the periods themselves are short. We will ignore these

differences to simplify the discussion, but they can make a difference in practice when dealing with longer periods.

3.1.2 Loss/Profit Data

When estimating VaR and ES, it is sometimes more convenient to deal with data in loss/profit (L/P) form. L/P data are a simple transformation of P/L data:

$$L/P_t = -P/L_t \tag{3.4}$$

L/P observations assign a positive value to losses and a negative value to losses, and we will call these L/P data 'losses' for short. Dealing with losses is sometimes a little more convenient for risk measurement purposes because the risk measures are themselves denominated in loss terms.

3.1.3 Arithmetic Return Data

Data can also come in the form of arithmetic (or simple) returns. The arithmetic return r_t is defined as:

$$r_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t + D_t}{P_{t-1}} - 1 \tag{3.5}$$

which is the same as the P/L over period t divided by the value of the asset at the end of $t - 1$.

In using arithmetic returns, we implicitly assume that the interim payment D_t does not earn any return of its own. However, this assumption will seldom be appropriate over long periods because interim income is usually reinvested. Hence, arithmetic returns should not be used when we are concerned with long horizons.

3.1.4 Geometric Return Data

Returns can also be expressed in geometric (or compound) form. The geometric return R_t is:

$$R_t = \ln\left(\frac{P_t + D_t}{P_{t-1}}\right) \tag{3.6}$$

The geometric return implicitly assumes that interim payments are continuously reinvested. The geometric return is often more economically meaningful than the arithmetic return, because it ensures that the asset price (or portfolio value) can never become negative regardless of how negative the returns might be. With arithmetic returns, on the other hand, a very low realised return – or a high loss – implies that the asset value P_t can become negative, and a negative asset price seldom makes economic sense.[1]

The geometric return is also more convenient. For example, if we are dealing with foreign currency positions, geometric returns will give us results that are independent of the reference currency. Similarly, if we are dealing with multiple periods, the geometric return over those

[1] This is mainly a point of principle rather than practice. In practice, any distribution we fit to returns is only likely to be an approximation, and many distributions are ill-suited to extreme returns anyway.

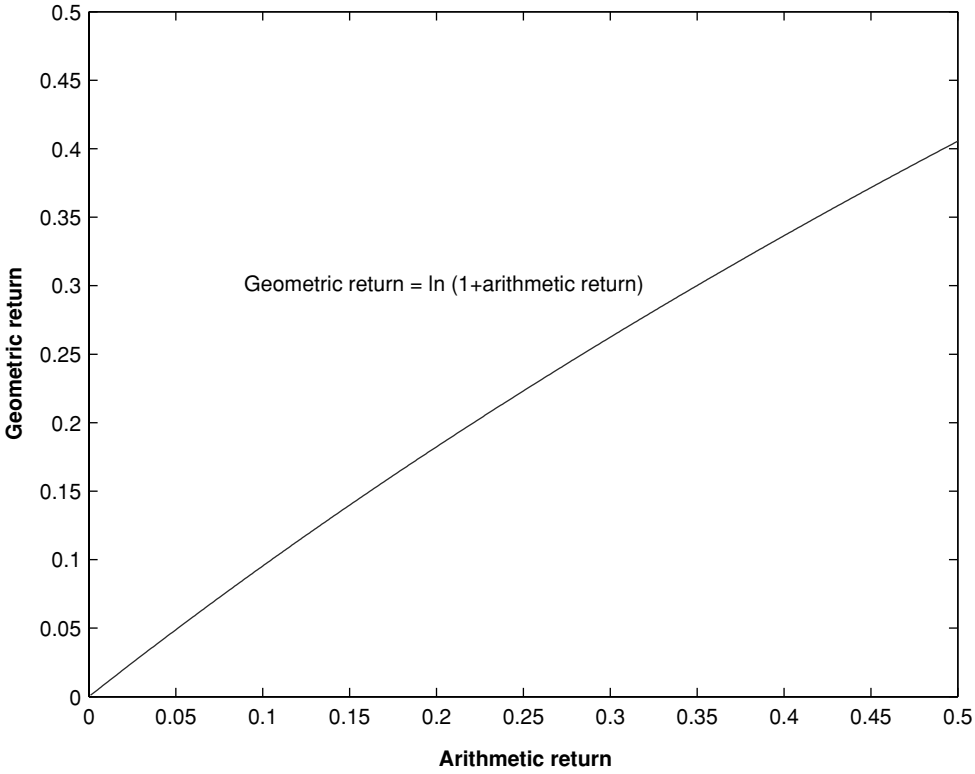

Figure 3.1 Geometric and arithmetic returns

periods is the sum of the one-period geometric returns. Arithmetic returns have neither of these convenient properties.

The relationship of the two types of return can be seen by rewriting Equation (3.6) (using a Taylor's series expansion for the natural log) as:

$$R_t = \ln\left(\frac{P_t + D_t}{P_{t-1}}\right) = \ln(1 + r_t) = r_t - \frac{1}{2}{r_t}^2 + \frac{1}{3}{r_t}^3 - \ldots \tag{3.7}$$

from which we can see that $R_t \approx r_t$ provided that returns are 'small'. This conclusion is illustrated by Figure 3.1, which plots the geometric return R_t against its arithmetic counterpart r_t. The difference between the two returns is negligible when both returns are small, but the difference grows as the returns get bigger – which is to be expected, as the geometric return is a log function of the arithmetic return. Since we would expect returns to be low over short periods and higher over longer periods, the difference between the two types of return is negligible over short periods but potentially substantial over longer ones. And since the geometric return takes account of earnings on interim income, and the arithmetic return does not, we should always use the geometric return if we are dealing with returns over longer periods.

Example 3.1 (Arithmetic and geometric returns)

If arithmetic returns r_t over some period are 0.05, Equation (3.7) tells us that the corresponding geometric returns are $R_t = \ln(1 + r_t) = \ln(1.05) = 0.0488$. Similarly, if geometric returns R_t are 0.05. Equation (3.7) implies that arithmetic returns are $1 + r_t = \exp(R_t) \Rightarrow r_t = \exp(R_t) - 1 = \exp(0.05) - 1 = 0.0513$. In both cases the arithmetic return is close to, but a little higher than, the geometric return – and this makes intuitive sense when one considers that the geometric return compounds at a faster rate.

3.2 ESTIMATING HISTORICAL SIMULATION VaR

The simplest way to estimate VaR is by means of historical simulation (HS). The HS approach estimates VaR by means of ordered loss observations.

Suppose we have 1000 loss observations and are interested in the VaR at the 95% confidence level. Since the confidence level implies a 5% tail, we know that there are 50 observations in the tail, and we can take the VaR to be the 51st highest loss observation.[2]

We can estimate the VaR on a spreadsheet by ordering our data and reading off the 51st largest observation from the spreadsheet. We can also estimate it more directly by using the 'Large' command in Excel, which gives us the kth largest value in an array. Thus, if our data are an array called 'Loss_data', then our VaR is given by the Excel command 'Large(Loss_data,51)'. If we are using MATLAB, we first order the loss/profit data using the 'Sort()' command (i.e., by typing 'Loss_data = Sort(Loss_data)'); and then derive the VaR by typing in 'Loss_data(51)' at the command line.

More generally, if we have n observations, and our confidence level is α, we would want the $(1 - \alpha) . n + 1$th highest observation, and we would use the commands 'Large(Loss_data,(1 − *alpha*)*n + 1)' using Excel, or 'Loss_data((1 − *alpha*)*n + 1)' using MATLAB, provided in the latter case that our 'Loss_data' array is already sorted into ordered observations.[3]

An example of an HS VaR is given in Figure 3.2. This figure shows the histogram of 1000 hypothetical loss observations and the 95% VaR. The figure is generated using the 'hsvarfigure' command in the MMR Toolbox. The VaR is 1.704 and separates the top 5% from the bottom 95% of loss observations.

In practice, it is often helpful to obtain HS VaR estimates from a cumulative histogram, or empirical cumulative frequency function. This is a plot of the ordered loss observations against their empirical cumulative frequency (e.g., so if there are n observations in total, the empirical cumulative frequency of the ith such ordered observation is i/n). The empirical cumulative frequency function of our earlier data set is shown in Figure 3.3. The empirical frequency function makes it very easy to obtain the VaR: we simply move up the cumulative frequency axis to where the cumulative frequency equals our confidence level, draw a horizontal line along to the curve, and then draw a vertical line down to the x-axis, which gives us our VaR.

[2] In theory, the VaR is the quantile that demarcates the tail region from the non-tail region, where the size of the tail is determined by the confidence level, but with finite samples there is a certain level of arbitrariness in how the ordered observations relate to the VaR itself – that is, do we take the VaR to be the 50th observation, the 51st observation, or some combination of them? However, this is just an issue of approximation, and taking the VaR to be the 51st highest observation is not unreasonable.

[3] We can also estimate HS VaR using percentile functions such as the 'Percentile' function in Excel or the 'prctile' function in MATLAB. However, such functions are less transparent (i.e., it is not obvious to the reader how the percentiles are calculated), and the Excel percentile function can be unreliable.

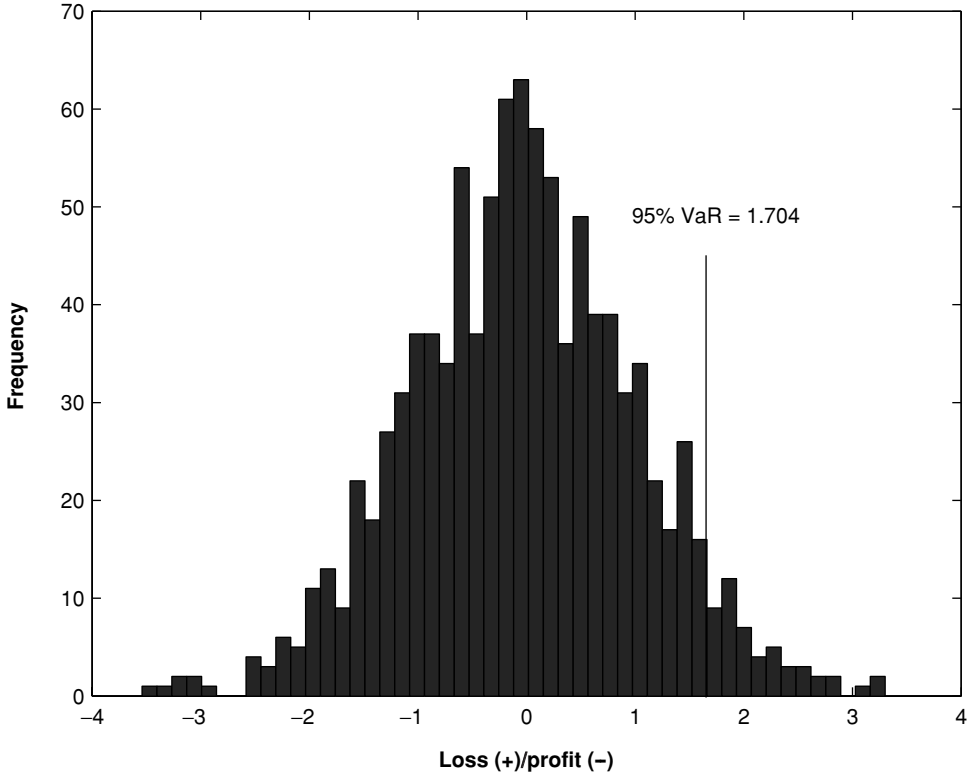

Figure 3.2 Historical simulation VaR
Note: Based on 1000 random numbers drawn from a standard normal L/P distribution, and estimated with the 'hsvarfigure' function.

3.3 ESTIMATING PARAMETRIC VaR

We can also estimate VaR using parametric approaches, the distinguishing feature of which is that they require us to explicitly specify the statistical distribution from which our data observations are drawn. We can also think of parametric approaches as fitting curves through the data and then reading off the VaR from the fitted curve.

In making use of a parametric approach, we therefore need to take account of both the statistical distribution and the type of data to which it applies.

3.3.1 Estimating VaR with Normally Distributed Profits/Losses

Suppose that we wish to estimate VaR under the assumption that P/L is normally distributed. In this case our VaR at the confidence level α is:

$$\alpha VaR = -\mu_{P/L} + \sigma_{P/L} z_\alpha \tag{3.8}$$

where z_α is the standard normal variate corresponding to α, and $\mu_{P/L}$ and $\sigma_{P/L}$ are the mean and standard deviation of P/L. Thus, z_α is the value of the standard normal variate such that α

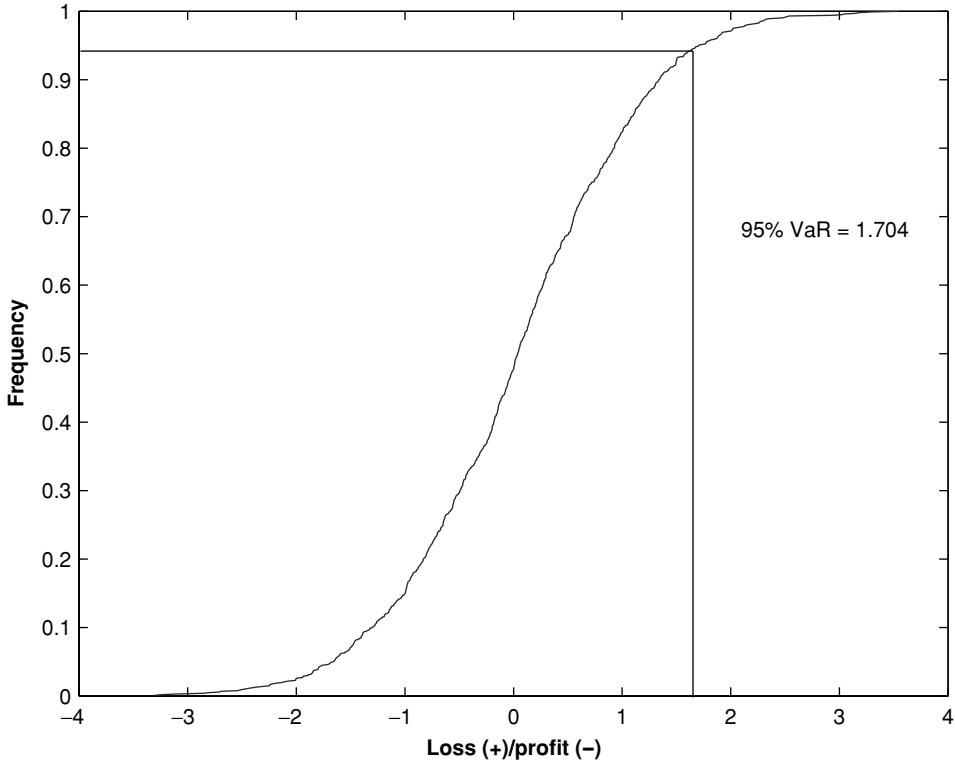

Figure 3.3 Historical simulation via an empirical cumulative frequency function
Note: Based on the same data as Figure 3.2.

of the probability density mass lies to its left, and $1 - \alpha$ of the probability density mass lies to its right. For example, if our confidence level is 95%, $z_\alpha = z_{0.95}$ will be 1.645.

In practice, $\mu_{P/L}$ and $\sigma_{P/L}$ would be unknown, and we would have to estimate VaR based on estimates of these parameters. Our VaR estimate, αVaR^e, would then be:

$$\alpha VaR^e = -m_{P/L} + s_{P/L} z_\alpha \tag{3.9}$$

where $m_{P/L}$ and $s_{P/L}$ are estimates of the mean and standard deviation of P/L.

Figure 3.4 shows the 95% VaR for a normally distributed P/L with mean 0 and standard deviation 1. Since the data are in P/L form, the VaR is indicated by the negative of the cut-off point between the lower 5% and the upper 95% of P/L observations. The actual VaR is the negative of –1.645, and is therefore 1.645.

If we are working with normally distributed L/P data, then $\mu_{L/P} = -\mu_{P/L}$ and $\sigma_{L/P} = \sigma_{P/L}$, and it immediately follows that:

$$\alpha VaR = \mu_{L/P} + \sigma_{L/P} z_\alpha \tag{3.10a}$$

$$\alpha VaR^e = m_{L/P} + s_{L/P} z_\alpha \tag{3.10b}$$

Figure 3.5 illustrates the corresponding VaR. This figure gives the same information as Figure 3.4, but is a little more straightforward to interpret because the VaR is defined in units

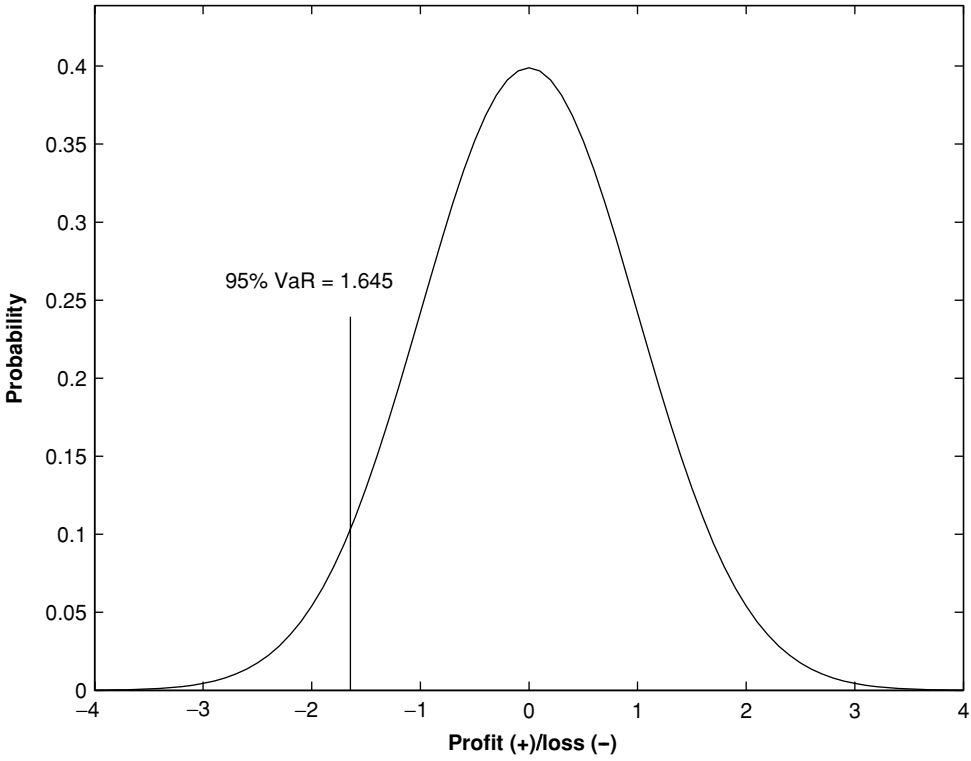

Figure 3.4 VaR with standard normally distributed profit/loss data
Note: Obtained from Equation (3.9) with $\mu_{P/L} = 0$ and $\sigma_{P/L} = 1$. Estimated with the 'normalvarfigure' function.

of losses (or 'lost money') rather than P/L. In this case, the VaR is given by the point on the x-axis that cuts off the top 5% of the pdf mass from the bottom 95% of pdf mass. If we prefer to work with the cumulative density function, the VaR is the x-value that corresponds to a cdf value of 95%. Either way, the VaR is again 1.645, as we would (hopefully) expect.

Example 3.2 (VaR with normal P/L)

If P/L over some period is normally distributed with mean 10 and standard deviation 20, then (by Equation (3.8)) the 95% VaR is $-10 + 20z_{0.95} = -10 + 20 \times 1.645 = 22.9$. The corresponding 99% VaR is $-10 + 20z_{0.99} = -10 + 20 \times 2.326 = 36.52$.

3.3.2 Estimating VaR with Normally Distributed Arithmetic Returns

We can also estimate VaR making assumptions about returns rather than P/L. Suppose then that we assume that arithmetic returns are normally distributed with mean μ_r and standard deviation σ_r. To derive the VaR, we begin by obtaining the critical value of r_t, r^*, such that

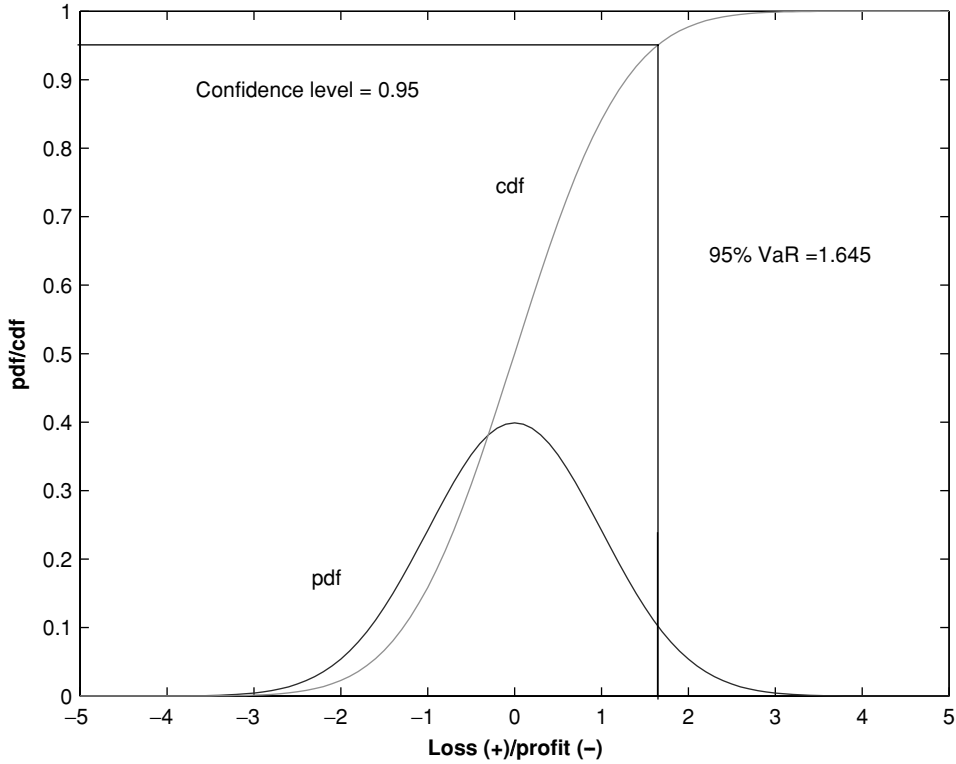

Figure 3.5 VaR with normally distributed loss/profit data
Note: Obtained from Equation (3.10a) with $\mu_{L/P} = 0$ and $\sigma_{L/P} = 1$.

the probability that r_t exceeds r^* is equal to our confidence level α. r^* is therefore:

$$r^* = \mu_r - \sigma_r z_\alpha \tag{3.11}$$

Since the actual return r_t is the loss/profit divided by the earlier asset value, P_{t-1}, it follows that:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} = -\frac{Loss_t}{P_{t-1}} \tag{3.12}$$

Substituting r^* for r_t then gives us the relationship between r^* and the VaR:

$$r_t^* = \frac{P_t^* - P_{t-1}}{P_{t-1}} = -\frac{VaR}{P_{t-1}} \tag{3.13}$$

Substituting Equation (3.11) into Equation (3.13) and rearranging then gives us the VaR itself:

$$\alpha VaR = -(\mu_r - \sigma_r z_\alpha) P_{t-1} \tag{3.14}$$

Equation (3.14) will give us equivalent answers to our earlier VaR equations. For example, if we set $\alpha = 0.95$, $\mu_r = 0$, $\sigma_r = 1$ and $P_{t-1} = 1$, which correspond to our earlier illustrative P/L and L/P parameter assumptions, αVaR is 1.645: the three approaches give the same results, because all three sets of underlying assumptions are equivalent.

Example 3.3 (VaR with normally distributed arithmetic returns)

Suppose arithmetic returns r_t over some period are distributed as normal with mean 0.1 and standard deviation 0.25, and we have a portfolio currently worth 1. Then (by Equation (3.14)) the 95% VaR is $-0.1 + 0.25 \times 1.645 = 0.331$, and the 99% VaR is $-0.1 + 0.25 \times 2.326 = 0.482$.

3.3.3 Estimating Lognormal VaR

Each of the previous approaches assigns a positive probability of the asset value, P_t, becoming negative, but we can avoid this drawback by working with geometric returns. Now assume that geometric returns are normally distributed with mean μ_R and standard deviation σ_R. If D_t is zero or reinvested continually in the asset itself (e.g., as with profits reinvested in a mutual fund), this assumption implies that the natural logarithm of P_t is normally distributed, or that P_t itself is lognormally distributed. The lognormal distribution is explained in Box 3.1, and a lognormal asset price is shown in Figure 3.6: observe that the price is always non-negative, and its distribution is skewed with a long right-hand tail.

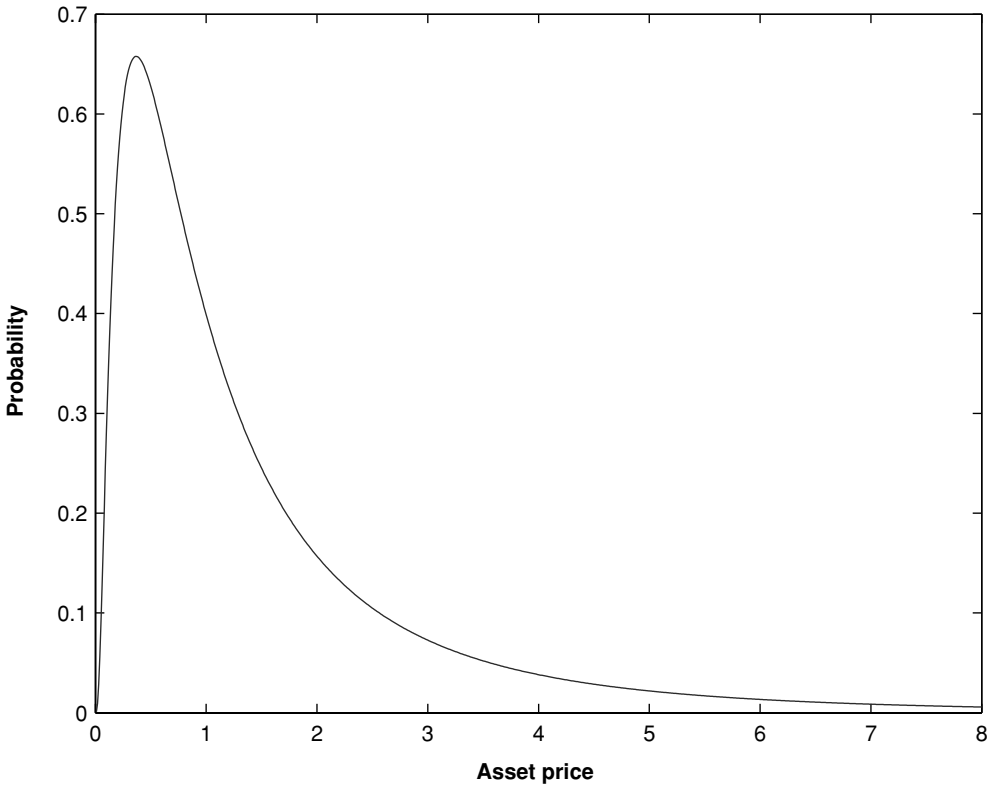

Figure 3.6 A lognormally distributed asset price
Note: Estimated using the 'lognpdf' function in the Statistics Toolbox.

Box 3.1 The Lognormal Distribution

A random variate X is said to be lognormally distributed if the natural log of X is normally distributed. The lognormal distribution can be specified in terms of the mean and standard deviation of $\ln X$. Call these parameters μ and σ. The lognormal is often also represented in terms of m and σ, where m is the median of x, and $m = \exp(\mu)$.

The pdf of X can written:

$$\phi(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log(x) - \mu}{\sigma}\right)^2\right\}$$

for $x > 0$. Thus, the lognormal pdf is only defined for positive values of x and is skewed to the right as in Figure 3.6.

Let $\omega = \exp(\sigma^2)$ for convenience. The mean and variance of the lognormal can be written as:

$$\text{mean} = m\exp(\sigma^2/2) \quad \text{and} \quad \text{variance} = m^2\omega(\omega - 1)$$

Turning to higher moments, the skewness of the lognormal is

$$skewness = (\omega + 2)(\omega - 1)^{1/2}$$

and is always positive, which confirms the lognormal has a long right-hand tail. The kurtosis of the lognormal is

$$kurtosis = \omega^4 + 2\omega^3 + 3\omega^2 - 3$$

and therefore varies from a minimum of (just over) 3 to a potentially large value depending on the value of s.

Since the VaR is a loss, and since the loss is the difference between P_t (which is random) and P_{t-1} (which we can take here as given), then the VaR itself has the same distribution as P_t. Normally distributed geometric returns imply that the VaR is lognormally distributed.

If we proceed as we did earlier with the arithmetic return, we begin by deriving the critical value of R, R^*, such that the probability that $R > R^*$ is α:

$$R^* = \mu_R - \sigma_R z_\alpha \tag{3.15}$$

We then use the definition of the geometric return to unravel the critical value of P^* (i.e., the value of P_t corrresponding to a loss equal to our VaR), and thence infer our VaR:

$$\begin{aligned} R^* &= \ln(P^*/P_{t-1}) = \ln P^* - \ln P_{t-1} \\ &\Rightarrow \ln P^* = R^* + \ln P_{t-1} \\ \Rightarrow P^* &= P_{t-1}\exp[R^*] = P_{t-1}\exp[\mu_R - \sigma_R z_\alpha] \\ \Rightarrow \alpha VaR &= P_{t-1} - P^* = P_{t-1}(1 - \exp[\mu_R - \sigma_R z_\alpha]) \end{aligned} \tag{3.16}$$

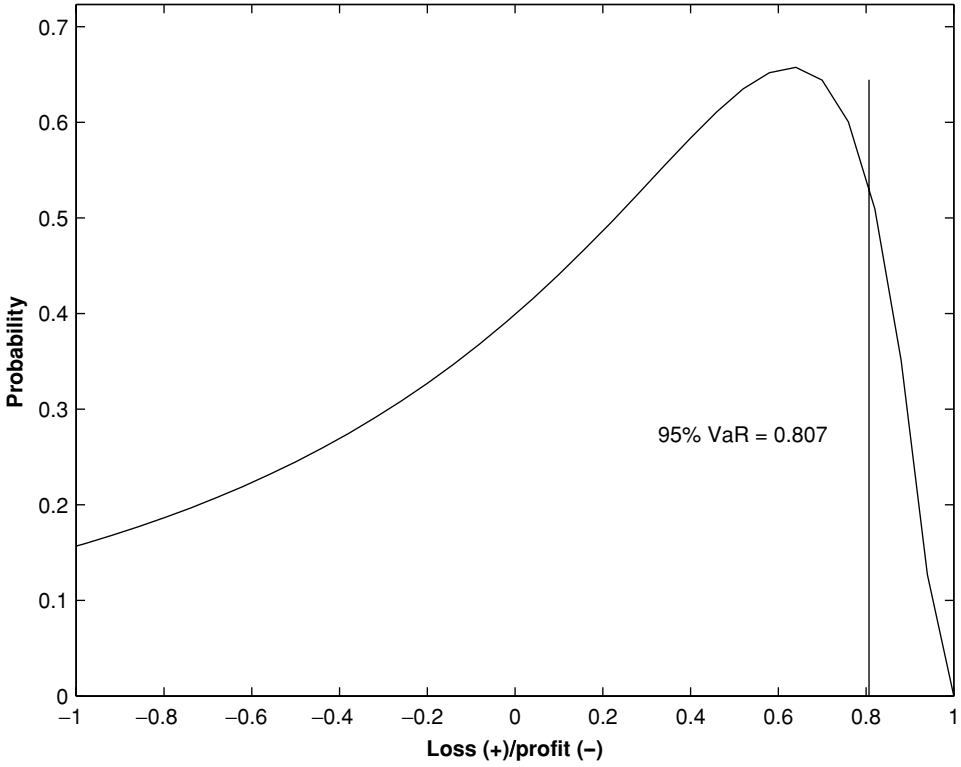

Figure 3.7 Lognormal VaR
Note: Estimated assuming the mean and standard deviation of geometric returns are 0 and 1, and for an initial investment of 1. The figure is produced using the 'lognormalvarfigure' function.

This gives us the lognormal VaR, which is consistent with normally distributed geometric returns.

The lognormal VaR is illustrated in Figure 3.7, based on the standardised (but typically unrealistic) assumptions that $\mu_R = 0$, $\sigma_R = 1$, and $P_{t-1} = 1$. In this case, the VaR at the 95% confidence level is 0.807. The figure also shows that the distribution of L/P is a reflection of the distribution of P_t shown earlier in Figure 3.6.

Example 3.4 (Lognormal VaR)

Suppose that geometric returns R_t over some period are distributed as normal with mean 0.05, standard deviation 0.20, and we have a portfolio currently worth 1. Then (by Equation (3.16)) the 95% VaR is $1 - \exp(0.05 - 0.20 \times 1.645) = 0.244$. The corresponding 99% VaR is $1 - \exp(0.05 - 0.20 \times 2.326) = 0.340$. Observe that these VaRs are quite close to those obtained in Example 3.3, where the arithmetic return parameters were the same as the geometric return parameters assumed here.

Example 3.5 (Lognormal VaR vs normal VaR)

Suppose that we make the empirically not too unrealistic assumptions that the mean and volatility of annualised returns are 0.10 and 0.40. We are interested in the 95% VaR at the 1-day holding period for a portfolio worth \$1. Assuming 250 trading days to a year, the daily return has a mean $0.1/250 = 0.00040$ and standard deviation $0.40/\sqrt{250} = 0.0253$. The normal 95% VaR is $-0.0004 + 0.0253 \times 1.645 = 0.0412$. If we assume a lognormal, then the 95% VaR is $1 - \exp(0.0004 - 0.0253 \times 1.645) = 0.0404$. The normal VaR is 4.12% and the lognormal VaR is 4.04% of the value of the portfolio. This illustrates that normal and lognormal VaRs are much the same if we are dealing with short holding periods and realistic return parameters.

3.4 ESTIMATING COHERENT RISK MEASURES

3.4.1 Estimating Expected Shortfall

We turn now to the estimation of coherent risk measures, and the easiest of these to estimate is the expected shortfall (ES). The ES is the probability-weighted average of tail losses, and a normal ES is illustrated in Figure 3.8. In this case, the 95% ES is 2.063, corresponding to our earlier normal 95% VaR of 1.645.

The fact that the ES is a probability-weighted average of tail losses suggests that we can estimate ES as an average of 'tail VaRs'.[4] The easiest way to implement this approach is to slice the tail into a large number n of slices, each of which has the same probability mass, estimate the VaR associated with each slice, and take the ES as the average of these VaRs.

To illustrate the method, suppose we wish to estimate a 95% ES on the assumption that losses are normally distributed with mean 0 and standard deviation 1. In practice, we would use a high value of n and carry out the calculations on a spreadsheet or using appropriate software. However, to show the procedure manually, let us work with a very small n value of 10. This value gives us 9 (i.e., $n - 1$) tail VaRs, or VaRs at confidence levels in excess of 95%. These VaRs are shown in Table 3.1, and vary from 1.6954 (for the 95.5% VaR) to 2.5758 (for the 99.5% VaR). Our estimated ES is the average of these VaRs, which is 2.0250.

Of course, in using this method for practical purposes, we would want a value of n large enough to give accurate results. To give some idea of what this might be, Table 3.2 reports some alternative ES estimates obtained using this procedure with varying values of n. These results show that the estimated ES rises with n, and gradually converges to the true value of 2.063. These results also show that our ES estimation procedure seems to be reasonably accurate even for quite small values of n. Any decent computer should therefore be able to produce accurate ES estimates quickly in real time.

3.4.2 Estimating Coherent Risk Measures

Other coherent risk measures can be estimated using modifications of this 'average VaR' method. Recall that a coherent risk measure is a weighted average of the quantiles (denoted

[4] The obvious alternative is to seek a 'closed-form' solution, which we could use to estimate the ES, but ES formulas seem to be known only for a limited number of parametric distributions (e.g., elliptical, including normal, and generalised Pareto distributions), whereas the 'average-tail-VaR' method is easy to implement and can be applied to any 'well-behaved' ESs that we might encounter, parametric or otherwise. Formulas for parametric ES are given in Chapters 6 and 7.

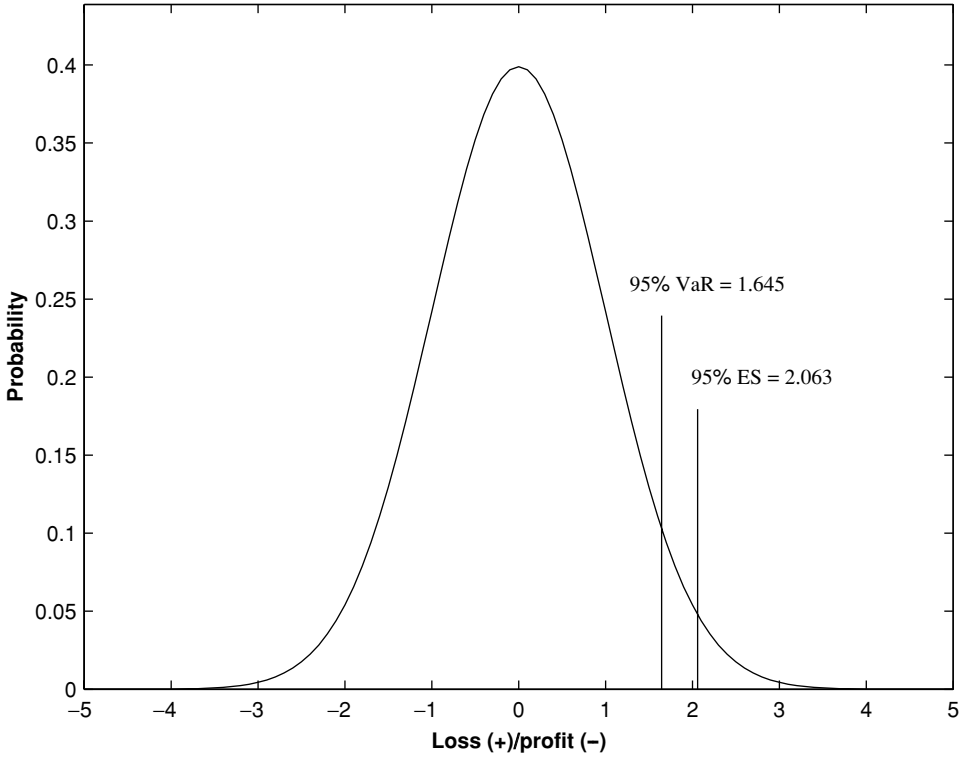

Figure 3.8 Normal VaR and ES
Note: Estimated with the mean and standard deviation of P/L equal to 0 and 1 respectively, using the 'normalesfigure' function.

Table 3.1 Estimating ES as a weighted average of tail VaRs

Confidence level	Tail VaR
95.5%	1.6954
96.0%	1.7507
96.5%	1.8119
97.0%	1.8808
97.5%	1.9600
98.0%	2.0537
98.5%	2.1701
99.0%	2.3263
99.5%	2.5738
Average of tail VaRs	2.0250

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox.

Table 3.2 ES estimates as a function of the number of tail slices

Number of tail slices (n)	ES
10	2.0250
25	2.0433
50	2.0513
100	2.0562
250	2.0597
500	2.0610
1000	2.0618
2500	2.0623
5000	2.0625
10 000	2.0626
True value	2.0630

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1.

by q_p) of our loss distribution:

$$M_\phi = \int_0^1 \phi(p) q_p dp \tag{3.17}$$

where the weighting function or risk-aversion function $\phi(p)$ is specified by the user. As discussed in Chapter 2, the ES gives all tail-loss quantiles an equal weight, and other quantiles a weight of 0. Thus the ES is a special case of M_ϕ obtained by setting $\phi(p)$ to the following:

$$\phi(p) = \begin{cases} 0 & \text{if } p < \alpha \\ 1/(1-\alpha) & \text{if } p \geq \alpha \end{cases} \tag{3.18}$$

The more general coherent risk measure, M_ϕ, involves a potentially more sophisticated weighting function $\phi(p)$. We can therefore estimate any of these measures by replacing the equal weights in the 'average VaR' algorithm with the $\phi(p)$ weights appropriate to the risk measure being estimated.

To show how this might be done, suppose we have the exponential weighting function (Equation (2.9)) from Chapter 2:

$$\phi_\gamma(p) = \frac{e^{-(1-p)/\gamma}}{\gamma(1-e^{-1/\gamma})} \tag{3.19}$$

and we believe that we can represent the degree of our risk-aversion by setting $\gamma = 0.05$. To illustrate the procedure manually, we continue to assume that losses are standard normally distributed and we set $n = 10$ (i.e., we divide the complete losses density function into 10 equal-probability slices). With $n = 10$, we have $n - 1 = 9$ (i.e., $n - 1$) loss quantiles or VaRs spanning confidence levels from 0.1 to 0.90. These VaRs are shown in the second column of Table 3.3, and vary from -1.2816 (for the 10% VaR) to 1.2816 (for the 90% VaR). The third column shows the $\phi(p)$ weights corresponding to each confidence level, and the fourth column shows the products of each VaR and corresponding weight. Our estimated exponential spectral risk measure is the $\phi(p)$-weighted average of the VaRs, and is therefore equal to 0.4228.

Table 3.3 Estimating exponential spectral risk measure as a weighted average of VaRs

Confidence level (α)	αVaR	Weight $\phi(\alpha)$	$\phi(\alpha) \times \alpha VaR$
10%	−1.2816	0	0.0000
20%	−0.8416	0	0.0000
30%	−0.5244	0	0.0000
40%	−0.2533	0.0001	0.0000
50%	0	0.0009	0.0000
60%	0.2533	0.0067	0.0017
70%	0.5244	0.0496	0.0260
80%	0.8416	0.3663	0.3083
90%	1.2816	2.7067	3.4689
Risk measure = mean ($\phi(\alpha)$ times αVaR) =			0.4226

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (3.19)) with $\gamma = 0.05$.

As when estimating the ES earlier, when using this type of routine in practice we would want a value of n large enough to give accurate results. Table 3.4 reports some alternative estimates obtained using this procedure with increasing values of n. These results show that the estimated risk measure rises with n, and gradually converges to a value in the region of about 1.854. The estimates in this table indicate that we may need a considerably larger value of n than we did earlier to get results of the same level of accuracy. Even so, a good computer should still be able to produce accurate estimates of spectral risk measures fairly quickly.

When estimating ES or more general coherent risk measures in practice, it also helps to have some guidance on how to choose the value of n. Granted that the estimate does eventually converge to the true value as n gets large, one useful approach is to start with some small value

Table 3.4 Estimates of exponential spectral coherent risk measure as a function of the number of tail slices

Number of tail slices	Estimate of exponential spectral risk measure
10	0.4227
50	1.3739
100	1.5853
250	1.7338
500	1.7896
1000	1.8197
2500	1.8392
5000	1.8461
10 000	1.8498
50 000	1.8529
100 000	1.8533
500 000	1.8536

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (3.19)) with $\gamma = 0.05$.

Table 3.5 Estimated risk measures and halving errors

Number of tail slices	Estimated spectral risk measure	Halving error
100	1.5853	0.2114
200	1.7074	0.1221
400	1.7751	0.0678
800	1.8120	0.0368
1600	1.8317	0.0197
3200	1.8422	0.0105
6400	1.8477	0.0055
12 800	1.8506	0.0029
25 600	1.8521	0.0015
51 200	1.8529	0.0008

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (3.19)) with $\gamma = 0.05$.

of n, and then double n repeatedly until we feel the estimates have settled down sufficiently. Each time we do so, we halve the width of the discrete slices, and we can monitor how this 'halving' process affects our estimates. This suggests that we look at the 'halving error' ε_n given by:

$$\varepsilon_n = \hat{M}^{(n)} - \hat{M}^{(n/2)} \tag{3.20}$$

where $\hat{M}^{(n)}$ is our estimated risk measure based on n slices. We stop doubling n when ε_n falls below some tolerance level that indicates an acceptable level of accuracy. The process is shown in Table 3.5. Starting from an arbitrary value of 100, we repeatedly double n (so it becomes 200, 400, 800, etc.). As we do so, the estimated risk measure gradually converges, and the halving error gradually falls. So, for example, for $n = 6400$, the estimated risk measure is 1.8477, and the halving error is 0.0055. If we double n to 12 800, the estimated risk measure becomes 1.8506, and the halving error falls to 0.0029, and so on.

However, this 'weighted average quantile' procedure is rather crude, and (bearing in mind that the risk measure (Equation (3.17)) involves an integral) we can in principle expect to get substantial improvements in accuracy if we resorted to more 'respectable' numerical integration or quadrature methods (e.g., trapezoidal and Simpson's rules, etc.; see Appendix 2 to this chapter) and/or by making use of a Richardson extrapolation method (see Box 3.2). This said, the crude 'weighted average quantile' method actually seems to perform well for spectral exponential risk measures when compared against some of these alternatives, so one is not necessarily better off with the more sophisticated methods.[5]

Thus, the key to estimating any coherent risk measure is to be able to estimate quantiles or VaRs: the coherent risk measures can then be obtained as appropriately weighted averages of quantiles. From a practical point of view, this is extremely helpful as all the building blocks that go into quantile or VaR estimation – databases, calculation routines, etc. – are exactly

[5] There is an interesting reason for this: the spectral weights give the highest loss the highest weight, whereas the quadrature methods such as the trapezoidal and Simpson's rules involve algorithms in which the two most extreme quantiles have their weights specifically cut (see Appendix 2 to this chapter), and this undermines the accuracy of the algorithm relative to the crude approach. However, there are ways round these sorts of problems, and in principle versions of the sophisticated approaches should give better results.

Box 3.2 Richardson Extrapolation

We can sometimes generate considerable increases in the accuracy of estimated risk measures if we apply a Richardson extrapolation method: this is a method that transforms an original sequence into another with the same limit but which converges faster than the original one. To show how this might be applied, suppose $\hat{\theta}_n$ is an estimator of a risk measure based on a sample of size n. Now let n_1, n_2 and n_3 be three sample sizes, where $n_1 < n_2 < n_3$. We then estimate $\hat{\theta}_{n_1}$, $\hat{\theta}_{n_2}$ and $\hat{\theta}_{n_3}$ and take our extrapolated estimator, $\hat{\theta}^*$, as the following weighted average of them:

$$\hat{\theta}^* = \frac{1}{2}\hat{\theta}_{n_1} - 4\hat{\theta}_{n_2} + \frac{9}{2}\hat{\theta}_{n_3}$$

Results reported by Inui and Kijima (2003) suggest that this extrapolation can considerably reduce bias in both VaR and ES estimators, and is especially good at reducing the significant biases of the former.[6]

what we need for the estimation of coherent risk measures as well. If an institution already has a VaR engine, then that engine needs only small adjustments to produce estimates of coherent risk measures: indeed, in many cases, all that needs changing is the last few lines of code in a long data processing system. The costs of switching from VaR to more sophisticated risk measures are therefore very low.

3.5 ESTIMATING THE STANDARD ERRORS OF RISK MEASURE ESTIMATORS

We should always bear in mind that any risk measure estimates that we produce are just that – estimates. We never know the true value of any risk measure, and an estimate is only as good as its precision: if a risk measure is very imprecisely estimated, then the estimator is virtually worthless, because its imprecision tells us that true value could be almost anything; on the other hand, if we know that an estimator is fairly precise, we can be confident that the true value is fairly close to the estimate, and the estimator has some value. Hence, we should always seek to supplement any risk estimates we produce with some indicator of their precision. This is a fundamental principle of good risk measurement practice.

We can evaluate the precision of estimators of risk measures by means of their standard errors, or (generally better) by producing confidence intervals for them. We shall have much more to say on confidence intervals in later chapters, but in this introductory chapter we focus on the more basic indicator, the standard error of a risk measure estimator.

3.5.1 Standard Errors of Quantile Estimators

We first consider the standard errors of quantile (or VaR) estimators. Following Kendall and Stuart,[7] suppose we have a distribution (or cumulative density) function $F(x)$, which might be a parametric distribution function or an empirical distribution function (i.e., a cumulative

[6] See Inui and Kijima (2003).
[7] Kendall and Stuart (1972), pp. 251–252.

histogram) estimated from real data. Its corresponding density or relative-frequency function is $f(x)$. Suppose also that we have a sample of size n, and we select a bin width h. Let dF be the probability that $(k-1)$ observations fall below some value $q - h/2$, that one observation falls in the range $q \pm h/2$, and that $(n-k)$ observations are greater than $q + h/2$. dF is proportional to

$$\{F(q)\}^{k-1} f(q)dq\{1 - F(q)\}^{n-k} \tag{3.21}$$

This gives us the frequency function for the quantile q not exceeded by a proportion k/n of our sample, i.e., the $100(k/n)$th percentile.

If this proportion is p, Kendall and Stuart show that Equation (3.21) is approximately equal to $p^{np}(1-p)^{n(1-p)}$ for large values of n. If ε is a very small increment to p, then

$$p^{np}(1-p)^{n(1-p)} \approx (p+\varepsilon)^{np}(1-p-\varepsilon)^{n(1-p)} \tag{3.22}$$

Taking logs and expanding, Equation (3.22) is itself approximately

$$(p+\varepsilon)^{np}(1-p-\varepsilon)^{n(1-p)} \approx -\frac{n\varepsilon^2}{2p(1-p)} \tag{3.23}$$

which implies that the distribution function dF is approximately proportional to

$$\exp\left(\frac{-n\varepsilon^2}{2p(1-p)}\right) \tag{3.24}$$

Integrating this out,

$$dF = \frac{1}{\sqrt{2\pi}\sqrt{p(1-p)/n}} \exp\left(\frac{-n\varepsilon^2}{2p(1-p)}\right) d\varepsilon \tag{3.25}$$

which tells us that ε is normally distributed in the limit with variance $p(1-p)/n$. However, we know that $d\varepsilon = dF(q) = f(q)dq$, so the variance of q is

$$var(q) \approx \frac{p(1-p)}{n[f(q)]^2} \tag{3.26}$$

This gives us an approximate expression for the variance, and hence its square root, the standard error, of a quantile estimator q.

This expression shows that the quantile standard error depends on p, the sample size n and the pdf value $f(q)$. The way in which the (normal) quantile standard errors depend on these parameters is apparent from Figure 3.9. This shows that:

- The standard error falls as the sample size n rises.
- The standard error rises as the probabilities become more extreme and we move further into the tail – hence, the more extreme the quantile, the less precise its estimator.

In addition, the quantile standard error depends on the probability density function $f(.)$ – so the choice of density function can make a difference to our estimates – and also on the bin width h, which is essentially arbitrary.

The standard error can be used to construct confidence intervals around our quantile estimates in the usual textbook way. For example, a 90% confidence interval for a quantile q is

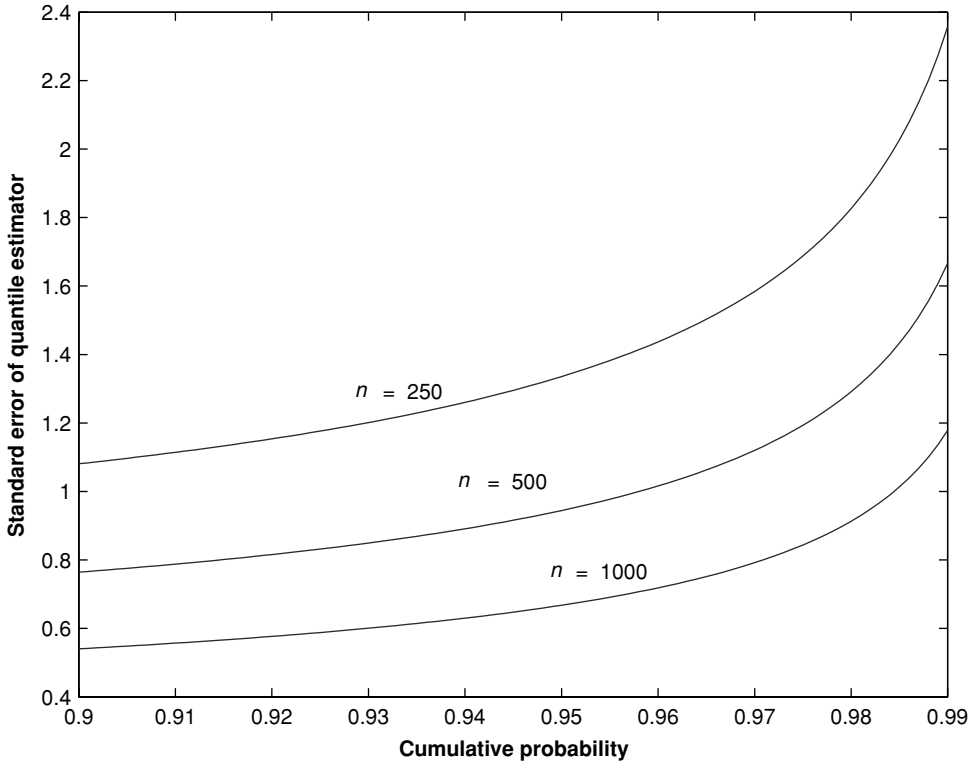

Figure 3.9 Standard errors of quantile estimators
Note: Based on random samples of size n drawn from a standard normal distribution. The bin width h is set to 0.1.

given by

$$[q - 1.645se(q), q + 1.645se(q)] = \left[q - 1.645\frac{\sqrt{p(1-p)/n}}{f(q)}, q + 1.645\frac{\sqrt{p(1-p)/n}}{f(q)}\right] \tag{3.27}$$

Example 3.7 (Obtaining VaR confidence intervals using quantile standard errors)

Suppose we wish to estimate the 90% confidence interval for a 95% VaR estimated on a sample of size of $n = 1000$ to be drawn from a standard normal distribution, based on an assumed bin width $h = 0.1$.

We know that the 95% VaR of a standard normal is 1.645. We take this to be q in Equation (3.27), and we know that q falls in the bin spanning $1.645 \pm 0.1/2 = [1.595, 1.695]$. The probability of a loss exceeding 1.695 is 0.045, and this is also equal to p, and the probability of profit or a loss less than 1.595 is 0.9446. Hence $f(q)$, the probability mass in the q range, is $1 - 0.0450 - 0.9446 = 0.0104$. We now plug the relevant values into Equation (3.27) to

obtain the 90% confidence interval for the VaR:

$$\left[1.645 - 1.645\frac{\sqrt{0.045(1-0.045)/1000}}{0.0104},\ 1.645 + 1.645\frac{\sqrt{0.045(1-0.045)/1000}}{0.0104}\right] = [0.6081, 2.6819]$$

This is a wide confidence interval, especially when compared to the OS and bootstrap confidence intervals given in Chapter 4.

The confidence interval narrows if we take a wider bin width, so suppose that we now repeat the exercise using a bin width $h = 0.2$, which is probably as wide as we can reasonably go with these data. q now falls into the range $1.645 \pm 0.2/2 = [1.545, 1.745]$. p, the probability of a loss exceeding 1.745, is 0.0405, and the probability of profit or a loss less than 1.545 is 0.9388. Hence $f(q) = 1 - 0.0405 - 0.9388 = 0.0207$. Plugging these values into Equation (3.27) now gives us a new estimate of the 90% confidence interval:

$$\left[1.645 - 1.645\frac{\sqrt{0.0405(1-0.0405)/1000}}{0.0207}, 1.645 + 1.645\frac{\sqrt{0.0405(1-0.0405)/1000}}{0.0207}\right] = [1.1496, 2.1404]$$

This is still a rather wide confidence interval.

This example illustrates that although we can use quantile standard errors to estimate VaR confidence intervals, the intervals can be wide and also sensitive to the arbitrary choice of bin width.

The quantile-standard-error approach is easy to implement and has some plausibility with large sample sizes. However, it also has weaknesses relative to other methods of assessing the precision of quantile (or VaR) estimators – it relies on asymptotic theory and requires large sample sizes; it can produce imprecise estimators, or wide confidence intervals; it depends on the arbitrary choice of bin width; and the symmetric confidence intervals it produces are misleading for extreme quantiles whose 'true' confidence intervals are asymmetric reflecting the increasing sparsity of extreme observations as we move further out into the tail.

3.5.2 Standard Errors in Estimators of Coherent Risk Measures

We now consider standard errors in estimators of coherent risk measures. One of the first studies to examine this issue (Yamai and Yoshiba (2001b) did so by investigating the relative accuracy of VaR and ES estimators for Lévy distributions with varying α stability parameters. Their results suggested that VaR and ES estimators had comparable standard errors for near-normal Lévy distributions, but the ES estimators had much bigger standard errors for particularly heavy-tailed distributions. They explained this finding by saying that as tails became heavier, ES estimators became more prone to the effects of large but infrequent losses. This finding suggests the depressing conclusion that the presence of heavy tails might make ES estimators in general less accurate than VaR estimators.

Fortunately, there are grounds to think that such a conclusion might be overly pessimistic. For example, Inui and Kijima (2003) present alternative results showing that the application of a Richardson extrapolation method can produce ES estimators that are both unbiased and

have comparable standard errors to VaR estimators.[8] Acerbi (2004) also looked at this issue and, although he confirmed that tail heaviness did increase the standard errors of ES estimators relative to VaR estimators, he concluded that the relative accuracies of VaR and ES estimators were roughly comparable in empirically realistic ranges.

However, the standard error of any estimator of a coherent risk measure will vary from one situation to another, and the best practical advice is to get into the habit of always estimating the standard error whenever one estimates the risk measure itself. Estimating the standard error of an estimator of a coherent risk measure is also relatively straightforward. One way to do so starts from recognition that a coherent risk measure is an L-estimator (i.e., a weighted average of order statistics), and L-estimators are asymptotically normal. If we take N discrete points in the density function, then as N gets large the variance of the estimator of the coherent risk measure (Equation (3.17)) is approximately

$$\begin{aligned}\sigma(M_{\phi}^{(N)}) &\to \frac{2}{N}\int\limits_{p<q}\phi(p)\phi(q)\frac{p(1-q)}{f(F^{-1}(p))f(F^{-1}(q))}dpdq \\ &= \frac{2}{N}\int\limits_{x<y}\phi(F(x))\phi(F(y))F(x)(1-F(y))dxdy \end{aligned} \tag{3.28}$$

and this can be computed numerically using a suitable numerical integration procedure. Where the risk measure is the ES, the standard error becomes

$$\sigma(ES^{(N)}) \to \frac{1}{N\alpha^2}\int\limits_{0}^{F^{-1}(\alpha)}\int\limits_{0}^{F^{-1}(\alpha)}[\min(F(x),F(y))-F(x)F(y)]dxdy \tag{3.29}$$

and used in conjunction with a suitable numerical integration method, this gives good estimates even for relatively low values of N.[9] If we wish to obtain confidence intervals for our risk measure estimators, we can make use of the asymptotic normality of these estimators to apply textbook formulas (e.g., such as Equation (3.27)) based on the estimated standard errors and centred around a 'good' best estimate of the risk measure.

An alternative approach to the estimation of standard errors for estimators of coherent risk measures is to apply a bootstrap: we bootstrap a large number of estimators from the given distribution function (which might be parametric or non-parametric, e.g., historical); and we estimate the standard error of the sample of bootstrapped estimators. Even better, we can also use a bootstrapped sample of estimators to estimate a confidence interval for our risk measure. We have more to say on how we carry out these calculations in Chapter 4, Appendix 2.

3.6 THE CORE ISSUES: AN OVERVIEW

Before proceeding to more detailed issues in the succeeding chapters, it might be helpful to pause for a moment to take an overview of the structure, as it were, of the subject matter itself. This is very useful, as it gives the reader a mental frame of reference within which the 'detailed' material that follows can be placed. Essentially, there are three core issues, and all the material that follows can be related to these. They also have a natural sequence, so we can think of them as providing a roadmap that leads us to where we want to be.

[8] See Inui and Kijima (2003).
[9] See Acerbi (2004, pp. 200–201).

Which risk measure? The first and most important is to choose the type of risk measure: do we want to a VaR, an ES, etc.? This is logically the first issue, because we need to know *what* we are trying to estimate before we start thinking about *how* we are going to estimate it.

Which level? The second issue is the *level* of analysis. Do we wish to estimate our risk measure at the level of the portfolio as a whole or at the level of the individual positions in it? The former would involve us taking the portfolio as our basic unit of analysis (i.e., we take the portfolio to have a specified composition, which is taken as given for the purposes of our analysis), and this will lead to a *univariate* stochastic analysis. The alternative is to work from the position level, and this has the advantage of allowing us to accommodate changes in the portfolio composition within the analysis itself. The disadvantage is that we then need a *multivariate* stochastic framework, and this is considerably more difficult to handle: we have to get to grips with the problems posed by variance–covariance matrices, copulas, and so on, all of which are avoided if we work at the portfolio level. There is thus a trade-off: working at the portfolio level is more limiting, but easier, while working at the position level gives us much more flexibility, but can involve much more work.

Which method? Having chosen our risk measure and level of analysis, we then choose a suitable estimation method. To decide on this, we would usually think in terms of the classic 'VaR trinity':

- Non-parametric methods
- Parametric methods
- Monte Carlo simulation methods

Each of these involves some complex issues, and we now proceed to deal with these in the next few chapters.

Appendix 1
Preliminary Data Analysis

When confronted with a new data set, we should *never* proceed straight to estimation without some preliminary analysis to get to know our data. Preliminary data analysis is useful because it gives us a feel for our data, and because it can highlight problems with our data set. Remember that we never really know where our data come from, so we should always be a little wary of any new data set, regardless of how reputable the source might appear to be. For example, how do you know that a clerk hasn't made a mistake somewhere along the line in copying the data and, say, put a decimal point in the wrong place? The answer is that you don't, and never can. Even the most reputable data providers provide data with errors in them, however careful they are. Everyone who has ever done any empirical work will have encountered such problems at some time or other: the bottom line is that real data must always be viewed with a certain amount of suspicion.

Such preliminary analysis should consist of at least the first two and preferably all three of the following steps:

- The first and by far the most important step is to eyeball the data to see if they 'look right' – or, more to the point, we should eyeball the data to see if anything looks *wrong*. Does the pattern of observations look right? Do any observations stand out as questionable? And so on. The interocular trauma test is the most important test ever invented and also the easiest to carry out, and we should always perform it on any new data set.
- We should plot our data on a histogram and estimate the relevant summary statistics (i.e., mean, standard deviation, skewness, kurtosis, etc.). In risk measurement, we are particularly interested in any non-normal features of our data: skewness, excess kurtosis, outliers in our data, and the like. We should therefore be on the lookout for any evidence of non-normality, and we should take any such evidence into account when considering whether to fit any parametric distribution to the data.
- Having done this initial analysis, we should consider what kind of distribution might fit our data, and there are a number of useful diagnostic tools available for this purpose, the most popular of which are QQ plots – plots of empirical quantiles against their theoretical equivalents.

A3.1 PLOTTING THE DATA AND EVALUATING SUMMARY STATISTICS

To get to know our data, we should first obtain their histogram and see what might stand out. Do the data look normal, or non-normal? Do they show one pronounced peak, or more than one? Do they seem to be skewed? Do they have fat tails or thin tails? Are there outliers? And so on.

As an example, Figure A3.1 shows a histogram of 100 random observations. In practice, we would usually wish to work with considerably longer data sets, but a data set this small helps to highlight the uncertainties one often encounters in practice. These observations show a

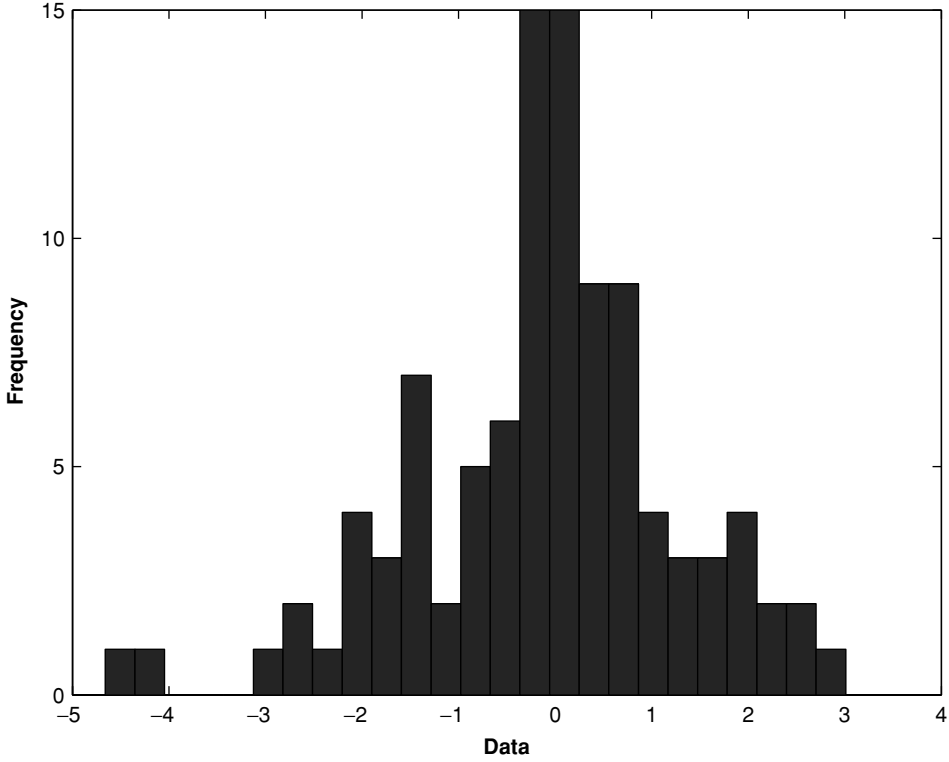

Figure A3.1 A histogram
Note: Data are 100 observations randomly drawn from a Student-*t* with 5 degrees of freedom.

dominant peak in the centre, which suggests that that they are probably drawn from a unimodal distribution. On the other hand, there may be a negative skew, and there are some large outlying observations on the extreme left of the distribution, which might indicate fat tails on at least the left-hand side. In fact, these particular observations are drawn from a Student-*t* distribution with 5 degrees of freedom, so in this case we know that the underlying true distribution is unimodal, symmetric and heavy tailed. However, we would not know this in a situation with 'real' data, and it is precisely because we do not know the distributions of real-world data sets that preliminary analysis is so important.

Some summary statistics for this data set are shown in Table A3.1. The sample mean (−0.099) and the sample mode differ somewhat (−0.030), but this difference is small relative to the sample standard deviation (1.363). However, the sample skew (−0.503) is somewhat negative and the sample kurtosis (3.985) is a little bigger than normal. The sample minimum (−4.660) and the sample maximum (3.010) are also not symmetric about the sample mean or mode, which is further evidence of asymmetry. If we encountered these results with 'real' data, we would be concerned about possible skewness and kurtosis. However, in this hypothetical case we know that the sample skewness is merely a product of sample variation, because we happen to know that the data are drawn from a symmetric distribution.

Depending on the context, we might also seriously consider carrying out some formal tests. For example, we might test whether the sample parameters (mean, standard deviation,

Table A3.1 Summary statistics

Parameter	Value
Mean	−0.099
Mode	−0.030
Standard deviation	1.363
Skewness	−0.503
Kurtosis	3.985
Minimum	−4.660
Maximum	3.010
Number of observations	100

Note: Data are the same observations shown in Figure A3.1.

etc.) are consistent with what we might expect under a null hypothesis (e.g., such as normality).

The underlying principle is very simple: since we *never* know the true distribution in practice, all we ever have to work with are *estimates* based on the *sample* at hand; it therefore behoves us to make the best use of the data we have, and to extract as much information as possible from them.

A3.2 QQ PLOTS

Having done our initial analysis, it is often good practice to ask what distribution might fit our data, and a very useful device for identifying the distribution of our data is a quantile–quantile or QQ plot – a plot of the quantiles of the empirical distribution against those of some specified distribution. The shape of the QQ plot tells us a lot about how the empirical distribution compares to the specified one. In particular, if the QQ plot is linear, then the specified distribution fits the data, and we have identified the distribution to which our data belong. In addition, departures of the QQ from linearity in the tails can tell us whether the tails of our empirical distribution are fatter, or thinner, than the tails of the reference distribution to which it is being compared.

To illustrate, Figure A3.2 shows a QQ plot for a data sample drawn from a normal distribution, compared to a reference distribution that is also normal. The QQ plot is obviously close to linear: the central mass observations fit a linear QQ plot very closely, and the extreme tail observations somewhat less so. However, there is no denying that the overall plot is approximately linear. Figure A3.2 is a classic example of a QQ plot in which the empirical distribution matches the reference population.

By contrast, Figure A3.3 shows a good example of a QQ plot where the empirical distribution does not match the reference population. In this case, the data are drawn from a Student-t with 5 degrees of freedom, but the reference distribution is standard normal. The QQ plot is now clearly non-linear: although the central mass observations are close to linear, the tails show steeper slopes indicative of the tails being heavier than those of the reference distribution.

A QQ plot is useful in a number of ways. First, as noted already, if the data are drawn from the reference population, then the QQ plot should be linear. This remains true if the data are drawn from some linear transformation of the reference distribution (i.e., are drawn from the same distribution but with different location and scale parameters). We can therefore use a

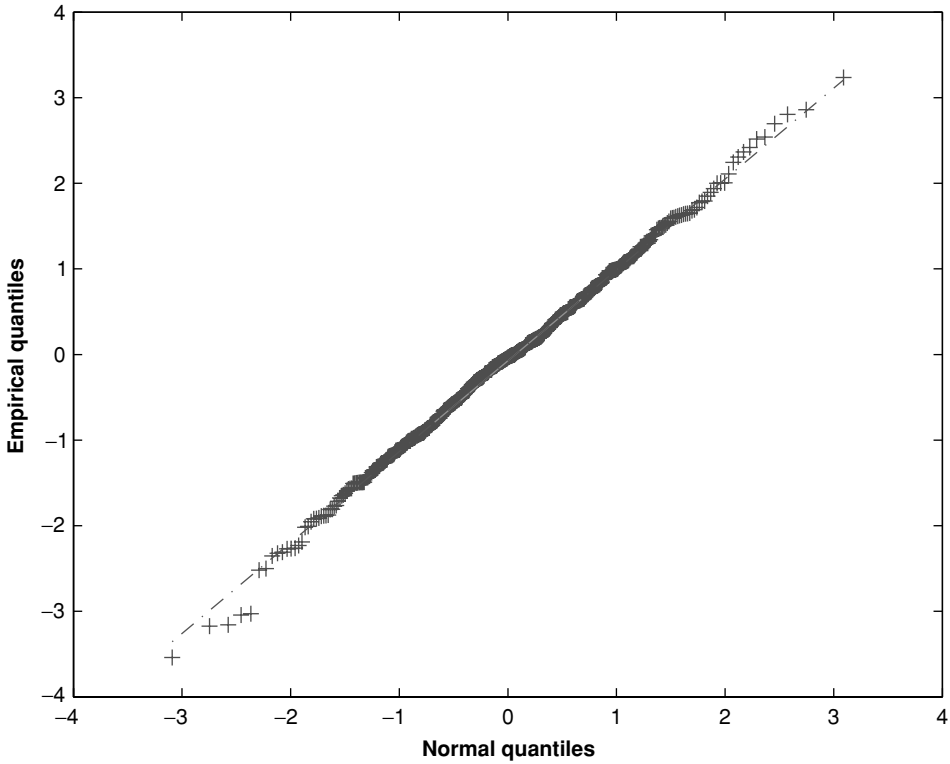

Figure A3.2 QQ plot: normal sample against normal reference distribution
Note: The empirical sample is a random sample of 500 observations drawn from a standard normal. The reference distribution is standard normal.

QQ plot to form a tentative view of the distribution from which our data might be drawn: we specify a variety of alternative distributions, and construct QQ plots for each. Any reference distributions that produce non-linear QQ plots can then be dismissed, and any distribution that produces a linear QQ plot is a good candidate distribution for our data.

Second, because a linear transformation in one of the distributions in a QQ plot merely changes the intercept and slope of the QQ plot, we can use the intercept and slope of a linear QQ plot to give us a rough idea of the location and scale parameters of our sample data. For example, the reference distribution in Figure A3.2 is a standard normal. The linearity of the QQ plot in this figure suggests that the data are normal, as mentioned already, but Figure A3.2 also shows that the intercept and slope are approximately 0 and 1 respectively, and this indicates that the data are drawn from a standard normal, and not just any normal. Such rough approximations give us a helpful yardstick against which we can judge more 'sophisticated' estimates of location and scale, and also provide useful initial values for iterative algorithms.

Third, if the empirical distribution has heavier tails than the reference distribution, the QQ plot will have steeper slopes at its tails, even if the central mass of the empirical observations are approximately linear. Figure A3.3 is a good case in point. A QQ plot where the tails have slopes different than the central mass is therefore suggestive of the empirical distribution having heavier, or thinner, tails than the reference distribution.

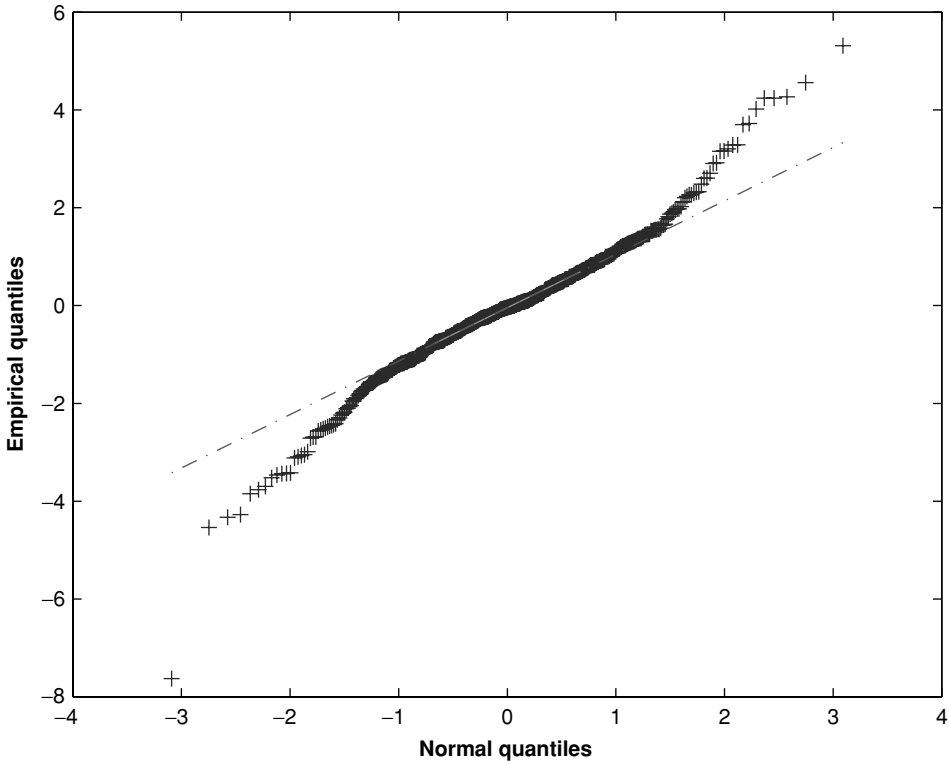

Figure A3.3 QQ plot: *t* sample against normal reference distribution
Note: The empirical sample is a random sample of 500 observations drawn from Student-*t* with 5 degrees of freedom. The reference distribution is standard normal.

Finally, a QQ plot is good for identifying outliers (e.g., observations contaminated by large errors): such observations will stand out in a QQ plot, even if the other observations are broadly consistent with the reference distribution.[10]

[10] Another useful tool, especially when dealing with the tails, is the mean excess function (MEF): the expected amount by which a random variable X exceeds some threshold u, given that $X > u$. The usefulness of the MEF stems from the fact that each distribution has its own distinctive MEF. A comparison of the empirical MEF with the theoretical MEF associated with some specified distribution function therefore gives us an indication of whether the chosen distribution fits the tails of our empirical distribution. However, the results of MEF plots need to be interpreted with some care, because data observations become more scarce as X gets larger. For more on these and how they can be used, see Embrechts *et al.* (1997, Chapters 3.4 and 6.2).

Appendix 2
Numerical Integration Methods

It is sometimes the case that we wish to compute the value of an integral

$$\int_a^b f(x)dx \tag{A3.1}$$

between some bounds a and b. The value of this integral might be a risk measure, such as a coherent risk measure:

$$M_\phi = \int_0^1 \phi(p)q_p dp \tag{A3.2}$$

for some appropriate weighting function $\phi(p)$. In this case, $\phi(p)q_p = f(x)$ and p replaces x as the random variable. Alternatively, the integral might represent the value of a financial instrument, in which case the probability measure used would be a risk-neutralised or distorted probability measure.

If $f(x)$ is straightforward, we can calculate the integral analytically. However, in most practical situations this will not be possible, and in these cases we need to resort to numerical integration (sometimes also known as numerical quadrature). These methods are based on the principle that we can approximate Equation (A3.1) by a discrete sum

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i) \tag{A3.3}$$

where x_i (sometimes also called quadrature nodes) are discrete points taken over the range $[a,b]$ and w_i is some set of weights. The right-hand side of Equation (A3.3) is straightforward to calculate, and the approximation should be fairly close to the true value if the w_i are suitably chosen and n sufficiently large. Our problem now reduces to selecting the w_i and choosing n.

There are many methods we can use to solve this problem. Some of the most popular are Newton–Cotes methods, which seek to approximate the value of the integrand $f(x)$ using low-order polynomials. The most basic of these is the trapezoidal rule, which approximates $f(x)$ linearly, implying that the area under $f(x)$ is approximated by a series of trapezoids. To apply this rule, we divide $[a,b]$ into $n-1$ equal intervals of length h. Thus:

$$h = \frac{b-a}{n-1} \tag{A3.4a}$$

$$x_i = a + (i-1)h \text{ for } i = 1, \ldots, n \tag{A3.4b}$$

The trapezoidal approximation implies the following weights:

$$w_1 = w_n = h/2; w_i = h \text{ for } i = 2, \ldots, n-1 \tag{A3.5}$$

This algorithm is clearly easy to program. It is also robust, and if the integrand is smooth, produces approximation errors of order $O(h^2)$.

This leaves the question of how to choose n. One way to deal with this issue is to estimate the 'halving error' – the change in the estimate resulting from a halving of h (or more or less equivalently, a doubling of n). For the trapezoidal rule it can be shown that the halving error is approximately:

$$\varepsilon_{h/2} \approx \frac{1}{3}(J_{h/2} - J_h) \tag{A3.6}$$

where J_h is an estimate based on a width of h. (Naturally, Equation (A3.6) can also be expressed in terms of n rather than h.) This formula is very useful, because it allows us to evaluate the accuracy of an estimate by seeing how much it changes if we half h. Given that J_h converges to the true value as $h \to 0$, we can use it to select a value for h (and equivalently, for n) that gives us an estimate of acceptable accuracy.

An alternative is Simpson's rule, which uses a quadratic rather than piecewise linear approximation to $f(x)$. This leads to Equation (A3.5) being replaced with:

$$w_1 = w_n = h/3; w_i = 4h/3 \text{ for even } i; w_i = 2h/3 \text{ for odd } i \tag{A3.7}$$

where n is odd. (We can make straightforward modifications for even n.) For a smooth integrand, Simpson's rule leads to an approximation error of order $O(h^4)$ and the following halving error approximation:

$$\varepsilon_{h/2} \approx \frac{1}{15}(J_{h/2} - J_h) \tag{A3.8}$$

Simpson's rule is therefore to be preferred with smooth functions because it is more accurate, although the trapezoidal rule is sometimes more accurate when the integrand is discontinuous.

We could also use many other methods. We could use Newton–Cotes methods with higher order polynomials, Gaussian quadrature methods (in which the nodes and weights are chosen to satisfy 'moment-matching' conditions), Monte Carlo integration (in which we simulate large numbers of x using pseudo-random number methods and take the integral as the average value; we have more to say on Monte Carlo methods in Chapter 8), and quasi-Monte Carlo methods (which are similar to Monte Carlo methods, except that we use low-discrepancy or quasi-random numbers instead of pseudo-random number ones). Depending on the particular problem, these methods may (or may not) be superior to the trapezoidal and Simpson's rules. For example, Miranda and Fackler (2002, p. 88) suggest that a particular Gaussian quadrature method, the Gauss–Legendre method, is the method of choice when $f(x)$ possesses continuous derivatives, but can perform poorly if this is not the case. The bottom line is that one has to take care to choose a method that works well for the particular problem at hand.[11]

[11] For more on these methods, see, e.g., Borse (1997, Chapters 7, 11 and 12), Miranda and Fackler (2002, Chapter 5) or Kreyszig (1999, Chapter 17).

4

Non-parametric Approaches

This chapter looks at some of the most popular approaches to the estimation of risk measures – the non-parametric approaches, which seek to estimate risk measures without making strong assumptions about the relevant (e.g., P/L) distribution. The essence of these approaches is that we try to let the P/L data speak for themselves as much as possible, and use the recent *empirical* (or in some cases simulated) distribution of P/L – not some assumed theoretical distribution – to estimate our risk measures. All non-parametric approaches are based on the underlying assumption that the near future will be sufficiently like the recent past that we can use the data from the recent past to forecast risks over the near future – and this assumption may or may not be valid in any given context. In deciding whether to use any non-parametric approach, we must make a judgement about the extent to which data from the recent past are likely to give us a good guide about the risks we face over the horizon period we are concerned with.

To keep the discussion as clear as possible, we will focus on the estimation of non-parametric VaR and ES. However, the methods discussed here extend very naturally to the estimation of coherent and other risk measures as well. These can be estimated using an 'average quantile' approach along the lines discussed in section 3.4 of Chapter 3: we would select our weighting function $\phi(p)$, decide on the number of probability 'slices' n to take, estimate the associated quantiles, and take the weighted average using an appropriate numerical algorithm (see Chapter 3, Appendix 2).[1] We can then obtain standard errors or confidence intervals for our risk measures using suitably modified forms of the methods discussed in Chapter 3, section 3.5.2.

In this chapter we begin by discussing how to assemble the P/L data to be used for estimating risk measures. We then discuss the most popular non-parametric approach – historical simulation (HS). Loosely speaking, HS is a histogram-based approach: it is conceptually simple, easy to implement, very widely used, and has a fairly good historical record. We focus on the estimation of VaR and ES, but as explained in the previous chapter, more general coherent risk measures can be estimated using appropriately weighted averages of any non-parametric VaR estimates. We then discuss refinements to basic HS using bootstrap and kernel methods, and the estimation of VaR or ES curves and surfaces. Section 4.3 discusses how we can estimate confidence intervals for HS VaR and ES. Section 4.4 addresses weighted HS – how we might weight our data to capture the effects of observation age and changing market conditions. These methods introduce parametric formulas (such as GARCH volatility forecasting equations) into the picture, and in so doing convert hitherto non-parametric methods into what are best described as semi-parametric methods. Such methods are very useful because they allow us to retain the broad HS framework while also taking account of ways in which we think that

[1] Nonetheless, there is an important caveat. This method was explained in Chapter 3 in an implicit context where the risk measurer could choose n, and this is sometimes not possible in a non-parametric context. For example, a risk measurer might be working with an n determined by the HS data set, and even where he/she has some freedom to select n, their range of choice might be limited by the data available. Such constraints can limit the degree of accuracy of any resulting estimated risk measures. However, a good solution to such problems is to increase the sample size by bootstrapping from the sample data. (The bootstrap is discussed further in Appendix 2 to this chapter).

the risks we face over the foreseeable horizon period might differ from those in our sample period. Section 4.5 then reviews the main advantages and disadvantages of non-parametric and semi-parametric approaches, and section 4.6 offers some conclusions.

4.1 COMPILING HISTORICAL SIMULATION DATA

The first task is to assemble a suitable P/L series for our portfolio, and this requires a set of historical P/L or return observations on the positions in our current portfolio. These P/Ls or returns will be measured over a particular frequency (e.g., a day), and we want a reasonably large set of historical P/L or return observations over the recent past. Suppose we have a portfolio of n assets, and for each asset i we have the observed return for each of T subperiods (e.g., daily subperiods) in our historical sample period. If $R_{i,t}$ is the (possibly mapped) return on asset i in subperiod t, and if w_i is the amount currently invested in asset i, then the historically simulated portfolio P/L over the subperiod t is:

$$P/L_t = \sum_{i=1}^{n} w_i R_{i,t} \tag{4.1}$$

Equation (4.1) gives us an historically simulated P/L series for our current portfolio, and is the basis of HS VaR and ES. This series will *not* generally be the same as the P/L *actually* earned on our portfolio – because our portfolio may have changed in composition over time or be subject to mapping approximations, and so on. Instead, the historical simulation P/L is the P/L we *would have* earned on our current portfolio had we held it throughout the historical sample period.[2]

As an aside, the fact that multiple positions collapse into one single HS P/L as given by Equation (4.1) implies that it is very easy for non-parametric methods to accommodate high dimensions – unlike the case for some parametric methods (as we shall see in Chapters 6 and 7). With non-parametric methods, there are no problems dealing with variance–covariance matrices, curses of dimensionality, and the like. This means that non-parametric methods will often be the most natural choice for high-dimension problems.

4.2 ESTIMATION OF HISTORICAL SIMULATION VAR AND ES

4.2.1 Basic Historical Simulation

Having obtained our historical simulation P/L data, we can estimate VaR by plotting the P/L (or L/P) on a simple histogram and then reading off the VaR from the histogram. To illustrate, suppose we have 1000 observations in our HS P/L series and we plot the L/P histogram shown in Figure 4.1. If these were daily data, this sample size would be equivalent to four years' daily data at 250 trading days to a year. If we take our confidence level to be 95%, our VaR is given by the x-value that cuts off the upper 5% of very high losses from the rest of the distribution. Given 1000 observations, we can take this value (i.e., our VaR) to be the 51st highest loss value, or 1.704.[3] The ES is then the average of the 50 highest losses, or 2.196.

[2] To be more precise, the historical simulation P/L is the P/L we would have earned over the sample period had we rearranged the portfolio at the end of each trading day to ensure that the amount left invested in each asset was the same as at the end of the previous trading day: we take out our profits, or make up for our losses, to keep the w_i constant from one end-of-day to the next.

[3] As explained in the previous chapter, we can also estimate the HS VaR more directly (i.e., without bothering with the histogram) by using a spreadsheet function that gives us the 51st highest loss value (e.g., the 'Large' command in Excel), or we can sort our losses data with highest losses ranked first, and then obtain the VaR as the 51st observation in our sorted loss data. We could also take our

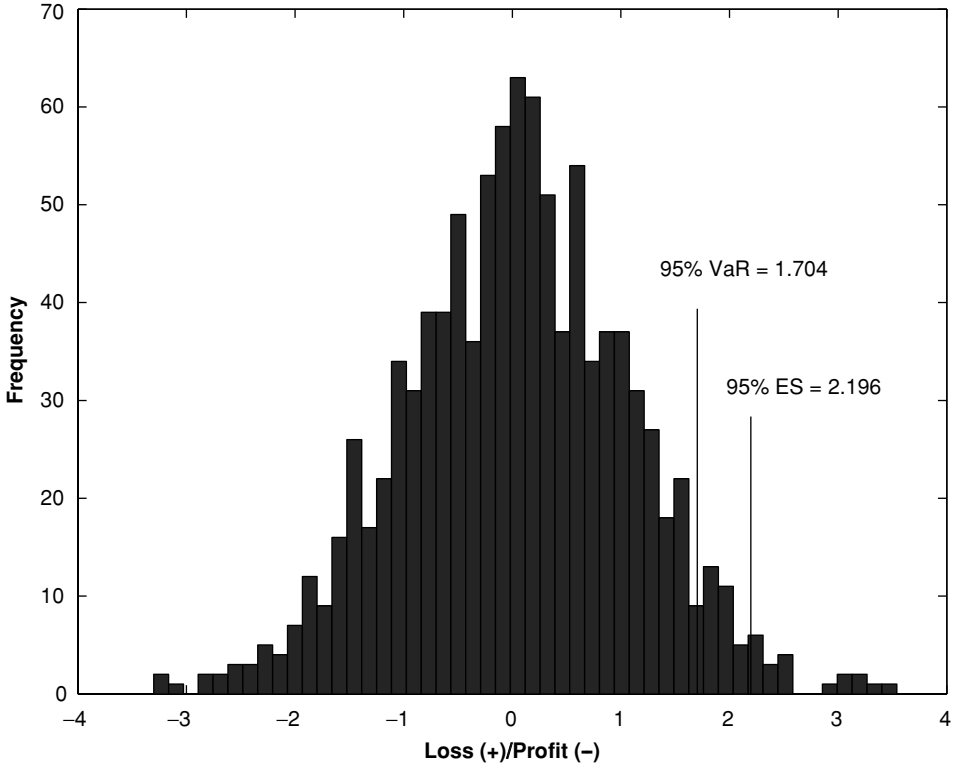

Figure 4.1 Basic historical simulation VaR and ES
Note: This figure and associated VaR and ES estimates are obtained using the 'hsesfigure' function.

The imprecision of these estimates should be obvious when we consider that the sample data set was drawn from a standard normal distribution. In this case the 'true' underlying VaR and ES are 1.645 and 2.063, and Figure 4.1 should (ideally) be normal. Of course, this imprecision underlines the need to work with large sample sizes where practically feasible.

4.2.2 Bootstrapped Historical Simulation

One simple but powerful improvement over basic HS is to estimate VaR and ES from bootstrapped data. As explained in Appendix 2 to this chapter, a bootstrap procedure involves resampling from our existing data set with replacement. The bootstrap is very intuitive and easy to apply. A bootstrapped estimate will often be more accurate than a 'raw' sample estimate, and bootstraps are also useful for gauging the precision of our estimates (see section 4.3.3 below). To apply the bootstrap, we create a large number of new samples, each observation of which is obtained by drawing at random from our original sample and replacing the observation after it has been drawn. Each new 'resampled' sample gives us a new VaR estimate, and

VaR to be any point between the 50th and 51st largest losses (e.g., such as their mid-point), but with a reasonable sample size (as here) there will seldom be much difference between these losses anyway. For convenience, we will adhere throughout to this convention of taking the VaR to be the highest loss observation outside the tail.

we can take our 'best' estimate to be the mean of these resample-based estimates. The same approach can also be used to produce resample-based ES estimates – each one of which would be the average of the losses in each resample exceeding the resample VaR – and our 'best' ES estimate would be the mean of these ES estimates. In our particular case, if we take 1000 resamples, then our best VaR and ES estimates are (because of bootstrap sampling variation) about 1.669 and 2.114 – and the fact that these are much closer to the known true values than our earlier basic HS estimates suggests that bootstraps estimates might be more accurate.

4.2.3 Historical Simulation using Non-parametric Density Estimation

Another potential improvement over basic HS sometimes suggested is to make use of non-parametric density estimation. (This subject is covered in Appendix 3 to this chapter.) To appreciate what this involves, we must recognise that basic HS does not make the best use of the information we have. It also has the practical drawback that it only allows us to estimate VaRs at discrete confidence intervals determined by the size of our data set. For example, if we have 100% HS P/L observations, basic HS allows us to estimate VaR at the 95% confidence level, but not the VaR at the 95.1% confidence level. The VaR at the 95% confidence level is given by the sixth largest loss, but the VaR at the 95.1% confidence level is a problem because there is no corresponding loss observation to go with it. We know that it should be greater than the sixth largest loss (or the 95% VaR), and smaller than the fifth largest loss (or the 96% VaR), but with only 100 observations there is no observation that corresponds to any VaR whose confidence level involves a fraction of 1%. With n observations, basic HS only allows us to estimate the VaRs associated with, at best, n different confidence levels.

Non-parametric density estimation offers a potential solution to both these problems. The idea is to treat our data as if they were drawings from some unspecified or unknown empirical distribution function. This approach also encourages us to confront potentially important decisions about the width of bins and where bins should be centred, and these decisions can sometimes make a difference to our results. Besides using a histogram, we can also represent our data using naïve estimators or, more generally, kernels, and the literature tells us that kernels are (or ought to be) superior. So, having assembled our 'raw' HS data, we need to make decisions on the widths of bins and where they should be centred, and whether to use a histogram, a naïve estimator, or some form of kernel. If we make good decisions on these issues, we can hope to get better estimates of VaR and ES (and more general coherent measures).

Non-parametric density estimation also allows us to estimate VaRs and ESs for any confidence levels we like and so avoid constraints imposed by the size of our data set. In effect, it enables us to draw lines through points on or near the edges of the 'bars' of a histogram. We can then treat the areas under these lines as a surrogate pdf, and so proceed to estimate VaRs for arbitrary confidence levels. The idea is illustrated in Figure 4.2. The left-hand side of this figure shows three bars from a histogram (or naïve estimator) close up. Assuming that the height of the histogram (or naïve estimator) measures relative frequency, then one option is to treat the histogram itself as a pdf. Unfortunately, the resulting pdf would be a strange one – just look at the corners of each bar – and it makes more sense to approximate the pdf by drawing lines through the upper parts of the histogram.

A simple way to do this is to draw in straight lines connecting the mid-points at the top of each histogram bar, as illustrated in the figure. Once we draw these lines, we can forget about the histogram bars and treat the area under the lines as if it were a pdf. Treating the area under the lines as a pdf then enables us to estimate VaRs at any confidence level, regardless of the

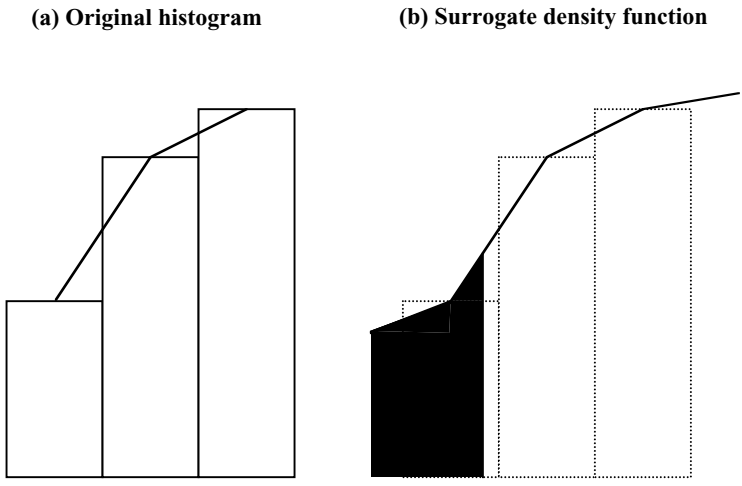

Figure 4.2 Histograms and surrogate density functions

size of our data set. Each possible confidence level would correspond to its own tail similar to the shaded area shown in Figure 4.2(b), and we can then use a suitable calculation method to estimate the VaR (e.g., we can carry out the calculations on a spreadsheet or, more easily, by using a purpose-built function such as the 'hsvar' function in the MMR Toolbox).[4] Of course, drawing straight lines through the mid-points of the tops of histogram bars is not the best we can do: we could draw smooth curves that meet up nicely, and so on. This is exactly the point of non-parametric density estimation, the purpose of which is to give us some guidance on how 'best' to draw lines through the data points we have. Such methods are also straightforward to apply if we have suitable software.

Some empirical evidence by Butler and Schachter (1998) using real trading portfolios suggests that kernel-type methods produce VaR estimates that are a little different to those we would obtain under basic HS. However, their work also suggests that the different types of kernel methods produce quite similar VaR estimates, although to the extent that there are differences among them, they also found that the 'best' kernels were the adaptive Epanechinikov and adaptive Gaussian ones. To investigate these issues myself, I applied four standard kernel estimators – based on normal, box, triangular and Epanechinikov kernels – to the test data used in earlier examples, and found that each of these gave the same VaR estimate of 1.735. In this case, these different kernels produced the same VaR estimate, which is a little higher (and, curiously, a little less accurate) than the basic HS VaR estimate of 1.704 obtained earlier. Other results not reported here suggest that the different kernels can give somewhat different estimates with smaller samples, but again suggest that the exact kernel specification does not make a great deal of difference.

So although kernel methods are better in theory, they do not necessarily produce much better estimates in practice. There are also practical reasons why we might prefer simpler

[4] The actual programming is a little tedious, but the gist of it is that if the confidence level is such that the VaR falls between two loss observations, then we take the VaR to be a weighted average of these two observations. The weights are chosen so that a vertical line drawn through the VaR demarcates the area under the 'curve' in the correct proportions, with α to one side and $1 - \alpha$ to the other. The details can be seen in the coding for the 'hsvar' and related functions.

non-parametric density estimation methods over kernel ones. Although the kernel methods are theoretically better, crude methods like drawing straight-line 'curves' through the tops of histograms are more transparent and easier to check. We should also not forget that our results are subject to a number of sources of error (e.g., due to errors in P/L data, mapping approximations, and so on), so there is a natural limit to how much real fineness we can actually achieve.

4.2.4 Estimating Curves and Surfaces for VaR and ES

It is straightforward to produce plots of VaR or ES against the confidence level. For example, our earlier hypothetical P/L data yields the curves of VaR and ES against the confidence level shown in Figure 4.3. Note that the VaR curve is fairly unsteady, as it directly reflects the randomness of individual loss observations, but the ES curve is smoother, because each ES is an average of tail losses.

It is more difficult constructing curves that show how non-parametric VaR or ES changes with the holding period. The methods discussed so far enable us to estimate the VaR or ES at a single holding period equal to the frequency period over which our data are observed (e.g.,

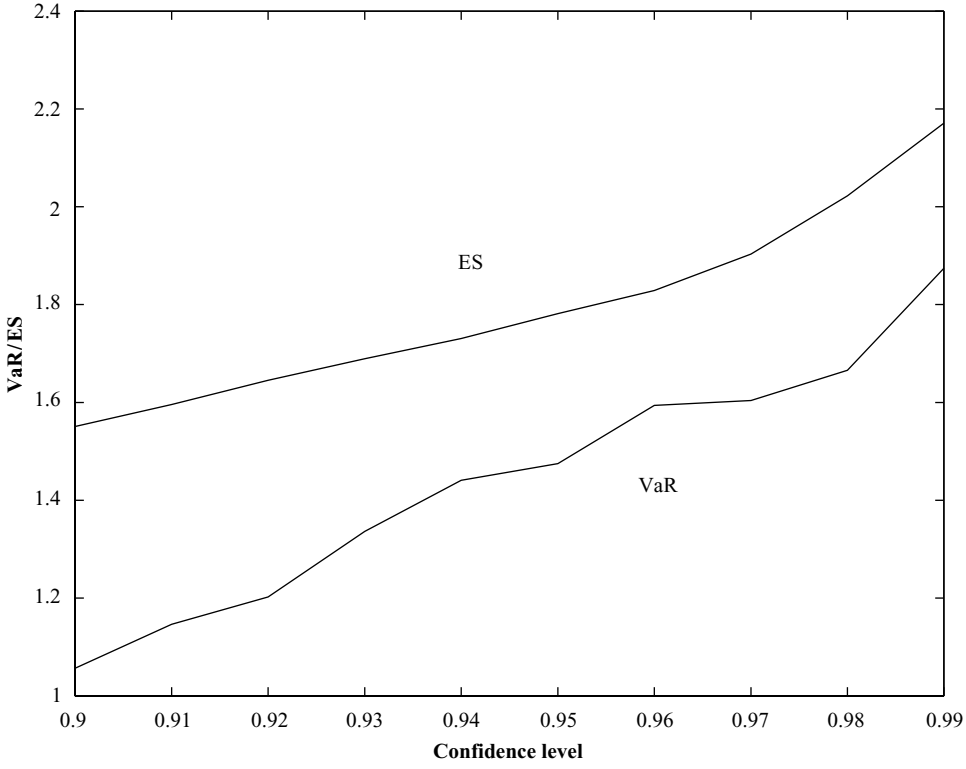

Figure 4.3 Plots of HS VaR and ES against confidence level
Note: Obtained using the 'hsvaresplot2D_cl' function and the same hypothetical P/L data used in Figure 4.1.

they give us VaR or ES for a daily holding period if P/L is measured daily). In theory, we can then estimate VaRs or ESs for any other holding periods we wish by constructing a HS P/L series whose frequency matches our desired holding period: if we wanted to estimate VaR over a weekly holding period, say, we could construct a weekly P/L series and estimate the VaR from that. There is, in short, no theoretical problem as such with estimating HS VaR or ES over any holding period we like.

However, there is a major practical problem: as the holding period rises, the number of observations rapidly falls, and we soon find that we don't have enough data. To illustrate, if we have 1000 observations of daily P/L, corresponding to four years' worth of data at 250 trading days a year, then we have 1000 P/L observations if we use a daily holding period. If we have a weekly holding period, with five days to a week, each weekly P/L will be the sum of five daily P/Ls, and we end up with only 200 observations of weekly P/L; if we have a monthly holding period, we have only 50 observations of monthly P/L; and so on. Given our initial data, the number of effective observations rapidly falls as the holding period rises, and the size of the data set imposes a major constraint on how large the holding period can practically be. In any case, even if we had a very long run of data, the older observations might have very little relevance for current market conditions.

4.3 ESTIMATING CONFIDENCE INTERVALS FOR HISTORICAL SIMULATION VaR AND ES

The methods considered so far are good for giving point estimates of VaR or ES, but they don't give us any indication of the precision of these estimates or any indication of VaR or ES confidence intervals. However, there are methods to get around this limitation and produce confidence intervals for our risk estimates.[5]

4.3.1 An Order-statistics Approach to the Estimation of Confidence Intervals for HS VaR and ES

One of the most promising methods is to apply the theory of order statistics, explained in Appendix 2 to this chapter. This approach gives us, not just a VaR (or ES) estimate, but a complete VaR (or ES) distribution function from which we can read off the VAR (or ES) confidence interval. (The central tendency parameters (mean, mode, median) also give us alternative point estimates of our VaR or ES, if we want them.) This approach is (relatively) easy to programme and very general in its application.

Applied to our earlier P/L data, the OS approach gives us estimates (obtained using the 'hsvarpdfperc' function) of the 5% and 95% points of the 95% VaR distribution function – that is, the bounds of the 90% confidence interval for our VaR – of 1.552 and 1.797. This tells us we can be 90% confident that the 'true' VaR lies in the range [1.552, 1.797].

The corresponding points of the ES distribution function can be obtained (using the 'hses-dfperc' function) by mapping from the VaR to the ES: we take a point on the VaR distribution function, and estimate the corresponding percentile point on the ES distribution function. Doing this gives us an estimated 90% confidence interval of [2.021, 2.224] [6]

[5] In addition to the methods considered in this section, we can also estimate confidence intervals for VaR using estimates of the quantile standard errors, as explained in Chapter 3. However, as made clear there, such confidence intervals are subject to a number of problems, and the methods suggested here are usually preferable.

[6] Naturally, the order-statistics approach can be combined with more sophisticated non-parametric density estimation approaches. Instead of applying the OS theory to the histogram or naïve estimator, we could apply it to a more sophisticated kernel estimator, and

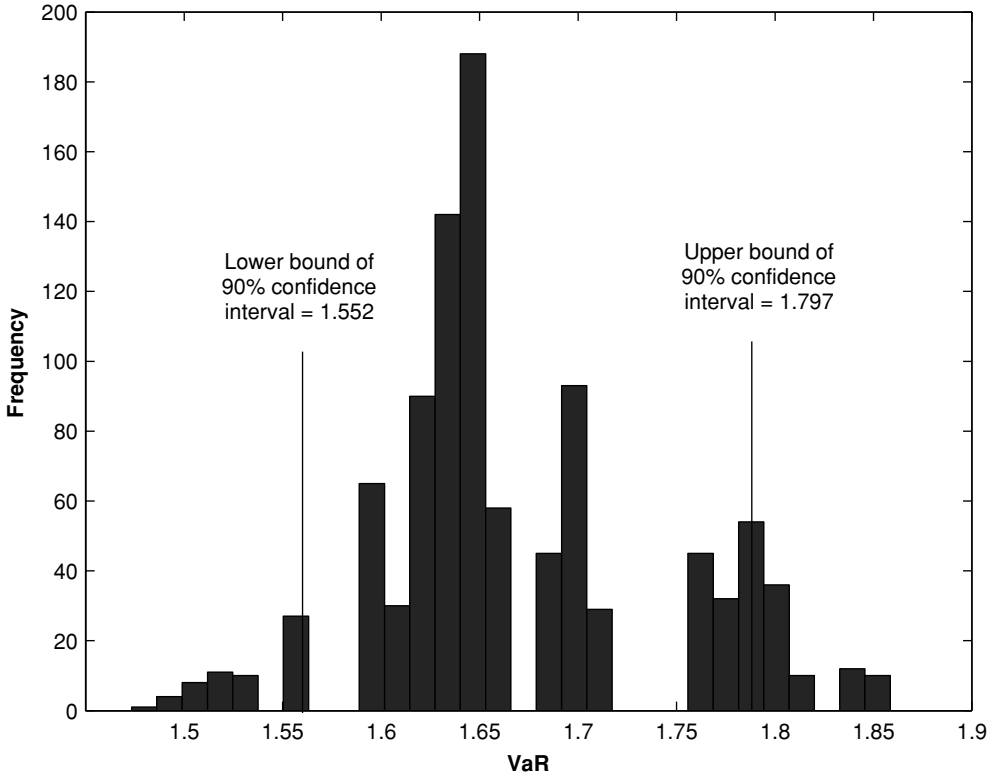

Figure 4.4 Bootstrapped VaR
Note: Results obtained using the 'bootstrapvarfigure' function with 1000 resamples, and the same hypothetical data as in earlier figures.

4.3.2 A Bootstrap Approach to the Estimation of Confidence Intervals for HS VaR and ES

We can also estimate confidence intervals using a bootstrap approach: we produce a bootstrapped histogram of resample-based VaR (or ES) estimates, and then read the confidence interval from the quantiles of this histogram. For example, if we take 1000 bootstrapped samples from our P/L data set, estimate the 95% VaR of each, and then plot them, we get the histogram shown in Figure 4.4. Using the basic percentile interval approach outlined in Appendix 2 to this chapter, the 90% confidence interval for our VaR is [1.554, 1.797]. The simulated histogram is surprisingly disjointed, although the bootstrap seems to give a relatively robust estimate of the confidence interval if we keep repeating the exercise.

We can also use the bootstrap to estimate ESs in much the same way: for each new resampled data set, we estimate the VaR, and then estimate the ES as the average of losses in excess of VaR. Doing this a large number of times gives us a large number of ES estimates, and we can plot them in the same way as the VaR estimates. The histogram of bootstrapped ES

thereby extract more information from our data. This approach has some merit and is developed in detail by Butler and Schachter (1998).

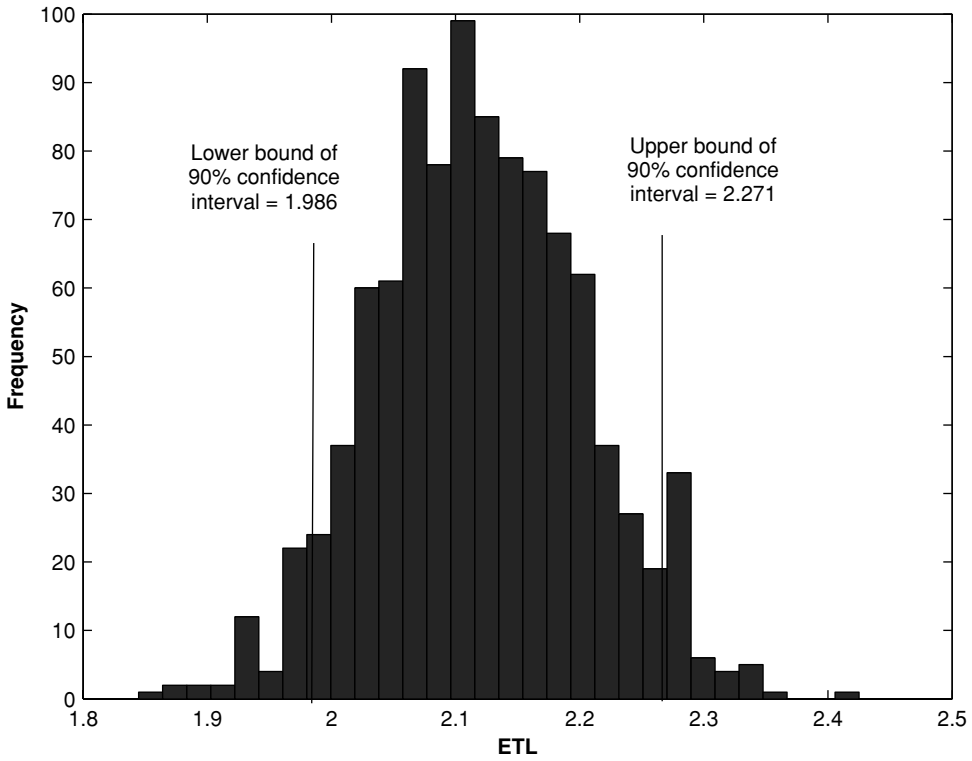

Figure 4.5 Bootstrapped ES
Note: Results obtained using the 'bootstrapesfigure' function with 1000 resamples, and the same hypothetical data as in earlier figures.

values is shown in Figure 4.5, and is better behaved than the VaR histogram in the last figure because the ES is an average of tail VaRs. The 90% confidence interval for our ES is [1.986, 2.271].

It is also interesting to compare the VaR and ES confidence intervals obtained by the two methods. These are summarised in Table 4.1, and we can see that the OS and bootstrap approaches give very similar results. This suggests that either approach is likely to be a reasonable one to use in practice.

Table 4.1 90% confidence intervals for non-parametric VaR and ES

Approach	Lower bound	Upper bound
	95% VaR	
Order statistics	1.552	1.797
Bootstrap	1.554	1.797
	95% ES	
Order statistics	2.021	2.224
Bootstrap	1.986	2.271

Note: Bootstrap estimates based on 1000 resamples.

4.4 WEIGHTED HISTORICAL SIMULATION

One of the most important features of traditional HS is the way it weights past observations. Recall that $R_{i,t}$ is the return on asset i in period t, and we are implementing HS using the past n observations. An observation $R_{i,t-j}$ will therefore belong to our data set if j takes any of the values $1, \ldots, t-n$, where j is the age of the observation (e.g., so $j = 1$ indicates that the observation is 1 day old, and so on). If we construct a new HS P/L series, P/L_t, each day, our observation $R_{i,t-j}$ will first affect P/L_t, then affect P/L_{t+1}, and so on, and finally affect P/L_{t+n}: our return observation will affect each of the next n observations in our P/L series. Also, other things (e.g., position weights) being equal, $R_{i,t-j}$ will affect each P/L in exactly the same way. But after n periods have passed, $R_{i,t-j}$ will fall out of the data set used to calculate the current HS P/L series, and will thereafter have no effect on P/L. In short, our HS P/L series is constructed in a way that gives any observation the *same* weight on P/L provided it is less than n periods old, and *no* weight (i.e., a zero weight) if it is older than that.

This weighting structure has a number of problems. One problem is that it is hard to justify giving each observation in our sample period the same weight, regardless of age, market volatility, or anything else. A good example of the difficulties this can create is given by Shimko *et al.* (1998). It is well known that natural gas prices are usually more volatile in the winter than in the summer, so a raw HS approach that incorporates both summer and winter observations will tend to average the summer and winter observations together. As a result, treating all observations as having equal weight will tend to underestimate true risks in the winter, and overestimate them in the summer.[7]

The equal-weight approach can also make risk estimates unresponsive to major events. For instance, a stock market crash might have no effect on VaRs except at a very high confidence level, so we could have a situation where everyone might agree that risk had suddenly increased, and yet that increase in risk would be missed by most HS VaR estimates. The increase in risk would only show up later in VaR estimates if the stock market continued to fall in subsequent days – a case of the stable door closing only well after the horse had long since bolted. That said, the increase in risk *would* show up in ES estimates just after the first shock occurred – which is, incidentally, a good example of how ES can be a more informative risk measure than the VaR.[8]

The equal-weight structure also presumes that each observation in the sample period is equally likely and independent of the others over time. However, this 'iid' assumption is unrealistic because it is well known that volatilities vary over time, and that periods of high and low volatility tend to be clustered together. The natural gas example just considered is a good case in point.

It is also hard to justify why an observation should have a weight that suddenly goes to zero when it reaches age n. Why is it that an observation of age $n-1$ is regarded as having a lot of value (and, indeed, the same value as any more recent observation), but an observation of age n is regarded as having no value at all? Even old observations usually have some information

[7] If we have data that show seasonal volatility changes, a solution – suggested by Shimko *et al.* (1998) – is to weight the data to reflect seasonal volatility (e.g., so winter observations get more weight, if we are estimating a VaR in winter).

[8] However, both VaR and ES suffer from a related problem. As Pritsker (2001, p. 5) points out, HS fails to take account of useful information from the upper tail of the P/L distribution. If the stock experiences a series of large falls, then a position that was long the market would experience large losses that should show up, albeit later, in HS risk estimates. However, a position that was short the market would experience a series of large profits, and risk estimates at the usual confidence levels would be completely unresponsive. Once again, we could have a situation where risk had clearly increased – because the fall in the market signifies increased volatility, and therefore a significant chance of losses due to large rises in the stock market – and yet our risk estimates had failed to pick up this increase in risk.

content, and giving them zero value tends to violate the old statistical adage that we should never throw information away.

This weighting structure also creates the potential for ghost effects – we can have a VaR that is unduly high (or low) because of a small cluster of high loss observations, or even just a single high loss, and the measured VaR will continue to be high (or low) until n days or so have passed and the observation has fallen out of the sample period. At that point, the VaR will fall again, but the fall in VaR is only a ghost effect created by the weighting structure and the length of sample period used.

We now address various ways in which we might 'adjust' our data to overcome some of these problems and take account of ways in which current market conditions might differ from those in our sample. These fall under the broad heading of 'weighted historical simulation' and can be regarded as semi-parametric methods because they combine features of both parametric and non-parametric methods.

4.4.1 Age-weighted Historical Simulation

One such approach is to weight the relative importance, of our observations by their age, as suggested by Boudoukh, Richardson and Whitelaw (BRW: 1998). Instead of treating each observation for asset i as having the same implied probability as any other (i.e., $1/n$), we could weight their probabilities to discount the older observations in favour of newer ones. Thus, if $w(1)$ is the probability weight given to an observation 1 day old, then $w(2)$, the probability given to an observation 2 days old, could be $\lambda w(1)$; $w(3)$ could be $\lambda^2 w(1)$; and so on. The λ term is between 0 and 1, and reflects the exponential rate of decay in the weight or value given to an observation as it ages: a λ close to 1 indicates a slow rate of decay, and a λ far away from 1 indicates a high rate of decay. $w(1)$ is set so that the sum of the weights is 1, and this is achieved if we set $w(1) = (1 - \lambda)/(1 - \lambda^n)$. The weight given to an observation i days old is therefore:

$$w(i) = \frac{\lambda^{i-1}(1 - \lambda)}{1 - \lambda^n} \tag{4.4}$$

and this corresponds to the weight of $1/n$ given to any in-sample observation under basic HS.

Our core information – the information inputted to the HS estimation process – is the paired set of P/L values and associated probability weights. To implement age-weighting, we merely replace the old equal weights $1/n$ with the age-dependent weights $w(i)$ given by (4.4). For example, if we are using a spreadsheet, we can order our P/L observations in one column, put their weights $w(i)$ in the next column, and go down that column until we reach our desired percentile. Our VaR is then the negative of the corresponding value in the first column. And if our desired percentile falls between two percentiles, we can take our VaR to be the (negative of the) interpolated value of the corresponding first-column observations.

This age-weighted approach has four major attractions. First, it provides a nice generalisation of traditional HS, because we can regard traditional HS as a special case with zero decay, or $\lambda \to 1$. If HS is like driving along a road looking only at the rear-view mirror, then traditional equal-weighted HS is only safe if the road is straight, and the age-weighted approach is safe if the road bends gently.

Second, a suitable choice of λ can make the VaR (or ES) estimates more responsive to large loss observations: a large loss event will receive a higher weight than under traditional HS, and the resulting next-day VaR would be higher than it would otherwise have been. This not only

means that age-weighted VaR estimates are more responsive to large losses, but also makes them better at handling clusters of large losses.

Third, age-weighting helps to reduce distortions caused by events that are unlikely to recur, and helps to reduce ghost effects. As an observation ages, its probability weight gradually falls and its influence diminishes gradually over time. Furthermore, when it finally falls out of the sample period, its weight will fall from $\lambda^n w(1)$ to zero, instead of from $1/n$ to zero. Since $\lambda^n w(1)$ is less than $1/n$ for any reasonable values of λ and n, then the shock – the ghost effect – will be less than it would be under equal-weighted HS.

Finally, we can also modify age-weighting in a way that makes our risk estimates more efficient and effectively eliminates any remaining ghost effects. Since age-weighting allows the impact of past extreme events to decline as past events recede in time, it gives us the option of letting our sample size grow over time. (Why can't we do this under equal-weighted HS? Because we would be stuck with ancient observations whose information content was assumed never to date.) Age-weighting allows us to let our sample period grow with each new observation, so we never throw potentially valuable information away. This would improve efficiency and eliminate ghost effects, because there would no longer be any 'jumps' in our sample resulting from old observations being thrown away.

However, age-weighting also reduces the effective sample size, other things being equal, and a sequence of major profits or losses can produce major distortions in its implied risk profile. In addition, Pritsker shows that even with age-weighting, VaR estimates can still be insufficiently responsive to changes in underlying risk.[9] Furthermore, there is the disturbing point that the BRW approach is ad hoc, and that except for the special case where $\lambda = 1$, we cannot point to any asset-return process for which the BRW approach is theoretically correct.

4.4.2 Volatility-weighted Historical Simulation

We can also weight our data by volatility. The basic idea – suggested by Hull and White (HW; 1998b) – is to update return information to take account of recent changes in volatility. For example, if the current volatility in a market is 1.5% a day, and it was only 1% a day a month ago, then data a month old understate the changes we can expect to see tomorrow, and this suggests that historical returns would underestimate tomorrow's risks; on the other hand, if last month's volatility was 2% a day, month-old data will overstate the changes we can expect tomorrow, and historical returns would overestimate tomorrow's risks. We therefore adjust the historical returns to reflect how volatility tomorrow is believed to have changed from its past values.

Suppose we are interested in forecasting VaR for day T. Let $r_{t,i}$ be the historical return in asset i on day t in our historical sample, $\sigma_{t,i}$ be the historical GARCH (or EWMA) forecast of the volatility of the return on asset i for day t, made at the end of day $t-1$, and $\sigma_{T,i}$ be our most recent forecast of the volatility of asset i. We then replace the returns in our data set, $r_{t,i}$, with volatility-adjusted returns, given by:

$$r_{t,i}^{*} = \left(\frac{\sigma_{T,i}}{\sigma_{t,i}}\right) r_{t,i} \tag{4.5}$$

[9] If VaR is estimated at the confidence level α, the probability of an HS estimate of VaR rising on any given day is equal to the probability of a loss in excess of VaR, which is of course $1-\alpha$. However, if we assume a standard GARCH(1,1) process and volatility is at its long-run mean value, then Pritsker's proposition 2 shows that the probability that HS VaR should increase is about 32% (Pritsker (2001, pp. 7–9)). In other words, most of the time HS VaR estimates should increase (i.e., when risk rises), they fail to.

Actual returns in any period t are therefore increased (or decreased), depending on whether the current forecast of volatility is greater (or less than) the estimated volatility for period t. We now calculate the HS P/L using (4.5) instead of the original data set $r_{t,i}$, and then proceed to estimate HS VaRs or ESs in the traditional way (i.e., with equal weights, etc.).[10]

The HW approach has a number of advantages relative to the traditional equal-weighted and/or the BRW age-weighted approaches:

- It takes account of volatility changes in a natural and direct way, whereas equal-weighted HS ignores volatility changes and the age-weighted approach treats volatility changes in a rather arbitrary and restrictive way.
- It produces risk estimates that are appropriately sensitive to current volatility estimates, and so enables us to incorporate information from GARCH forecasts into HS VaR and ES estimation.
- It allows us to obtain VaR and ES estimates that can exceed the maximum loss in our historical data set: in periods of high volatility, historical returns are scaled upwards, and the HS P/L series used in the HW procedure will have values that exceed actual historical losses. This is a major advantage over traditional HS, which prevents the VaR or ES from being any bigger than the losses in our historical data set.
- Empirical evidence presented by HW indicates that their approach produces superior VaR estimates to the BRW one.

The HW approach is also capable of various extensions. For instance, we can combine it with the age-weighted approach if we wished to increase the sensitivity of risk estimates to large losses, and to reduce the potential for distortions and ghost effects. We can also combine the HW approach with OS or bootstrap methods to estimate confidence intervals for our VaR or ES – that is, we would work with order statistics or resample with replacement from the HW-adjusted P/L, rather than from the traditional HS P/L.

4.4.3 Correlation-weighted Historical Simulation

We can also adjust our historical returns to reflect changes between historical and current correlations. Correlation-weighting is a little more involved than volatility-weighting. To see the principles involved, suppose for the sake of argument that we have already made any volatility-based adjustments to our HS returns along Hull–White lines, but also wish to adjust those returns to reflect changes in correlations.[11]

To make the discussion concrete, we have m positions and our (perhaps volatility adjusted) $1 \times m$ vector of historical returns $\mathbf{R}$ for some period t reflects an $m \times m$ variance–covariance matrix $\boldsymbol{\Sigma}$. $\boldsymbol{\Sigma}$ in turn can be decomposed into the product $\boldsymbol{\sigma}\mathbf{C}\boldsymbol{\sigma}^{\mathbf{T}}$, where $\boldsymbol{\sigma}$ is an $m \times m$ diagonal matrix of volatilities (i.e., so the ith element of $\boldsymbol{\sigma}$ is the ith volatility σ_i and the off-diagonal elements are zero), $\boldsymbol{\sigma}^{\mathbf{T}}$ is its transpose, and $\mathbf{C}$ is the $m \times m$ matrix of historical correlations. $\mathbf{R}$ therefore reflects an historical correlation matrix $\mathbf{C}$, and we wish to adjust $\mathbf{R}$ so that they become $\bar{\mathbf{R}}$ reflecting a current correlation matrix $\bar{\mathbf{C}}$. Now suppose for convenience that both correlation matrices are positive definite. As explained later in Chapter 8, this means that each correlation matrix has an $m \times m$ 'matrix square root', $\mathbf{A}$ and $\bar{\mathbf{A}}$ respectively, given by a Choleski

[10] Naturally, volatility weighting presupposes that one has estimates of the current and past volatilities to work with. We discuss the estimation of volatilities (and covariances and correlations) in Chapter 5.

[11] The correlation adjustment discussed here is based on a suggestion by Duffie and Pan (1997).

decomposition (which also implies that they are easy to obtain). We can now write $\mathbf{R}$ and $\bar{\mathbf{R}}$ as matrix products of the relevant Cholesky matrices and an uncorrelated noise process $\boldsymbol{\varepsilon}$:

$$\mathbf{R} = \mathbf{A}\boldsymbol{\varepsilon} \tag{4.6a}$$

$$\bar{\mathbf{R}} = \bar{\mathbf{A}}\boldsymbol{\varepsilon} \tag{4.6b}$$

We then invert Equation (4.6a) to obtain $\boldsymbol{\varepsilon} = \mathbf{A}^{-1}\mathbf{R}$, and substitute this into (Equation 4.6b) to obtain the correlation-adjusted series $\bar{\mathbf{R}}$ that we are seeking:

$$\bar{\mathbf{R}} = \bar{\mathbf{A}}\mathbf{A}^{-1}\mathbf{R} \tag{4.7}$$

The returns adjusted in this way will then have the currently prevailing correlation matrix $\bar{\mathbf{C}}$ and, more generally, the currently prevailing covariance matrix $\bar{\boldsymbol{\Sigma}}$. This approach is a major generalisation of the HW approach, because it gives us a weighting system that takes account of correlations as well as volatilities.

Example 4.1 (Correlation-weighted HS)

Suppose we have only two positions in our portfolio, so $m = 2$. The historical correlation between our two positions is 0.3, and we wish to adjust our historical returns $\mathbf{R}$ to reflect a current correlation of 0.9.

If a_{ij} is the i, jth element of the 2×2 matrix $\mathbf{A}$, then applying the Cholesky decomposition tells us that

$$a_{11} = 1, a_{12} = 0, a_{21} = \rho, a_{22} = \sqrt{1 - \rho^2}$$

where $\rho = 0.3$. The matrix $\bar{\mathbf{A}}$ is similar except for having $\rho = 0.9$. Standard matrix theory also tells us that

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22}, -a_{12} \\ -a_{21}, a_{11} \end{bmatrix}$$

Substituting these into Equation (4.7), we find that

$$\begin{aligned} \bar{\mathbf{R}} = \bar{\mathbf{A}}\mathbf{A}^{-1}\mathbf{R} &= \begin{bmatrix} 1, 0 \\ 0.9, \sqrt{1 - 0.9^2} \end{bmatrix} \frac{1}{\sqrt{1 - 0.3^2}} \begin{bmatrix} \sqrt{1 - 0.3^2}, 0 \\ -0.3, 1 \end{bmatrix} \mathbf{R} \\ &= \frac{1}{\sqrt{1 - 0.3^2}} \begin{bmatrix} \sqrt{1 - 0.3^2}, 0 \\ 0.9\sqrt{1 - 0.3^2} - 0.3\sqrt{1 - 0.9^2}, \sqrt{1 - 0.9^2} \end{bmatrix} \mathbf{R} \\ &= \begin{bmatrix} 1, 0 \\ 0.7629, 0.4569 \end{bmatrix} \mathbf{R} \end{aligned}$$

4.4.4 Filtered Historical Simulation

Another promising approach is filtered historical simulation (FHS).[12] This is a form of semi-parametric bootstrap which aims to combine the benefits of HS with the power and

[12] This approach is suggested in Barone-Adesi *et al.* (1998), Barone-Adesi *et al.* (1999), Barone-Adesi and Giannopoulos (2000) and in other papers by some of the same authors.

flexibility of conditional volatility models such as GARCH. It does so by bootstrapping returns within a conditional volatility (e.g., GARCH) framework, where the bootstrap preserves the non-parametric nature of HS, and the volatility model gives us a sophisticated treatment of volatility.

Suppose we wish to use FHS to estimate the VaR of a single-asset portfolio over a 1-day holding period. The first step in FHS is to fit, say, a GARCH model to our portfolio-return data. We want a model that is rich enough to accommodate the key features of our data, and Barone-Adesi and colleagues recommend an asymmetric GARCH, or AGARCH, model. This not only accommodates conditionally changing volatility, volatility clustering, and so on, but also allows positive and negative returns to have differential impacts on volatility, a phenomenon known as the leverage effect. The AGARCH postulates that portfolio returns obey the following process:

$$r_t = \mu + \varepsilon_t \tag{4.8a}$$

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1} + \gamma)^2 + \beta\sigma_{t-1}^2 \tag{4.8b}$$

The daily return in Equation (4.8a) is the sum of a mean daily return (which can often be neglected in volatility estimation) and a random error ε_t. The volatility in Equation (4.8b) is the sum of a constant and terms reflecting last period's 'surprise' and last period's volatility, plus an additional term γ that allows for the surprise to have an asymmetric effect on volatility, depending on whether the surprise term is positive or negative.

The second step is to use the model to forecast volatility for each of the days in a sample period. These volatility forecasts are then divided into the realised returns to produce a set of standardised returns. These standardised returns should be independently and identically distributed (iid), and therefore be suitable for HS.

Assuming a 1-day VaR holding period, the third stage involves bootstrapping from our data set of standardised returns: we take a large number of drawings from this data set, which we now treat as a sample, replacing each one after it has been drawn, and multiply each random drawing by the AGARCH forecast of tomorrow's volatility. If we take M drawings, we therefore get M simulated returns, each of which reflects current market conditions because it is scaled by today's forecast of tomorrow's volatility.

Finally, each of these simulated returns gives us a possible end-of-tomorrow portfolio value, and a corresponding possible loss, and we take the VaR to be the loss corresponding to our chosen confidence level.[13]

We can easily modify this procedure to encompass the obvious complications of a multi-asset portfolio or a longer holding period. If we have a multi-asset portfolio, we would fit a multivariate GARCH (or AGARCH) to the set or vector of asset returns, and we would standardise this vector of asset returns. The bootstrap would then select, not just a standardised portfolio return for some chosen past (daily) period, but the standardised vector of asset returns for the chosen past period. This is important because it means that our simulations would keep any correlation structure present in the raw returns. The bootstrap thus maintains existing correlations, without our having to specify an explicit multivariate pdf for asset returns.

The other obvious extension is to a longer holding period. If we have a longer holding period, we would first take a drawing and use Equation (4.8) to get a return for tomorrow; we

[13] The FHS approach can also be extended easily to allow for the estimation of ES as well as VaR. For more on how this might be done, see Giannopoulos and Tunaru (2004).

would then use this drawing to update our volatility forecast for the day after tomorrow, and take a fresh drawing to determine the return for that day; and we would carry on in the same manner – taking a drawing, updating our volatility forecasts, taking another drawing for the next period, and so on – until we had reached the end of our holding period. At that point we would have enough information to produce a single simulated P/L observation; and we would repeat the process as many times as we wished in order to produce the histogram of simulated P/L observations from which we can estimate our VaR.

FHS has a number of attractions: (i) It enables us to combine the non-parametric attractions of HS with a sophisticated (e.g., GARCH) treatment of volatility, and so take account of changing market volatility conditions. (ii) It is fast, even for large portfolios. (iii) As with the earlier HW approach, FHS allows us to get VaR and ES estimates that can exceed the maximum historical loss in our data set. (iv) It maintains the correlation structure in our return data without relying on knowledge of the variance–covariance matrix or the conditional distribution of asset returns. (iv) It can be modified to take account of autocorrelation or past cross-correlations in asset returns. (v) It can be modified to produce estimates of VaR or ES confidence intervals by combining it with an OS or bootstrap approach to confidence interval estimation.[14] (vi) There is evidence that FHS works well.[15]

Box 4.1 Neural Network Approaches

Neural (or learning) network methods provide a very different approach to the estimation of non-parametric or semi-parametric VaR and ES. Neural networks are adaptive statistical processing models whose structure is analogous to that of the brain. The processing units of these models are known as neurons, which are connected to each other by weighted links. The system is initially assigned arbitrary connection weights, inputs are fed into it, and the network determines the initial outputs; these are then compared to target outputs, and if the outputs are not sufficiently close to the target ones, the weights are revised to reduce the discrepancy; new outputs are then calculated using the new weights and compared to the targets, and if these new outputs are also too far from target, the weights are revised again; the process is then repeated again and again until the outputs are close enough to the target to be acceptable. At this point, the network has been 'trained' and can be used to estimate/forecast probabilities or quantities using a 'live' data set. Neural network approaches are attractive because they do not rely on restrictive parametric assumptions, are extremely flexible, highly adaptive, and respond well to changes in the data set, and their results are robust to misspecification errors.

[14] The OS approach would require a set of paired P/L and associated probability observations, so we could apply this to FHS by using a P/L series that had been through the FHS filter. The bootstrap is even easier, since FHS already makes use of a bootstrap. If we want B bootstrapped estimates of VaR, we could produce, say, $100*B$ or $1000*B$ bootstrapped P/L values; each set of 100 (or 1000) P/L series would give us one HS VaR estimate, and the histogram of M such estimates would enable us to infer the bounds of the VaR confidence interval.

[15] Barone-Adesi and Giannopoulos (2000), p. 17. However, FHS does have problems. In his thorough simulation study of FHS, Pritsker (2001, pp. 22–24) comes to the tentative conclusions that FHS VaR might not pay enough attention to extreme observations or time-varying correlations, and Barone-Adesi and Giannopoulos (2000, p. 18) largely accept these points. A partial response to the first point would be to use ES instead of VaR as our preferred risk measure, and the natural response to the second concern is to develop FHS with a more sophisticated past cross-correlation structure. Pritsker (2001, p. 22) also presents simulation results that suggest that FHS-VaR tends to underestimate 'true' VaR over a 10-day holding period by about 10%, but this finding conflicts with results reported by Barone-Adesi *et al.* (2000) based on real data. The evidence on FHS is thus mixed.

There are many different neural network approaches, and a variety of ways of using them to forecast financial risks. One of these is the radial basis function approach, which is particularly suited to the estimation of probabilities. We can use this approach to estimate the probabilities associated with particular P/L values and, once we have the matching probability and P/L series, we can estimate VaR or ES using standard non-parametric methods. Other neural network methods suitable for VaR or ES estimation are the closed-form integration approach (Gottschling *et al.* (1999)), the normal-mixture approach (Prinzler (1999), Bartlmae and Rauscher (2000)), and the quantile-regression approach (Taylor (2000)). However, the application of neural network approaches to risk estimation is still a relatively new field, and much remains to be done.[16]

4.5 ADVANTAGES AND DISADVANTAGES OF NON-PARAMETRIC METHODS

4.5.1 Advantages

In drawing our discussion to a close, it is perhaps a good idea to summarise the main advantages and disadvantages of non-parametric approaches. The advantages include:

- Non-parametric approaches are intuitive and conceptually simple.
- Since they do not depend on parametric assumptions about P/L, they can accommodate fat tails, skewness, and any other non-normal features that can cause problems for parametric approaches.
- They can in theory accommodate any type of position, including derivatives positions.
- There is a widespread perception among risk practitioners that HS works quite well empirically, although formal empirical evidence on this issue is inevitably mixed.
- They are (in varying degrees, fairly) easy to implement on a spreadsheet.
- Non-parametric methods are free of the operational problems to which parametric methods are subject when applied to high-dimensional problems: no need for covariance matrices, no curses of dimensionality, etc.
- They use data that are (often) readily available, either from public sources (e.g., Bloomberg) or from in-house data sets (e.g., collected as a by-product of marking positions to market).
- They provide results that are easy to report and communicate to senior managers and interested outsiders (e.g., bank supervisors or rating agencies).
- It is easy to produce confidence intervals for non-parametric VaR and ES.
- Non-parametric approaches are capable of considerable refinement and potential improvement if we combine them with parametric 'add-ons' to make them semi-parametric: such refinements include age-weighting (as in BRW), volatility-weighting (as in HW and FHS), and correlation-weighting.

[16] Another promising non-parametric method is context modelling as recently suggested by Denecker *et al.* (2001). As its name suggests, this method estimates VaR using the theory of context modelling, which is a very efficient data compression technique often used in text and image processing. Esentially, this approach builds up a series of contexts or patterns that can be used to forecast the values of future variables based on past observations. Each context describes a state of the world, and we can use past data to determine the relevant contexts. New observations are then used to determine the context class and the VaR can be determined from the asset return distribution for that context class.

4.5.2 Disadvantages

Perhaps their biggest potential weakness is that their results are very (and in most cases, completely) dependent on the historical data set.[17] There are various other related problems:

- If our data period was unusually quiet, non-parametric methods will often produce VaR or ES estimates that are too low for the risks we are actually facing; and if our data period was unusually volatile, they will often produce VaR or ES estimates that are too high.
- Non-parametric approaches can have difficulty handling shifts that take place during our sample period. For example, if there is a permanent change in exchange rate risk, it will usually take time for the HS VaR or ES estimates to reflect the new exchange rate risk. Similarly, such approaches are sometimes slow to reflect major events, such as the increases in risk associated with sudden market turbulence.
- If our data set incorporates extreme losses that are unlikely to recur, these losses can dominate non-parametric risk estimates even though we don't expect them to recur.
- Most (if not all) non-parametric methods are subject (to a greater or lesser extent) to the phenomenon of ghost or shadow effects.
- In general, non-parametric estimates of VaR or ES make no allowance for plausible events that might occur, but did not actually occur, in our sample period.
- Non-parametric estimates of VaR and ES are to a greater or lesser extent constrained by the largest loss in our historical data set. In the simpler versions of HS, we cannot extrapolate from the largest historical loss to anything larger that might conceivably occur in the future. More sophisticated versions of HS can relax this constraint, but even so, the fact remains that non-parametric estimates of VaR or ES are still constrained by the largest loss in a way that parametric estimates are not. This means that such methods are not well suited to handling extremes, particularly with small- or medium-sized samples.

However, we can often ameliorate these problems by suitable refinements. For example, we can ameliorate volatility, market turbulence, correlation and other problems by semi-parametric adjustments, and we can ameliorate ghost effects by age-weighting our data and allowing our sample size to rise over time.

There can also be problems associated with the length of the sample window period. We need a reasonably long window to have a sample size large enough to get risk estimates of acceptable precision, and as a broad rule of thumb, most experts believe that we usually need at least a couple of year's worth of daily observations (i.e., 500 observations, at 250 trading days to the year), and often more. On the other hand, a very long window can also create its own problems. The longer the window:

- the greater the problems with aged data;
- the longer the period over which results will be distorted by unlikely-to-recur past events, and the longer we will have to wait for ghost effects to disappear;
- the more the news in current market observations is likely to be drowned out by older observations – and the less responsive will be our risk estimates to current market conditions; and

[17] There can also be problems getting the data set. We need time series data on all current positions, and such data are not always available (e.g., if the positions are in emerging markets). We also have to ensure that data are reliable, compatible, and delivered to the risk estimation system on a timely basis.

- the greater the potential for data-collection problems. This is a particular concern with new or emerging market instruments, where long runs of historical data don't exist and are not necessarily easy to proxy.

4.6 CONCLUSIONS

Non-parametric methods are widely used and in many respects highly attractive approaches to the estimation of financial risk measures. They have a reasonable track record and are often superior to parametric approaches based on simplistic assumptions such as normality. They are also capable of considerable refinement to deal with some of the weaknesses of more basic non-parametric approaches. As a general rule, they work fairly well if market conditions remain reasonably stable, and are capable of considerable refinement. However, they have their limitations and it is often a good idea to supplement them with other approaches. Wherever possible, we should also complement non-parametric methods with stress testing to gauge our vulnerability to 'what if' events. We should never rely on non-parametric methods alone.

Appendix 1
Estimating Risk Measures with Order Statistics

The theory of order statistics is very useful for risk measurement because it gives us a practical and accurate means of estimating the distribution function for a risk measure – and this is useful because it enables us to estimate confidence intervals for them.

A4.1 USING ORDER STATISTICS TO ESTIMATE CONFIDENCE INTERVALS FOR VaR

If we have a sample of n P/L observations, we can regard each observation as giving an estimate of VaR at an implied confidence level. For example, if $n = 1000$, we might take the 95% VaR as the negative of the 51st smallest P/L observation, we might take the 99% VaR as the negative of the 11th smallest, and so on. We therefore take the α VaR to be equal to the negative of the rth lowest observation, where r is equal to $100(1 - \alpha) + 1$. More generally, with n observations, we take the VaR as equal to the negative of the rth lowest observation, where $r = n(1 - \alpha) + 1$.

The rth order statistic is the rth lowest (or, alternatively, highest) in a sample of n observations, and the theory of order statistics is well established in the statistical literature. Suppose our observations $x_1, x_2, \ldots, x_n$ come from some known distribution (or cumulative density) function $F(x)$, with rth order statistic $x_{(r)}$. Now suppose that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. The probability that j of our n observations do not exceed a fixed value x must obey the following binomial distribution:

$$\Pr\{j \text{ observations} \leq x\} = \binom{n}{j} \{F(x)\}^j (1 - F(x))^{n-j} \tag{A4.1}$$

It follows that the probability that at least r observations in the sample do not exceed x is also a binomial:

$$G_r(x) = \sum_{j=r}^{n} \binom{n}{j} \{F(x)\}^j (1 - F(x))^{n-j} \tag{A4.2}$$

$G_r(x)$ is therefore the distribution function of our order statistic and, hence, of our quantile or VaR.[18]

This VaR distribution function provides us with estimates of our VaR *and* of its associated confidence intervals. The median (i.e., 50 percentile) of the estimated VaR distribution function gives us a natural 'best' estimate of our VaR, and estimates of the lower and upper percentiles of the VaR distribution function give us estimates of the bounds of our VaR confidence interval. This is useful, because the calculations are accurate and easy to carry out on a spreadsheet. Equation (A4.2) is also very general and gives us confidence intervals for *any* distribution function $F(x)$, parametric (normal, t, etc.) or empirical.

[18] See, e.g., Kendall and Stuart (1973), p. 348, or Reiss (1989), p. 20.

Table A4.1 Order-statistics estimates of standard normal 95% VaRs and associated confidence intervals

(a) As n varies

No. of observations	100	500	1000	5000	10 000
Lower bound of confidence interval	1.267	1.482	1.531	1.595	1.610
Median of VaR distribution	1.585	1.632	1.639	1.644	1.644
Standard estimate of VaR	1.645	1.645	1.645	1.645	1.645
Upper bound of confidence interval	1.936	1.791	1.750	1.693	1.679
Width of interval/median	42.2%	18.9%	13.4%	6.0%	4.2%

(b) As VaR confidence level varies (with $n = 500$)

VaR confidence level	0.90	0.95	0.99
Lower bound of confidence interval	1.151	1.482	2.035
Median of VaR distribution	1.274	1.632	2.279
Standard estimate of VaR	1.282	1.645	2.326
Upper bound of confidence interval	1.402	1.791	2.560
Width of interval/median of interval	19.7%	18.9%	23.0%

Notes: The confidence interval is specified at a 90% level of confidence, and the lower and upper bounds of the confidence interval are estimated as the 5 and 95 percentiles of the estimated VaR distribution (Equation (A4.2)).

To use this approach, all we need to do is specify $F(x)$ (as normal, t, etc.), set our parameter values, and use Equation (A4.2) to estimate our VaR distribution function.

To illustrate, suppose we want to apply the order-statistics (OS) approach to estimate the distribution function of a standard normal VaR. We then assume that $F(x)$ is standard normal and use Equation (A4.2) to estimate three key parameters of the VaR distribution: the median or 50 percentile of the estimated VaR distribution, which can be interpreted as an OS estimate of normal VaR; and the 5 and 95 percentiles of the estimated VaR distribution, which can be interpreted as the OS estimates of the bounds of the 90% confidence interval for standard normal VaR.

Some illustrative estimates for the 95% VaR are given in Table A4.1. To facilitate comparison, the table also shows the estimates of standard normal VaR based on the conventional normal VaR formula as explained in Chapter 3. The main results are:

- The confidence interval – the gap between the 5 and 95 percentiles – is quite wide for low values of n, but narrows as n gets larger.
- As n rises, the median of the estimated VaR distribution converges to the conventional estimate.
- The confidence interval is (in this case, a little) wider for more extreme VaR confidence levels than it is for the more central ones.

The same approach can also be used to estimate the percentiles of other VaR distribution functions. If we wish to estimate the percentiles of a non-normal parametric VaR, we replace the normal distribution function $F(x)$ by the non-normal equivalent – the t-distribution function, the Gumbel distribution function, and so on. We can also use the same approach to estimate the confidence intervals for an empirical distribution function (i.e., for historical simulation VaR), where $F(x)$ is some empirical distribution function.

A4.2 USING ORDER STATISTICS TO ESTIMATE CONFIDENCE INTERVALS FOR EXPECTED SHORTFALL

We can also apply an OS approach to the estimation of ES confidence intervals too. We can do so in one of two ways: we can derive an average tail loss (or ES) series from the original P/L series and then apply an OS approach to this ES series; or we can derive the percentiles of the VaR distribution function and take the ES percentiles as the average of losses that exceed the percentiles of the VaR distribution function.

A4.3 REFINING THE BASIC ORDER STATISTICS APPROACH

The OS approach outlined here can also be refined in several ways. One such avenue is to apply a Richardson extrapolation method, which is essentially a weighted average of estimates based on alternative sample sizes (see Chapter 3, Appendix 2). A second, complementary, improvement is to generalise the basic OS approach using *L*-estimators, which are weighted averages of order statistics. The class of *L*-estimators is quite a large one, and includes kernel estimators as special cases. Using *L*-estimators instead of basic order statistics increases the robustness of quantile estimators. *L*-estimators are also simple to construct and are asymptotically normal, and this last property both facilitates the construction of quantile confidence intervals and makes it easier to decompose risks.[19]

A4.4 CONCLUSIONS

The OS approach provides an ideal method for estimating the confidence intervals for our VaRs and ESs. In particular, the OS approach is:

- Completely general, in that it can be applied to any parametric or non-parametric VaR or ES.
- Reasonable even for relatively small samples, because it is not based on asymptotic theory – although it is also the case that estimates based on small samples will also be less accurate, precisely because the samples are small.
- Easy to implement in practice.

The OS approach is also superior to confidence-interval estimation methods based on estimates of quantile standard errors (see Chapter 3, section 3.5.1), because it does not rely on asymptotic theory and or force estimated confidence intervals to be symmetric (which can be a problem for extreme VaRs and ESs). And, finally, the OS approach is capable of considerable refinement (and improvement) using Richardson extrapolation and/or *L*-estimator methods.

[19] For more on *L*-estimators and their properties, see Mausser (2001).

Appendix 2
The Bootstrap

The bootstrap is a simple and useful method for assessing uncertainty in estimation procedures. Its distinctive feature is that it replaces mathematical or statistical analysis with simulation-based resampling from a given data set. It therefore provides a means of assessing the accuracy of parameter estimators without having to resort to strong parametric assumptions or closed-form confidence-interval formulas. The roots of the bootstrap go back a couple of centuries, but the idea only took off in the last three decades after it was developed and popularised by the work of Bradley Efron. It was Efron, too, who first gave it its name, which refers to the phrase 'to pull oneself up by one's bootstraps'. The bootstrap is a form of statistical 'trick', and is therefore very aptly named.

The main purpose of the bootstrap is to assess the accuracy of parameter estimates. The bootstrap is ideally suited for this purpose, as it can provide such estimates without having to rely on potentially unreliable assumptions (e.g., assumptions of normality or large samples).[20] The bootstrap is also easy to use because it does not require the user to engage in any difficult mathematical or statistical analysis. In any case, such traditional methods only work in a limited number of cases, whereas the bootstrap can be applied more or less universally. So the bootstrap is easier to use, more powerful and (as a rule) more reliable than traditional means of estimating confidence intervals for parameters of interest. In addition, the bootstrap can be used to provide alternative 'point' estimates of parameters as well.[21]

A4.5 LIMITATIONS OF CONVENTIONAL SAMPLING APPROACHES

The bootstrap is best appreciated by considering the limitations of conventional sampling approaches. Suppose we have a sample of size n drawn from a population. The parameters of the population distribution are unknown – and, more likely than not, so too is the distribution itself. We are interested in a particular parameter θ, where θ might be a mean, variance (or standard deviation), quantile, or some other parameter. The obvious approach is to estimate θ using a suitable sample estimator – so if θ is the mean, our estimator $\hat{\theta}$ would be the sample mean, if θ is the variance, our estimator $\hat{\theta}$ would be based on some sample variance, and so on. Obtaining an estimator for θ is therefore straightforward, but how do we obtain a confidence interval for it?

To estimate confidence intervals for θ using traditional closed-form approaches requires us to resort to statistical theory, and the theory available is of limited use. For example, suppose we

[20] The bootstrap is also superior to the jackknife, which was often used for similar purposes before the advent of powerful computers. The jackknife is a procedure in which we construct a large number of subsamples from an original sample by taking the original sample and leaving one observation out at a time. For each such subsample, we estimate the parameter of interest, and the jackknife estimator is the average of the subsample-based estimators. The jackknife can also be regarded as an approximation to the bootstrap, but it can provide a very poor approximation when the parameter estimator is a non-smooth function of the data. The bootstrap is therefore more reliable and easier to implement.

[21] The bootstrap also has other uses too. For example, it can be used to relax and check assumptions, to give quick approximations and to check the results obtained using other methods.

wish to obtain a confidence interval for a variance. If we assume that the underlying distribution is normal, then we know that $(n-1)\hat{\sigma}^2/\sigma^2$ is distributed as χ^2 with $n-1$ degrees of freedom, and this allows us to obtain a confidence interval for σ^2. If we denote the α point of this distribution as $\chi^2_{\alpha,n-1}$, then the 90% confidence interval for $(n-1)\hat{\sigma}^2/\sigma^2$ is:

$$\left[\chi^2_{0.05,n-1}, \chi^2_{0.95,n-1}\right] \tag{A4.3}$$

This implies that the 90% confidence interval for σ^2 is:

$$\left[\frac{(n-1)\hat{\sigma}^2}{\chi_{0.95,n-1}}, \frac{(n-1)\hat{\sigma}^2}{\chi_{0.05,n-1}}\right] \tag{A4.4}$$

On the other hand, if we cannot assume that the underlying distribution is normal, then obtaining a confidence interval for σ^2 can become very difficult: the problem is that although we can estimate σ^2 itself, under more general conditions we would often not know the distribution of σ^2, or have expressions for standard errors, and we cannot usually obtain closed-form confidence intervals without them.

We can face similar problems with other parameters as well, such as medians, correlations, and tail probabilities.[22] So in general, closed-form confidence intervals are of limited applicability, and will not apply to many of the situations we are likely to meet in practice.

A4.6 THE BOOTSTRAP AND ITS IMPLEMENTATION

The bootstrap frees us of this type of limitation, and is also much easier to implement. It enables us to estimate a confidence interval for any parameter that we can estimate, regardless of whether we have any formulas for the distribution function for that parameter or for the standard error of its estimator. The bootstrap also has the advantage that it comes with less baggage, in the sense that the assumptions needed to implement the bootstrap are generally less demanding than the assumptions needed to estimate confidence intervals using more traditional (i.e., closed-form) methods.

The basic bootstrap procedure is very simple.[23] We start with a given original sample of size n.[24] We now draw a new random sample of the same size from this original sample, taking care to replace each chosen observation back in the sample pool after it has been drawn. (This random sampling, or resampling, is the very heart of the bootstrap. It requires that we have a uniform random number generator to select a random number between 1 and n, which determines the particular observation that is chosen each time.) When constructing the new sample, known as a resample, we would typically find that some observations get chosen more than once, and others don't get chosen at all: so the resample would typically be different from the original one, even though every observation included in it was drawn from the original sample. Once we have our resample, we use it to estimate the parameter we are interested in. This gives us a resample estimate of the parameter. We then repeat the 'resampling' process again and again, and obtain a set of B resample parameter estimates. This set of B resample estimates can also be regarded as a bootstrapped sample of parameter estimates.

[22] However, in the case of quantiles, we can use order statistics to write down their distribution functions.

[23] This application of the bootstrap can be described as a non-parametric one because we bootstrap from a given data sample. The bootstrap can also be implemented parametrically, where we bootstrap from the assumed distribution. When used in parametric mode, the bootstrap provides more accurate answers than textbook formulas usually do, and it can provide answers to problems for which no textbook formulas exist. The bootstrap can also be implemented semi-parametrically and a good example of this is the FHS approach discussed in Chapter 4.

[24] In practice, it might be possible to choose the value of n, but we will assume for the sake of argument that n is given.

We can then use the bootstrapped sample to estimate a confidence interval for our parameter θ. For example, if each resample i gives us a resample estimator $\hat{\theta}^B(i)$ we might construct a simulated density function from the distribution of our $\hat{\theta}^B(i)$ values and infer the confidence intervals from its percentile points. If our confidence interval spans the central $1 - 2\alpha$ of the probability mass, then it is given by:

$$\textit{confidence interval} = [\hat{\theta}^B_{\alpha}, \hat{\theta}^B_{1-\alpha}] \tag{A4.5}$$

where $\hat{\theta}^B_{\alpha}$ is the α quantile of the distribution of bootstrapped $\hat{\theta}^B(i)$ values. This 'percentile interval' approach is very easy to apply and does not rely on any parametric theory, asymptotic or otherwise.

Nonetheless, this basic percentile interval approach is limited itself, particularly if parameter estimators are biased. It is therefore often better to use more refined percentile approaches, and perhaps the best of these is the bias-corrected and accelerated (or BC_a) approach, which generates a substantial improvement in both theory and practice over the basic percentile interval approach. To use this approach we replace the α and $1 - \alpha$ subscripts in Equation (A4.5) with α_1 and α_2, where

$$\alpha_1 = \Phi\left(\hat{z}^0 + \frac{\hat{z}^0 + z_{\alpha}}{1 - \hat{a}(\hat{z}^0 + z_{\alpha})}\right), \alpha_2 = \Phi\left(\hat{z}^0 + \frac{\hat{z}^0 + z_{1-\alpha}}{1 - \hat{a}(\hat{z}^0 + z_{1-\alpha})}\right) \tag{A4.6}$$

If the parameters $\hat{a}$ and $\hat{z}^0$ are zero, this BC_a confidence interval will coincide with the earlier percentile interval. However, in general, they will not be 0, and we can think of the BC_a method as correcting the end-points of the confidence interval. The parameter $\hat{a}$ refers to the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter θ, and it can be regarded as a correction for skewness. This parameter can be estimated from the following, which would be based on an initial bootstrap or jackknife exercise:

$$\hat{a} = \frac{\sum_{i=1}^{M} (\hat{\theta} - \hat{\theta}^B(i))^3}{6\left[\sum_{i=1}^{M} (\hat{\theta} - \hat{\theta}^B(i))^2\right]^{3/2}} \tag{A4.7}$$

The parameter $\hat{z}^0$ can be estimated as the standard normal inverse of the proportion of bootstrap replications that is less than the original estimate $\hat{\theta}$. The BC_a method is therefore (relatively) straightforward to implement, and it has the theoretical advantages over the percentile interval approach of being both more accurate and of being transformation-respecting, the latter property meaning that if we take a transformation of θ (e.g., if θ is a variance, we might wish to take its square root to obtain the standard deviation), then the BC_a method will automatically correct the end-points of the confidence interval of the transformed parameter.[25]

We can also use a bootstrapped sample of parameter estimates to provide an alternative point estimator of a parameter that is often superior to the raw sample estimator $\hat{\theta}$. Given that there are B resample estimators, we can take our bootstrapped point estimator $\hat{\theta}^B$ as the sample

[25] For more on BC_a and other refinements to the percentile interval approach, see Efron and Tibshirani (1993, Chapters 14 and 22) or Davison and Hinkley (1997, Chapter 5).

mean of our B $\hat{\theta}^B(i)$ values:[26]

$$\hat{\theta}^B = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}^B(i) \tag{A4.8}$$

Relatedly, we can also use a bootstrap to estimate the bias in an estimator. The bias is the difference between the expectation of an estimator and the quantity estimated (i.e., the bias equals $E[\hat{\theta}] - \theta)$), and can be estimated by plugging Equation (A4.8) and a basic sample estimator $\hat{\theta}$ into the bias equation:

$$\text{bias} = E[\hat{\theta}] - \theta \Rightarrow \text{estimated bias} = \hat{\theta}^B - \hat{\theta} \tag{A4.9}$$

We can use an estimate of bias for various purposes (e.g., to correct a biased estimator, to correct prediction errors, etc.). However, the bias can have a (relatively) large standard error. In such cases, correcting for the bias is not always a good idea, because the bias-corrected estimate can have a larger standard error than the unadjusted, biased, estimate.

The programs to compute bootstrap statistics are easy to write and the most obvious price of the bootstrap, increased computation, is no longer a serious problem.[27]

A4.7 STANDARD ERRORS OF BOOTSTRAP ESTIMATORS

Naturally, bootstrap estimates are themselves subject to error. Typically, bootstrap estimates have little bias, but they often have substantial variance. The latter comes from basic sampling variability (i.e., the fact that we have a sample of size n drawn from our population, rather than the population itself) and from resampling variability (i.e., the fact that we take only B bootstrap resamples rather than an infinite number of them). The estimated standard error for $\hat{\theta}$, $\hat{s}_B$, can be obtained from:

$$\hat{s}_B = \left(\frac{1}{B}\sum_{i=1}^{B}(\hat{\theta}^B(i) - \hat{\theta}^B)^2\right)^{1/2} \tag{A4.10}$$

where $\hat{\theta}^B = (1/B)\sum_{i=1}^{B}\hat{\theta}^B(i)$. $\hat{s}_B$ is of course also easy to estimate. However, $\hat{s}_B$ is itself variable, and the variance of $\hat{s}_B$ is:

$$\text{var}(\hat{s}_B) = \text{var}[E(\hat{s}_B)] + E[\text{var}(\hat{s}_B)] \tag{A4.11}$$

Following Efron and Tibshirani (1993, Chapter 19), this can be rearranged as:

$$\text{var}(\hat{s}_B) = \text{var}\left[\hat{m}_2^{1/2}\right] + E\left[\frac{\hat{m}_2}{4B}\left(\frac{\hat{m}_4}{\hat{m}_2^2} - 1\right)\right] \tag{A4.12}$$

where $\hat{m}_i$ is the ith moment of the bootstrap distribution of the $\hat{\theta}^B(i)$. In the case where θ is the mean, Equation (A4.12) reduces to:

$$\text{var}(\hat{s}_B) = \frac{\hat{m}_4/\hat{m}_2 - \hat{m}_2}{4n^2} + \frac{\sigma^2}{2nB} + \frac{\sigma^2\left(\hat{m}_4/\hat{m}_2^2 - 3\right)}{4n^2B} \tag{A4.13}$$

[26] This basic bootstrap estimation method can also be supplemented by variance-reduction methods (e.g., importance sampling) to improve accuracy at a given computational cost. See Efron and Tibshirani (1993, Chapter 23) or Davison and Hinkley (1997, Chapter 9).

[27] An example of the bootstrap approach applied to VaR is given in Chapter 4, sections 4.2.2 and 4.3.2, the former discussing the bootstrap point estimator, and the latter discussing bootstrapped confidence intervals for VaR.

If the distribution is normal, this further reduces to:

$$\text{var}(\hat{s}_B) = \frac{\sigma^2}{2n^2}\left(1 + \frac{n}{B}\right) \tag{A4.14}$$

We can then set B to reduce $\text{var}(\hat{s}_B)$ to a desired level, and so achieve a target level of accuracy in our estimate of $\hat{s}_B$. However, these results are limited, because Equation (A4.13) only applies to the mean and Equation (A4.14) presupposes normality as well.

We therefore face two related questions: (a) how we can estimate $\text{var}(\hat{s}_B)$ in general? and (b) how can we choose B to achieve a given level of accuracy in our estimate of $\hat{s}_B$? One approach to these problems is to apply brute force: we can estimate $\text{var}(\hat{s}_B)$ using a jackknife-after-bootstrap (in which we first bootstrap the data and then estimate $\text{var}(\hat{s}_B)$ by jackknifing from the bootstrapped data), or by using a double bootstrap (in which we estimate a sample of bootstrapped $\hat{s}_B$ values and then estimate their variance). We can then experiment with different values of B to determine the values of these parameters needed to bring $\text{var}(\hat{s}_B)$ down to an acceptable level.

If we are more concerned about the second problem (i.e., how to choose B), a more elegant approach is the following, suggested by Andrews and Buchinsky (1997). Suppose we take as our 'ideal' the value of $\hat{s}_B$ associated with an infinite number of resamples, i.e., $\hat{s}_\infty$. Let τ be a target probability that is close to 1, and let *bound* be a chosen bound on the percentage deviation of $\hat{s}_B$ from $\hat{s}_\infty$. We want to choose $B = B(bound, \tau)$ such that the probability that $\hat{s}_B$ is within the desired bound is τ:

$$\Pr\left[100\left|\frac{\hat{s}_B - \hat{s}_\infty}{\hat{s}_B}\right| \leq bound\right] = \tau \tag{A4.15}$$

If B is large, then the required number of resamples is approximately

$$B \approx \frac{2500(\kappa - 1)\chi^2_\tau}{bound^2} \tag{A4.16}$$

However, this formula is not operational because κ, the kurtosis of the distribution of $\hat{\theta}^B$, is unknown. To get around this problem, we replace κ with a consistent estimator of κ, and this leads Andrews and Buchinsky to suggest the following three-step method to determine B:

- We initially assume that $\kappa = 3$, and plug this into Equation (A4.16) to obtain a preliminary value of B, denoted by B_0, where

$$B_0 = \text{int}\left(\frac{5000\chi^2_\tau}{bound^2}\right) \tag{A4.17}$$

  and where $\text{int}(a)$ refers to the smallest integer greater than or equal to a.
- We simulate B_0 resamples, and estimate the sample kurtosis of the bootstrapped $\hat{\theta}^B$ values, $\hat{\kappa}$.
- We take the desired number of bootstrap resamples as equal to $\max(B_0, B_1)$, where

$$B_1 \approx \frac{2500(\hat{\kappa} - 1)\chi^2_\tau}{bound^2} \tag{A4.18}$$

This method does not directly tell us what the variance of $\hat{s}_B$ might be, but we already know how to estimate this in any case. Instead, this method gives us something more useful: it tells us how to set B to achieve a target level of precision in our bootstrap estimators, and (unlike

Equations (A4.13) and (A4.14)) it applies for any parameter θ and applies however $\hat{\theta}^B$ is distributed.[28]

A4.8 TIME DEPENDENCY AND THE BOOTSTRAP

Perhaps the main limitation of the bootstrap is that standard bootstrap procedures presuppose that observations are independent over time, and they can be unreliable if this assumption does not hold. Fortunately, there are various ways in which we can modify bootstraps to allow for such dependence:

- If we are prepared to make parametric assumptions, we can model the dependence parametrically (e.g., using a GARCH procedure). We can then bootstrap from the residuals, which should be independent. However, this solution requires us to identify the underlying stochastic model and estimate its parameters, and this exposes us to model and parameter risk.
- An alternative is to use a block approach: we divide sample data into non-overlapping blocks of equal length, and select a block at random. However, this approach can 'whiten' the data (as the joint observations spanning different blocks are taken to be independent), which can undermine our results. On the other hand, there are also various methods of dealing with this problem (e.g., making block lengths stochastic, etc.) but these refinements also make the block approach more difficult to implement.
- A third solution is to modify the probabilities with which individual observations are chosen. Instead of assuming that each observation is chosen with the same probability, we can make the probabilities of selection dependent on the time indices of recently selected observations: so, for example, if the sample data are in chronological order and observation i has just been chosen, then observation $i + 1$ is more likely to be chosen next than most other observations.

[28] This three-step method can also be improved and extended. For example, it can be improved by correcting for bias in the kurtosis estimator, and a similar (although more involved) three-step method can be used to achieve given levels of accuracy in estimates of confidence intervals as well. For more on these refinements, see Andrews and Buchinsky (1997).

Appendix 3
Non-parametric Density Estimation

Density estimation deals with the construction of an estimated density function from observed data. It also deals with associated issues such as how to present a data set, how to investigate its properties (e.g., such as possible bimodality, skewness, etc.), and how to use density estimates as inputs in other tasks. Density estimation comes in two basic forms. The first and most familiar is parametric density estimation, in which we fit a particular (possibly conditional) distribution onto the data: we carry out some preliminary data analysis, select a suitable distribution, estimate its parameters, and fit the relevant curve to the data. The second is non-parametric density estimation, in which we deliberately avoid fitting a curve to the data, but instead let the data speak for themselves as much as possible. In the one case, the emphasis is on *fitting a chosen curve to the data*; in the other, the emphasis is on *coaxing information out of the data*.

Non-parametric density estimation is quite a subtle subject and has given rise to an extensive specialist literature. We begin here with an overview of the main non-parametric density estimators: histograms, naïve estimators, kernel estimators, and adaptive kernel estimators. We then look at the kernel method applied to VaR.

A4.9 NON-PARAMETRIC DENSITY ESTIMATORS: AN OVERVIEW

A4.9.1 The Histogram

The most common way of representing data is the histogram. Given an origin x_0 and a bin width (or bandwidth) h, the bins of the histogram can be defined as the intervals

$$[x_0 + mh, x_0 + (m + 1)h) \tag{A4.19}$$

which have been arbitrarily chosen to be closed on the left (i.e., so data on the left boundary are included in the interval) and open on the right (i.e., so data on the right boundary are excluded from the interval). The histogram itself is then defined as

$$\hat{f}(x) = \frac{1}{nh} \cdot \#X_i \tag{A4.20}$$

where $\#X_i$ is the number of observations in the same bin as x. To construct a histogram, we choose an origin x_0 and a bin width h, and then estimate Equation (A4.20).

However, the choice of both these parameters can make a difference to our results, particularly the bin width. To illustrate the difference the bin width can make, Figure A4.1 gives two examples of the same data represented by histograms of different bin sizes. Figure A4.1(a) shows a histogram with six bins, and gives the impression that the data have a reasonable-looking density with an apparently well-behaved but long right-hand tail. Figure A4.1(b) shows a histogram with 15 bins applied to the same data, and gives a rather different impression: it turns out that the data distribution is bimodal, with most data belonging to a well-defined

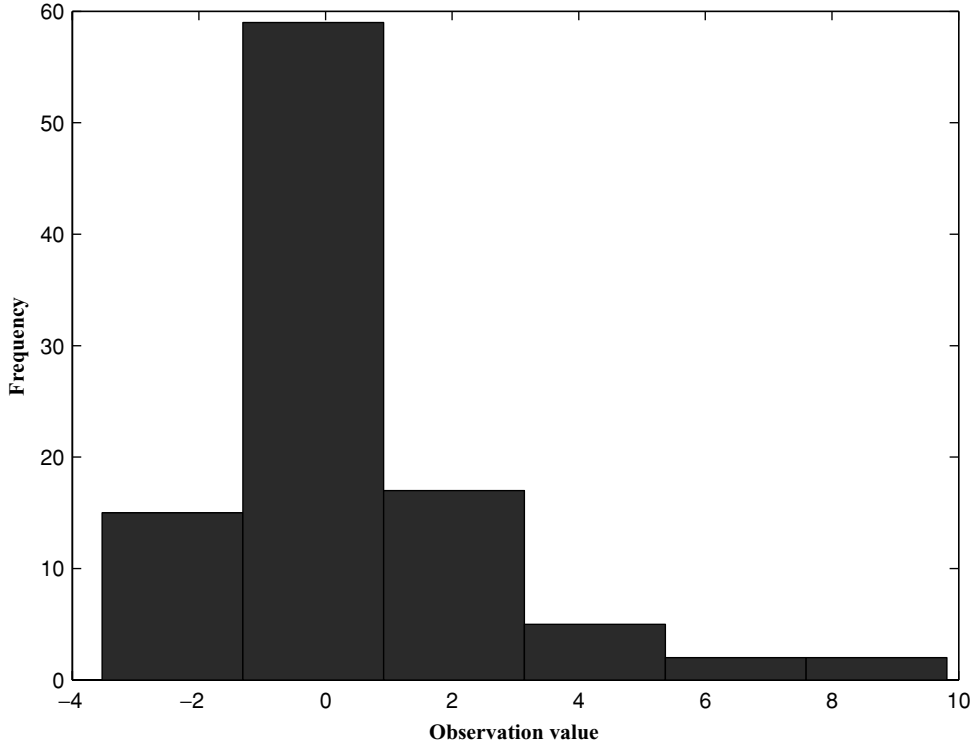

Figure A4.1(a) Histogram with six bins

density mass, and a small number of extreme observations forming its own mini-cluster on the right-hand side. Our earlier 'long tail' turns out to be nothing of the sort. The second histogram is clearly more informative about the tail, but it also shows a fair bit of irrelevant noise that the first histogram has smoothed over. This illustrates the trade-off in choosing bin width: a wider bin width smooths out irrelevant noise in our data, but a bin width that is too wide smooths out valuable information as well.

These examples show that histograms depend on arbitrary judgements, and can be misleading. Histograms can also be inefficient, and the discontinuities of histograms can sometimes cause problems when we try to use them as surrogate density functions in other routines.

A4.9.2 The Naïve Estimator

Fortunately, there are many alternatives to histograms, and one of these is the so-called naïve estimator. The naïve estimator replaces Equation (A4.20) with:

$$\hat{f}(x) = \frac{1}{2nh}.[\text{No. of } X_1, \ldots, X_n \text{ in range}(x - h, x + h)] \tag{A4.21}$$

This looks similar to Equation (A4.20), but does not depend on a choice of origin x_0. The naïve estimator constructs a histogram by treating each observation as falling at the centre of a sampling interval – which is what frees it of the need to specify the origin x_0 – and we can consider it as the sum of 'boxes' centred around our observations. However, the choice of bin

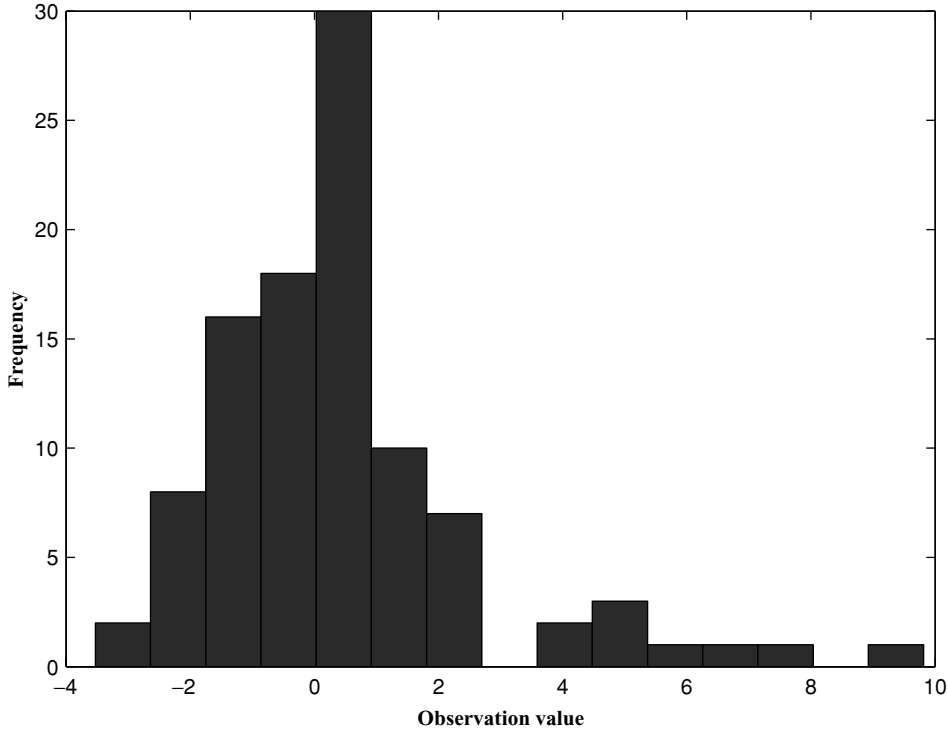

Figure A4.1(b) Histogram with 15 bins

width remains – and can make a big difference to the smoothness of our results – and our estimates will still be discontinuous.

For future reference, it is helpful to express the naïve estimator in terms of weighting functions. If we define the weight function by

$$w(x) = \begin{cases} 1/2 & \text{if} \quad |x| < 1 \\ 0 & \quad\quad\; |x| \geq 1 \end{cases} \tag{A4.22}$$

then we can write the naïve estimator as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w\left(\frac{x - X_i}{h} \right)_i \tag{A4.23}$$

We can then think of the naïve estimator as determined by a particular, rather naïve, choice of weight function.

A4.9.3 The Kernel Estimator

A superior alternative to both the histogram and naïve estimators is the kernel estimator. The kernel is a generalisation of the naïve estimator, and replaces the naïve weight function (Equation (A4.22)) with a kernel function $K(x)$ that satisfies the condition:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \tag{A4.24}$$

$K(x)$ will often be a probability density function, but it can also be a discrete or piecewise function such as (Equation (A4.22)). Given the kernel function, we can define the kernel estimator as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{A4.25}$$

Provided that $K(x)$ is non-negative everywhere, the kernel estimator (Equation (A4.25)) will be well behaved with smooth derivatives, which means that we can treat Equation (A4.25) as if it were a fully fledged probability density function.

We can also consider the kernel estimator as placing probability 'bumps' around each of our observations. The shape of these bumps is determined by the kernel function $K(x)$ and the bandwidth h determines their width. In choosing the kernel function and bandwidth, our objective is to find an estimator $\hat{f}$ that is 'close' in some sense to the true but unknown density function f, and a natural measure of the closeness of fit at a point x is the mean square error (MSE):

$$\begin{aligned} MSE_x(\hat{f}(x)) &= E[\hat{f}(x) - f(x)]^2 = [E\hat{f}(x) - f(x)]^2 + \text{var}(\hat{f}(x)) \\ &= [\text{bias}(\hat{f}(x))^2 + \text{var}(\hat{f}(x)) \end{aligned} \tag{A4.26}$$

where $\text{var}(\hat{f}(x))$ is the variance of $\hat{f}(x)$. The MSE is thus the bias squared plus the variance, which indicates that we face a trade-off between the bias and variance of $\hat{f}(x)$. We can then produce a global measure of the closeness of fit of $\hat{f}(x)$ to $f(x)$ by integrating the MSE over x. This gives us the mean integrated square error (MISE):

$$\begin{aligned} MISE(\hat{f}) &= E\int [\hat{f}(x) - f(x)]^2 dx \\ &= \int [E\hat{f}(x) - f(x)]^2 dx + \int \text{var}(\hat{f}(x))\, dx \end{aligned} \tag{A4.27}$$

so the MISE is the sum of integrated square bias and integrated variance.

We now wish to choose a kernel function to minimise the MISE, and we can show – after a considerable amount of work[29] – that the optimal kernel is the Epanechinikov kernel:

$$K_e(x) = \begin{cases} [3/(4\sqrt{5})](1 - 0.2x^2) & -\sqrt{5} \le x \le \sqrt{5} \\ 0 & \textit{otherwise} \end{cases} \tag{A4.28}$$

However, there are also many other useful kernels that are nearly as efficient as the Epanechinikov kernel. These include the triangular, Gaussian and box kernels, which are respectively:

$$\begin{cases} 1 - |x| & |x| < 1 \\ 0 & \textit{otherwise} \end{cases} \tag{A4.29a}$$

$$(1/\sqrt{2\pi})e^{-0.5x^2} \tag{A4.29b}$$

$$\begin{cases} 1/2 & |x| < 1 \\ 0 & \textit{otherwise} \end{cases} \tag{A4.29c}$$

Any of these ought to produce reasonable results in practice.

[29] See Silverman (1986, pp. 39–40).

Having chosen a kernel, there still remains the issue of how to choose h, the bandwidth. One solution is to choose h subjectively, by plotting out kernel estimates for different bandwidths, and choosing the one that seems right for the data at hand. However, there are also helpful guidance rules for choosing h if we have suitable kernel functions.[30] For example, if we have a Gaussian kernel, the optimal bandwidth is:

$$h_{opt} = 1.06\sigma n^{-1/5} \tag{A4.30}$$

which allows us to estimate the optimal bandwidth directly from the data, using a sample estimate of σ.[31] This bandwidth will be optimal if our data are normally distributed, but can oversmooth non-normal data. To compensate for this, Silverman suggests the modified optimal bandwidth:

$$h = 1.06An^{-1/5} \tag{A4.31}$$

where

$$A = \text{minimum}[\sigma, \text{interquartile range}/1.34] \tag{A4.32}$$

This bandwidth should provide a close fit for a wide range of distributions.

Some illustrative kernels are shown in Figure A4.2, applied to a random sample of 100 observations drawn from a standard normal distribution. As we can see, there is not a great deal to choose from between them: any of these kernels would yield much the same VaR.

Box A4.1 Variable Kernel Estimators

The kernel approach is intuitive and straightforward to apply, and its properties are fairly well understood. However, it also suffers from a problem. Since the bandwidth is fixed across the whole sample, a kernel that provides a good degree of smoothing over the central part of the distribution will often leave spurious noise in the tail. But if we smooth this noise, there is a danger that we will oversmooth the more central part of the distribution and eradicate useful information from it: it is difficult to deal with the tail properly without oversmoothing the main part of the distribution.

A solution is to use adaptive methods in which the bandwidth is allowed to depend on how the observations are distributed.[32] One such method is based on the variable kernel estimator:

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{hd_{j,k}}K\left(\frac{x - X_j}{hd_{j,k}}\right)$$

where $d_{j,k}$ is the distance between X_j and the kth nearest of the other data points. The bandwidth of the kernel placed on the observation X_j is proportional to $d_{j,k}$, so more sparse data will have flatter kernels placed on them.

[30] These automatic methods also include least squares and likelihood cross-validation, and the test graph method. For more on these, see Silverman (1986, pp. 48–57).

[31] See Silverman (1986, p. 45).

[32] There are also other solutions. These include: nearest (or more accurately, near) neighbour method, in which the bandwidth applied to an observation depends on the distance between that observation and its near neighbours; maximum penalised likelihood methods, which are maximum likelihood methods adjusted for the roughness of the empirical density function; and orthogonal series estimators, which are Fourier transform methods. For more on these methods, see Silverman (1986, Chapters 2 and 5).

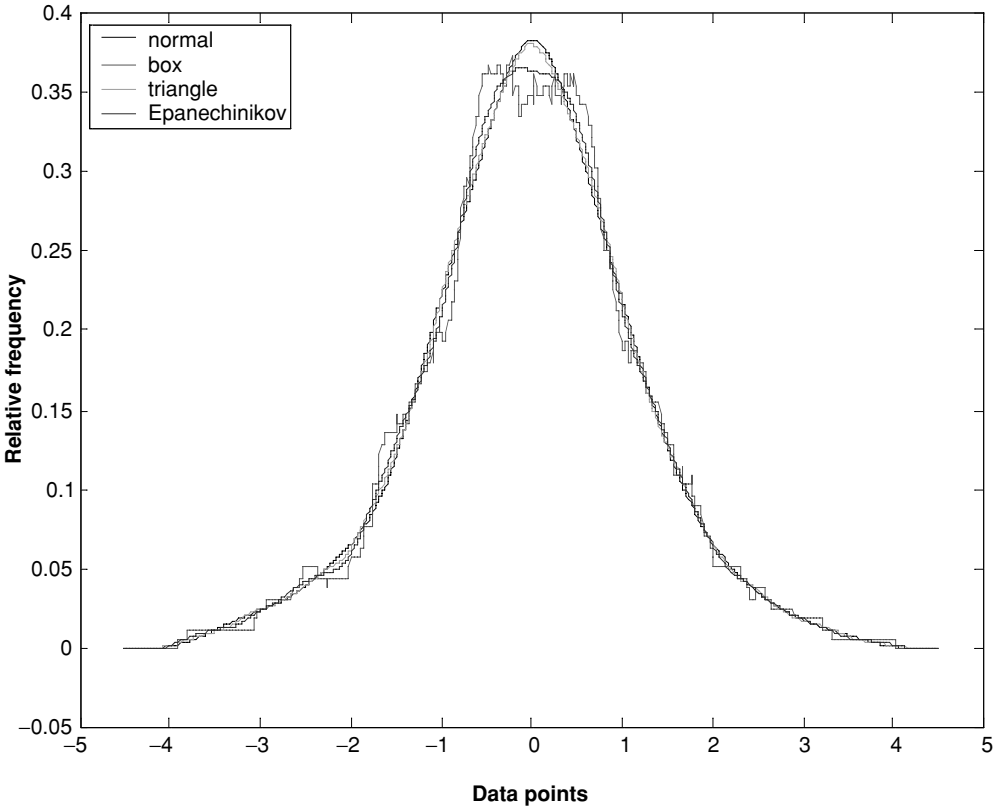

Figure A4.2 Alternative kernel density functions
Note: Estimated using the MATLAB 'ksdensity' Statistical Toolbox function, using Equation (A4.30) to determine the binwidth.

A4.10 KERNEL APPROACHES TO THE ESTIMATION OF RISK MEASURES

Density estimation theory suggests that kernel methods should produce better non-parametric estimates of VaR and ES than those we would obtain under historical simulation. The HS approach is essentially density estimation with a histogram and, as we have seen, a histogram is rarely the best way to handle a data set.

To estimate VaR with kernels, we fit a suitable kernel density function to the data, construct the cumulative density function from the kernel density, and then invert the cumulative density function to obtain the quantile (or percentile) that corresponds to our VaR. Estimates of ES or more general coherent risk measures can be obtained by applying a suitable version of the 'average tail VaR' method to kernel VaR estimates, and confidence intervals for our risk measures can be obtained using bootstrap methods.

Some evidence on how these estimators actually perform is provided by Butler and Schachter (1998). They applied their historical kernel approach to some real trading portfolios, and in those particular cases they found that the adaptive kernel approaches generally led to higher

VaRs, but the choice of kernel otherwise made relatively little difference to VaR estimates. To the extent that the choice of kernel did make a difference, they concluded that the best ones were the adaptive Epanechinikov and adaptive Gaussian kernels.

All said, it is still unclear how much of an improvement we can expect kernel approaches to make over traditional historical simulation: although kernel methods are superior in theory, they are more difficult to use, their results are less transparent and there is in any case a limit to how much accuracy we can realistically obtain given errors in our data and elsewhere in our risk measurement systems.

Appendix 4
Principal Components Analysis and Factor Analysis

This appendix discusses principal components analysis (PCA) and factor analysis (FA), which are methods of dealing with and gaining insights into the characteristics of a multivariate data set. More specifically, they are multidimensional methods that provide a simpler representation of the risk factors present in a data set. They are used to reduce the dimensionality of highly correlated data sets by finding hypothetical variables – which are in fact linear combinations of the original data – that 'explain' a large proportion of the variability of the original data. Such methods are very useful when we have large dimensionality risk measurement problems. Two common examples are the following:

- We might wish to estimate the risk measures of portfolios with hundreds of different assets.
- We might be dealing with fixed-income portfolios, and have 10 or more points on the spot rate curve (or 'yield curve') to consider for each currency denominated in our portfolio.

These methods can also be useful for other tasks, such as cleaning data, developing mapping systems and 'adjusting' covariance or correlation matrices to ensure that they satisfy desirable properties of positive definiteness or positive semi-definiteness (see Chapter 5, section 5.3).

A4.11 PRINCIPAL COMPONENTS ANALYSIS

A4.11.1 Theory

We begin with PCA, and some definitions. Let $\mathbf{\Sigma}$ be a real, symmetric, positive definite matrix of dimension $m \times m$. (By positive definite, we mean that $\mathbf{w\Sigma w^T} > 0$ for any $1 \times m$ vector $\mathbf{w} \neq 0$ and corresponding $m \times 1$ transpose vector $\mathbf{w^T}$. This condition ensures that a portfolio volatility $\mathbf{w\Sigma w^T}$ is always positive.) An eigenvector of $\mathbf{\Sigma}$ is an $m \times 1$ vector $\boldsymbol{\upsilon_i} \neq 0$ that exists if there is a real number λ_i, known as an eigenvalue, such that $\mathbf{\Sigma}\boldsymbol{\upsilon_i} = \boldsymbol{\upsilon_i}\lambda_i$. $\mathbf{\Sigma}$ has rank m, which implies that it has m non-trivially distinct eigenvectors, $\boldsymbol{\upsilon}_i$, each of which has its own positive eigenvalue λ_i. However, any scalar multiple of an eigenvector is also an eigenvector, so a given eigenvector is unique only up to a linear transformation.

Now suppose that $\mathbf{x}$ is an $m \times 1$ random vector (e.g., a vector of returns), with covariance matrix $\mathbf{\Sigma}$, and consider $\mathbf{p}$, an $m \times 1$ linear combination of the individual $\mathbf{x}$-variables produced by premultiplying $\mathbf{x}$ by the $m \times m$ matrix $\mathbf{A}$:

$$\mathbf{p} = \mathbf{Ax} \tag{A4.33}$$

The ith element of $\mathbf{p}$, p_i, is therefore:

$$p_i = \mathbf{A_i x} = \sum_{j=1}^{m} a_{ij} x_j \tag{A4.34}$$

where $\mathbf{A_i}$ is the ith row of $\mathbf{A}$. So p_i is a linear combination or weighted average of the x_j terms, the elements of $\mathbf{x}$. Because the relative weights are not affected if we multiply all weights by

the same constant, we can normalise the weights so that $\mathbf{A_i A_i^T} = 1$. Now note that we can write the variance and covariance terms of $\mathbf{p}$ as:

$$\text{var}(p_i) = \mathbf{A_i \Sigma A_i^T}, \quad i = 1, \ldots, m \tag{A4.35a}$$

$$\text{cov}(p_i, p_j) = \mathbf{A_i \Sigma A_j^T}, \quad i, j = 1, \ldots, m \tag{A4.35b}$$

The idea motivating PCA is to find linear combinations $\mathbf{A_i}$ that maximise the variance of p_i but leave $\text{cov}(p_i, p_j) = 0$ for any $i \neq j$. The first principal component, p_1, is then the linear combination $p_1 = \mathbf{A}_1\mathbf{x}$ that maximises the variance of p_1 subject to the constraint that $\mathbf{A_1 A_1^T} = 1$; the second principal component is the linear combination $p_2 = \mathbf{A}_2\mathbf{x}$ that maximises the variance of p_2 subject to the constraints $\mathbf{A}_2\mathbf{A}_2^{\mathbf{T}} = 1$ and $\text{cov}(p_1, p_2) = 0$; the third principal component is the linear combination $p_3 = \mathbf{A}_3\mathbf{x}$ that maximises the variance of p_3 subject to the constraints $\mathbf{A}_3\mathbf{A}_3^{\mathbf{T}} = 1$, $\text{cov}(p_1, p_3) = 0$ and $\text{cov}(p_2, p_3) = 0$; and so on.

However, we also know that $\mathbf{\Sigma} \boldsymbol{\upsilon_i} = \boldsymbol{\upsilon_i} \lambda_\mathbf{i}$, and this implies:

$$\mathbf{\Sigma}\boldsymbol{\upsilon} = \boldsymbol{\upsilon}\mathbf{\Lambda} \Rightarrow \mathbf{\Sigma}\boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathbf{T}} = \mathbf{\Sigma} = \boldsymbol{\upsilon}\mathbf{\Lambda}\boldsymbol{\upsilon}^{\mathbf{T}} \tag{A4.36}$$

where $\boldsymbol{\upsilon}$ is the $m \times m$ matrix of eigenvectors of $\mathbf{\Sigma}$, $\mathbf{\Lambda}$ is an $m \times m$ diagonal matrix (i.e., a matrix whose off-diagonal terms are zero) whose diagonal elements, $\lambda_1, \lambda_2, \ldots, \lambda_m$, are the eigenvalues of $\mathbf{\Sigma}$, and bearing in mind that we can normalise $\boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathbf{T}}$ to equal the identity matrix $\mathbf{I}$. If we now transform $\mathbf{x}$ via $\mathbf{p} = \boldsymbol{\upsilon}^{\mathbf{T}}\mathbf{x}$, the variance–covariance matrix of $\mathbf{p}$, $\mathbf{VC(p)}$, is then

$$\mathbf{VC(p)} = \mathbf{VC}(\boldsymbol{\upsilon}^{\mathbf{T}}\mathbf{x}) = \boldsymbol{\upsilon}^{\mathbf{T}}\mathbf{\Sigma}\boldsymbol{\upsilon} = \boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathbf{T}}\mathbf{\Lambda}\boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathbf{T}} = \mathbf{\Lambda} \tag{A4.37}$$

Since $\mathbf{\Lambda}$ is a diagonal matrix, Equation (A4.37) tells us that the $\mathbf{p}$ are uncorrelated with each other, and that their variances are given by the diagonal elements of $\mathbf{\Lambda}$, the eigenvalues. It is then clear that these $\mathbf{p}$ are our principal components, and that the $\mathbf{\Lambda}$ matrix we are seeking is the eigenvector matrix $\boldsymbol{\upsilon}$.

In short, $\mathbf{x}$ has a covariance matrix $\mathbf{\Sigma}$, which will typically have non-zero off-diagonal terms, but the principal components of $\mathbf{x}$, $\mathbf{p}$ $(= \mathbf{Ax})$, have a diagonal covariance matrix $\mathbf{\Lambda}$, the diagonal elements of which are the eigenvalues of $\mathbf{x}$.

Thus, the principal components of our m original variables are m artificial variables constructed so that the first principal component 'explains' as much as it can of the variance of these variables; the second principal component 'explains' as much as it can of the remaining variance, but is uncorrelated with the first principal component; the third principal component explains as much as it can of the remaining variance, and is uncorrelated with the first two principal components; and so on. The way the principal components are constructed also implies that the eigenvalues are in declining order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m > 0 \tag{A4.38}$$

The first principal component typically 'explains' more of the variability of our original data than the second principal component, the second typically 'explains' more than the third, and so on. Furthermore, since the principal components are constructed as a linear transformation of the original variables, the complete set of m principal components will explain all (i.e., 100%) of the movement (or total variance) of our original variables. The proportions of the total variance explained by each principal component are therefore:

$$\frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i}, \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i}, \ldots, \frac{\lambda_m}{\sum_{i=1}^{m} \lambda_i} \tag{A4.39}$$

The first principal component explains $\lambda_1 / \sum_{i=1}^{m} \lambda_i$ of the total variance, and so forth.

However, it is important to appreciate that principal components are dependent on the way the original data are measured. If data are measured in different units, our results would depend on the measurement units used, and might be misleading and/or unreliable. In particular, principal components depend on the relative variances of the different series in our data set. If different series have very different variances, then the series with the highest variances will tend to dominate the first one or two principal components, and our results could be severely compromised. Similarly, if we change the metric used to measure a series, its correlations would remain unchanged but the principal components would change because they depend on the relative variances, which would alter. Thus, the PCA results are dependent on the metrics on which our data are specified. It is therefore important to ensure that data are measured in comparable units, and one common way to achieve this comparability is to standardise data by dividing them by their standard deviations. This is equivalent to applying PCA to the correlation matrix instead of the covariance matrix.

It will often be that the first few principal components will explain a very considerable proportion of the total variance of our original variables, and the remaining principal components explain relatively little. The standard financial example is where the original variables might be different spot (or interest) rates across the maturity spectrum, and where the first three principal components are commonly reported to explain over 95% of spot-rate behaviour. In this particular application, the first three principal components also have ready interpretations in terms of interest-rate dynamics: the first can be interpreted as reflecting the level of the spot-rate curve, the second can be interpreted as reflecting its steepness; and the third can be interpreted as reflecting its curvature. We can therefore model the spot rates using only a small number of spot-rate principal components – which shows that PCA can be useful for reducing the dimensionality of a multivariate problem.

There is no magic formula to tell us how many principal components we should use, but one standard (albeit imperfect) criterion is to specify some admissible proportion π of unexplained variance that we are willing to accept. Granted this proportion, we then want the minimum number of principal components that explains at least $1 - \pi$ of the total variance. We therefore choose the first k principal components such that

$$\frac{\lambda_1 + \ldots + \lambda_k}{\sum_{i=1}^{m} \lambda_i} > 1 - \pi \tag{A4.40}$$

So, if our admissible proportion of unexplained variance is 5%, say, we would choose the first k principal components such that the left-hand side of Equation (A4.40) is at least 95%. However, this criterion is somewhat ad hoc, and relies on the sometimes unwarranted assumption that principal components with small variances will have small impacts on key variables. It can therefore be unreliable. In any case, it still does not tell us how many principal components we need to get 'acceptably' accurate estimates of VaR or ES.

In using PCA, we also have to watch out for some pitfalls. One problem is that principal components based on historical data can be unstable; however, we can mitigate this instability using simple rules of thumb, such as taking moving averages of our principal components. We should also be careful about using too many principal components. While it is always true that adding more principal components will increase the fit of our PCA model to the historical data, we are usually more concerned in practice with the predictive ability of our model, and this is

a different matter. This means that we only want to add principal components that represent stable relationships that are good for forecasting, and there will often come a point where additional principal components merely lead to the model tracking noise – and so undermine the forecasting ability of our model. Modellers should therefore be careful not to use too many principal components.

Box A4.2 Principal Components Analysis Reduces the Number of Covariance Terms to be Handled

The diagonality of $\mathbf{\Lambda}$ implies a major reduction in the number of covariance terms we have to deal with. If we have m variables with a covariance matrix $\mathbf{\Sigma}$, then $\mathbf{\Sigma}$ will have $m(m+1)/2$ separate terms – reflecting m diagonal (or variance) and $m(m-1)/2$ off-diagonal (or covariance) terms, but the eigenvalue matrix $\mathbf{\Lambda}$ has only m terms, which is a very considerable reduction if m is large. For example, if $m = 10$, $\mathbf{\Sigma}$ has 55 separate terms, but $\mathbf{\Lambda}$ has 10, so PCA reduces the number of covariance terms by over 80%; and if $m = 100$, $\mathbf{\Sigma}$ has 5050 separate terms, $\mathbf{\Lambda}$ has 100, and PCA reduces the number of covariance terms by over 98%. PCA can lead to very substantial savings in the number of covariance parameters we need to handle.

We get even greater savings if we reduce the dimensionality of our problem and delete 'insignificant' principal components. Thus, if our original covariance matrix is 10×10, but we use only the first three principal components, then the number of covariance terms we have to handle falls from 55 to only three, which is a reduction of almost 95%; and if our original covariance matrix is 100×100 and we again use three principal components, then the number of covariance terms falls by almost 99.95%.

A4.11.2 Principal Components VaR and ES: An Illustration

To give a typical illustration, suppose we have a set of 1000 observations of the returns on each of 10 assets. These returns are randomly drawn from a multivariate standard normal distribution in which all returns have mean 0 and standard deviation 1, and the correlation between the returns to assets i and j is equal to $0.8^{|i-j|}$. We begin with a preliminary analysis of the principal components of our returns, the results of which are presented in Figure A4.3. These indicate that the first principal component explains about 55% of the movement in our data, the second about 20%, the third about 10%, and so on. The first two principal components therefore explain 75% of the total variance, the first three explain 85% of it, and so on.

We now estimate the VaRs and ESs of our portfolio, for the 95% and 99% confidence levels. These results are presented in Figures A4.4 and A4.5 respectively. In this particular case, both PCA VaR and ES results are very stable in absolute terms, especially when the number of principal components is three or more, and (as we might expect) the ES estimates are notably more stable than the VaR ones. Since the 'true' risk measures coincide with the PCA measures when we have 10 principal components, we can also see that three or more principal components – in other words, principal components accounting for as little of 85% of the overall explanatory power of our original returns – generally suffice here to give us risk estimates that are 'close' to the true values: the PCA estimates of the 99% VaR are perhaps a little off for three or so principal components, but the PCA estimates of the 95% VaR are very close, and the PCA ES estimates are extremely so. These examples illustrate the widespread view that

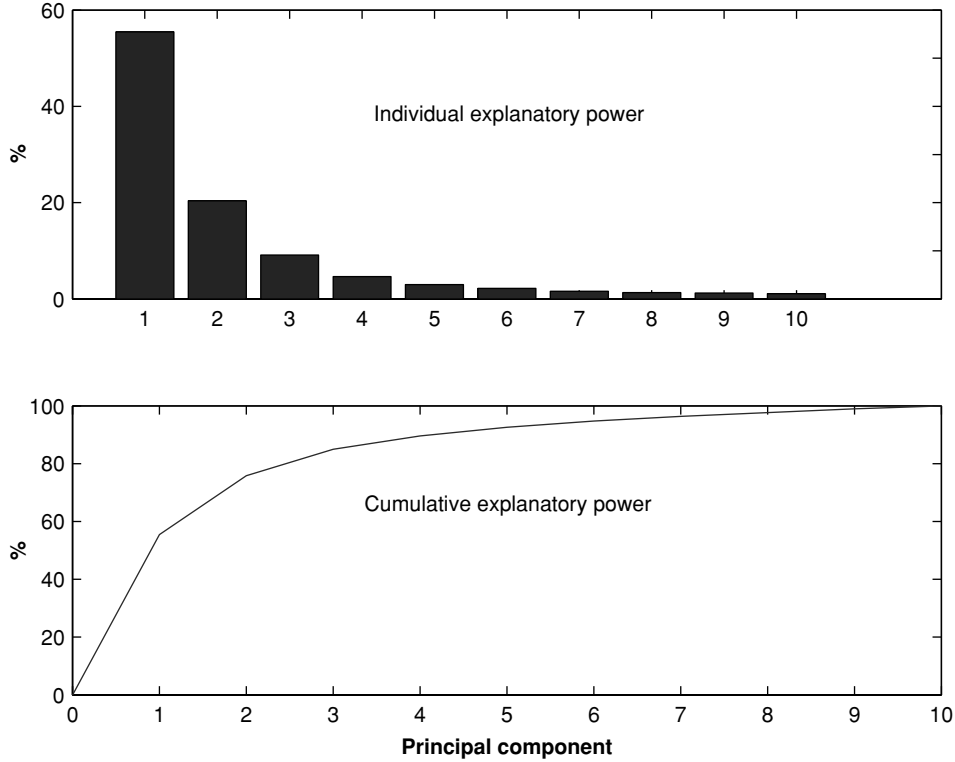

Figure A4.3 Explanatory power of the principal components
Note: These results are obtained using the 'pcaprelim' function.

a relatively small number of principal components is often sufficient to generate accurate risk measures for risk problems of considerable dimensionality. However, we cannot assume that PCA risk estimators will always be so well behaved – empirical return distributions are usually less accommodating than the multivariate normal distribution considered here.

A4.12 FACTOR ANALYSIS

Factor analysis (FA) is a family of methods often used in multivariate data analysis to investigate covariances or correlations among a large number of variables, and to identify clusters of closely related variables. More precisely, we might use FA when we wish to identify a small number of factors that can account for most of the covariance or correlation variation in a multivariate data set. So FA differs from PCA in that it places the emphasis on using a small number of hypothetical random variables to explain the *covariance or correlation* in a multivariate data set, whereas PCA uses a small number of hypothetical variables to explain the *variance* of the same data set.

The idea is that we start with a linear model that relates our $(m \times 1)$ vector of observed variables to a set of unobserved random factors:

$$\mathbf{x} = \mathbf{B}\mathbf{f} + \boldsymbol{\varepsilon} \tag{A4.41}$$

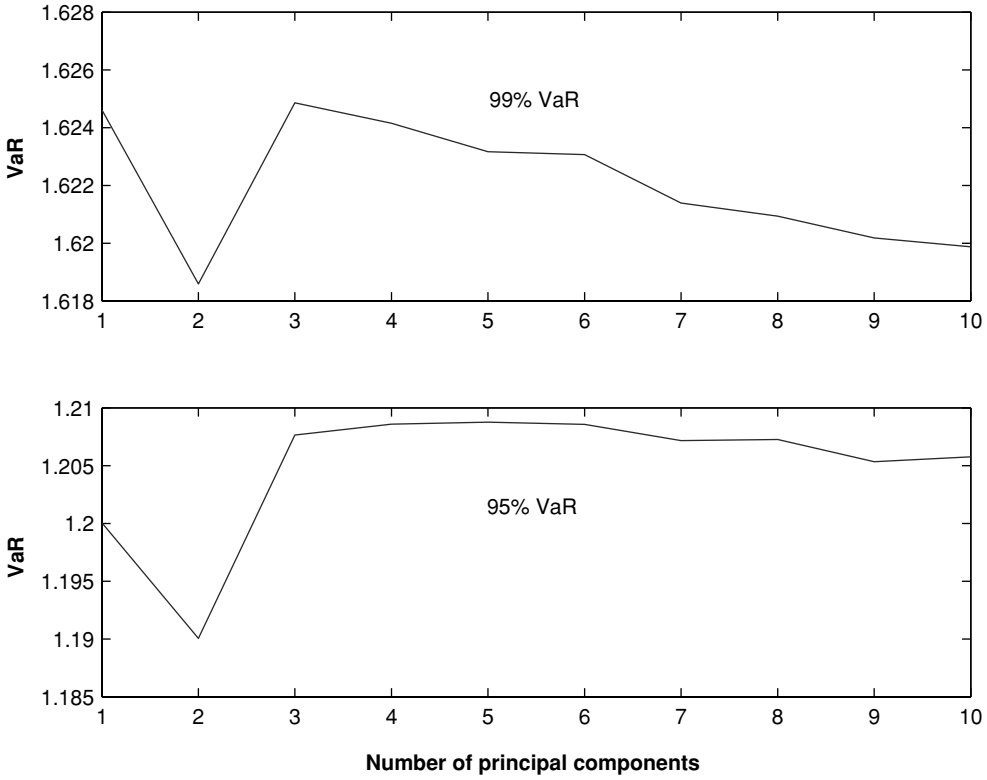

Figure A4.4 Principal components VaR
Note: Estimated using the same data explained in the text using the 'pcavar' function.

where $\mathbf{f}$ is our $q \times 1$ vector of unobserved factors; $\mathbf{B}$ is an $m \times q$ matrix of parameters often known as the 'factor loadings', which needs to be estimated; and $\boldsymbol{\varepsilon}$ is a vector of unobserved random errors. The purpose of the exercise is to determine the $\mathbf{f}$ factors. We now assume that $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are independent, and that the individual errors in $\boldsymbol{\varepsilon}$ are also independent of each other and have expected values of zero. These assumptions ensure that $\boldsymbol{\varepsilon}$ has a diagonal covariance matrix, $\boldsymbol{\Sigma}^{\varepsilon}$, say. If we let $\boldsymbol{\Sigma}$ be the covariance matrix of $\mathbf{x}$ and set the covariance matrix of $\mathbf{f}$ to be the identity matrix $\mathbf{I}$, then we can write $\boldsymbol{\Sigma}$ as:

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^{\mathbf{T}} + \boldsymbol{\Sigma}^{\varepsilon} \tag{A4.42}$$

We now estimate Equation (A4.38) by some appropriate method (e.g., least squares or maximum likelihood).[33] Once we have estimates of $\mathbf{B}$ and $\boldsymbol{\Sigma}^{\varepsilon}$, we plug them into Equation (A4.41) to estimate our factors $\mathbf{f}$ by regression methods.

Once we have the factors, we can then reduce the dimensionality of our data set by working with them instead of our original data.

As we might expect, FA gives rise to deeper issues that may not always be apparent at first sight. As with PCA, there are issues of factor non-uniqueness that require that the individual

[33] For details see, e.g., Adelman (1990), Kennedy and Gentle (1980) or Tsay (2002, Chapter 8.7).

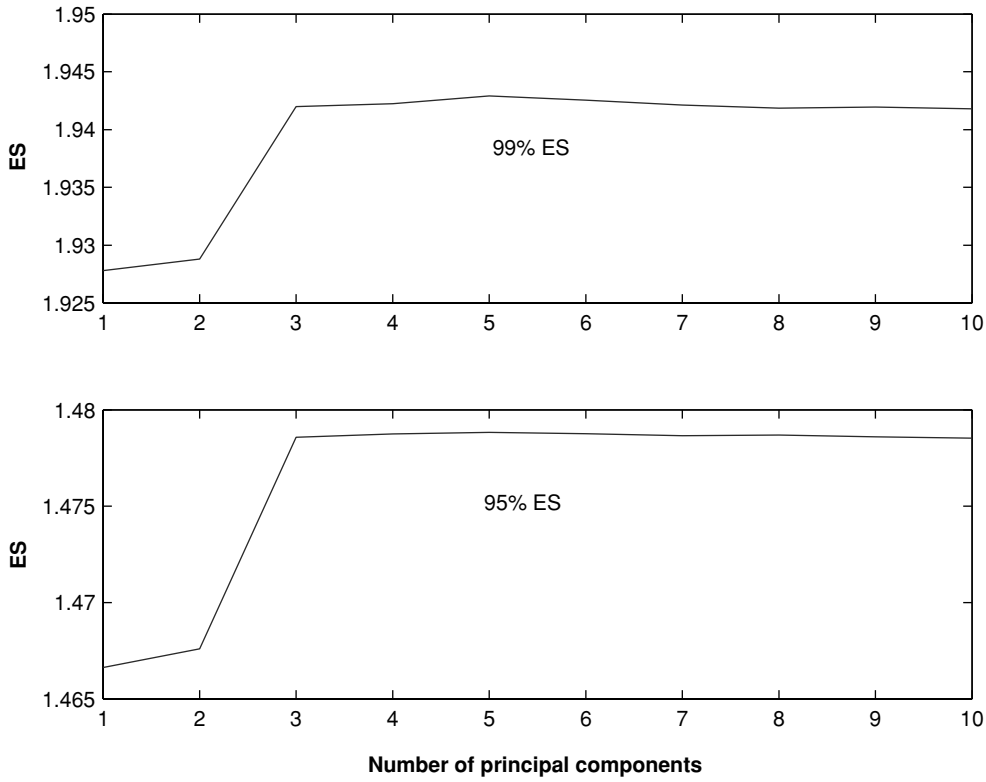

Figure A4.5 Principal components ES
Note: Estimated using the same data explained in the text using the 'pcaes' function.

series in our data set are suitably standardised in scale. Our choice of estimation methods also has implications about the distribution of underlying risk factors. If we use ML, we are making this choice explicitly, and if we use an LS method we are making it implicitly. Either way, we are taking this distribution to be multivariate normal or some close relative of multivariate normality. The reader should also be warned that FA methods are not as robust as PCA methods, and they are more difficult to program as well.

FA differs from PCA in that it focuses on correlations rather than volatilities. It also differs in the number of parameters it needs. While PCA would require only m parameters (i.e., one for each of the eigenvalues of $\mathbf{x}$), FA requires that we estimate $m \times q$ parameters in $\mathbf{B}$ and the m diagonal elements of $\mathbf{\Sigma}^{\varepsilon}$, making a total of $m(q+1)$ parameters to be estimated. Since there are only $m(m+1)/2$ elements in $\mathbf{\Sigma}$ anyway, this implies that we can only achieve a significant reduction in the number of variance–covariance parameters with FA if we work with significantly fewer factors than original variables (i.e., to be precise, if q is significantly less than $(m-1)/2$).

How do we choose between PCA and FA? Assuming that one has the means to implement either method, perhaps the best advice is that practitioners should use the method that best suits the particular problem they are dealing with. The difference between the two methods boils down to how they handle the residual, 'unexplained', elements of the variance–covariance

matrix, as shown by $\boldsymbol{\Sigma}^{\varepsilon}$ in Equation (A4.42). Broadly speaking, PCA tries to pick factors/principal components that maximise the 'explained' variance and so minimise the 'unexplained' variance terms. It operates on the diagonal terms, and leaves the off-diagonal terms as by-products. In contrast, FA tries to reduce the off-diagonal terms in $\boldsymbol{\Sigma}^{\varepsilon}$, and so treats the diagonal terms as by-products. This reasoning suggests that PCA should be relatively good at dealing with volatilities, and FA should be relatively good at dealing with correlations. This in turn suggests that we should use PCA when dealing with problems that are mainly volatility dependent (e.g., when pricing with interest-rate caps and floors) and FA when dealing with problems that are mainly correlation dependent (e.g., when pricing diff swaps and options on yield differentials). Whether PCA or FA is best suited to the estimation of measures of market risk would then appear to depend on whether the principal sources of variability in our data are variance factors or correlation factors, and this presumably varies from one situation to another.

5
Forecasting Volatilities, Covariances and Correlations

This chapter deals with the forecasting of volatilities, covariances and correlations.[1] This is one of the most important subjects in modern risk measurement, and is central to equity and derivatives pricing, hedging and portfolio management. It also provides a very important input to parametric (and sometimes semi-parametric) approaches to the estimation of financial risk measures and, as such, forms an essential preliminary to the two chapters on parametric approaches, which follow.

This chapter is divided into three main sections. Section 5.1 deals with the estimation of volatilities, and covers each of the four main approaches to volatility estimation: the equal-weighted moving average (or historical), exponentially weighted moving average (EWMA), GARCH, and implied volatility approaches. Section 5.2 then deals with the estimation of covariances and correlations, and parallels the earlier treatment of volatilities. Finally, section 5.3 deals with the estimation of variance–covariance and correlation matrices.

5.1 FORECASTING VOLATILITIES

5.1.1 Defining Volatility

We can define volatility as the standard deviation of returns. However, since returns increase with the time period over which they are measured, other things being equal, we need to standardise the period over which volatility is quoted, so we can make meaningful comparisons of volatilities across different horizon periods. The usual way of doing so is to quote volatility on an annualised percentage basis, i.e.:

$$volatility\ at\ time\ t = (100\sigma_t)\% \tag{5.1}$$

where σ_t is the annualised standard deviation of returns. If the volatility is constant, and the random variable follows a random walk, we can derive a straightforward rule to extrapolate the h-period variance from the one-period variance: given these assumptions, the variance over h periods will be equal to h times the variance over one period, and it follows that:

$$\sigma|_{h-period} = \sqrt{h}\sigma|_{1-period} \tag{5.2}$$

Hence, if $\sigma|_{1-period}$ is 1, the variance over two periods is $\sqrt{2}$, the variance over 10 periods is $\sqrt{10}$, and so on. This is the (in)famous 'square root of time' rule, which allows us to extrapolate a variance defined over one period to obtain a variance defined over another.

[1] In theory, we should be careful not to confuse forecasts and estimates. A forecast is the value that we we expect a certain parameter or other quantity to take at the end of a defined future horizon period, so all forecasts are estimates but not all estimates are forecasts. However, in the present context, we can usually use the terms 'forecast' and 'estimate' as if they were interchangeable.

5.1.2 Historical Volatility Forecasts

We turn now to volatility forecasting. If we assume for the sake of argument that the 'true' volatility is constant, and if we ignore scaling factors for convenience, one obvious choice is the historical (or equal-weighted) moving average estimate

$$\sigma_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{t-i} - \bar{x})^2 \tag{5.3}$$

which also provides an unbiased estimate of our volatility. However, if we are dealing with daily data the mean return will be very low, and we can dispense with the need to measure it by treating it as if it were zero. Setting the mean to zero often makes an insignificant difference to our estimates, and (usually) reduces their standard errors.[2] With large samples, it is also common to work with n rather than $n-1$ in the denominator. These modifications lead to

$$\sigma_t^2 = \frac{1}{n}\sum_{i=1}^{n}x_{t-i}^2 \tag{5.4}$$

as our volatility equation. The important point to note about either of these equations – (5.3) or (5.4) – is that they give equal weight to all observations up to the last nth observation, and no weight to any other, more distant, observations.

There are several problems with the historical approach. One problem is that if we stick with the assumption that 'true' volatility is constant, then any differences in our volatility estimates must arise solely from sample error. A short-period (or low n) estimate will then produce a more volatile volatility estimate than a long-period (or high n) estimate, but any such differences can only be ascribed to sampling error, because we have assumed that the true volatility is constant. So if we wish to allow the true volatility to change over time, we need to make less restrictive assumptions about it.

A second problem arises because this model implies that more distant events in the sample period will have the same effect as more recent ones. If an unusual event occurs at date t, this weighting scheme implies that it will continue to influence our volatility estimate for the next n periods, even though the event has passed and markets have returned to normal conditions. The result is a ghost effect – our volatility estimates are artificially high (or low) for the next n periods after the event has taken place, and then suddenly return to normal after the event has dropped out of our sample period. This dropping off at the end of n periods has nothing to do with 'true' volatility, but is entirely a consequence of the way that volatility is estimated.

These effects are illustrated in Figure 5.1 for Equation (5.3) estimated on hypothetical but empirically plausible return data for values of n equal to 20 and 60 supplemented by a big shock occurring at $t = 150$.[3] If n is large, the volatility estimate is more smooth over time and less responsive to individual observations. When our unusual event occurs, both estimates of volatility immediately jump. Volatility then remains high for a while, and eventually dies back down. These high peaks reflect ghost effects that depend on n: the smaller is n, the more rapid the jump and the higher the impact on volatility; however, a smaller n also means that

[2] See, e.g., Figlewski (1994).

[3] To be precise, the return data are simulated data drawn from a conditional normal distribution with a annualised mean of 7% and an annualised GARCH(1,1) volatility process with parameters $\omega = 0.025$, $\alpha = 0.1$ and $\beta = 0.85$. The shock is a negative return equal to four times the previous lowest return in the previous 649 days.

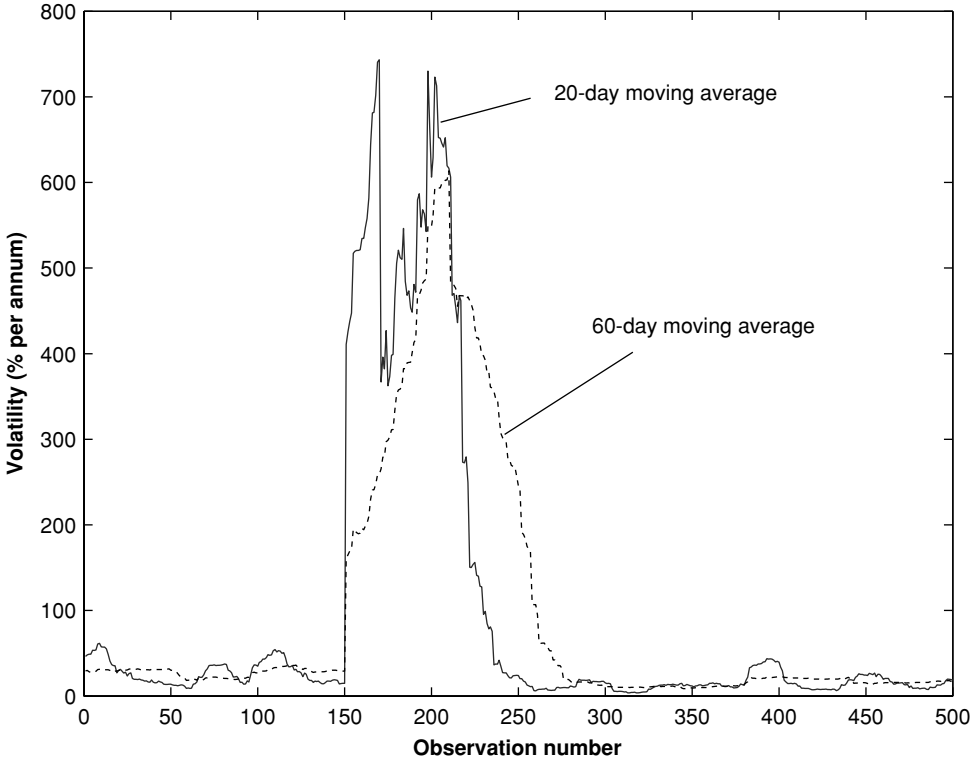

Figure 5.1 Historical volatilities

the shock disappears more quickly. So there is a trade-off between the severity and the length of the ghost effects on volatility.[4]

5.1.3 Exponentially Weighted Moving Average Volatility

One way to ameliorate the drawbacks of equal-weighted moving average schemes is to use a moving average scheme with declining weights, so we can give greater weight to more recent observations and less weight to more distant ones. This type of weighting scheme might (arguably) be justified by claiming that volatility tends to change over time in a stable way, which is certainly more reasonable than assuming it to be constant. Our volatility-forecasting model then has the form:

$$\sigma_t^2 = \sum_{i=1}^{n} \alpha_i x_{t-i}^2 \tag{5.5}$$

where the weights, the α_i terms, decline as i gets larger, and sum to 1. One of the simplest examples is the exponentially weighted moving average (EWMA) model, in which the weights

[4] However, it is also interesting to observe from the figure that the volatility of the volatility also increases after the shock. This reflects the fact that the true volatility process is assumed to be a GARCH – which is not unreasonable – and the shock increases the GARCH volatility. This increased volatility also makes returns more volatile, and creates a positive feedback loop between the volatility of returns and the volatility of volatility.

decline exponentially over time: this means that $\alpha_{i+1}/\alpha_i = \lambda$, where λ is a constant between 0 and 1. These assumptions lead to the following volatility forecasting equation:

$$\sigma_t^2 \approx (1-\lambda)\sum_{i=1}^{n}\lambda^{i-1}x_{t-i}^2 \tag{5.6}$$

The approximation is valid provided that n is sufficiently large. The EWMA scheme has the intuitively appealing property that the influence of any observation declines over time at a stable rate, and it is easy to apply because it relies on only one parameter, λ.

The EWMA also leads to a very straightforward volatility formula. If we lag Equation (5.6) by one period, and multiply throughout by λ, we get:

$$\lambda\sigma_{t-1}^2 \approx \lambda(1-\lambda)\sum_{i=1}^{n}\lambda^{i-1}x_{t-i-1}^2 = (1-\lambda)\sum_{i=1}^{n}\lambda^{i}x_{t-i-1}^2 \tag{5.7}$$

We now subtract Equation (5.7) from Equation (5.6) and rearrange to get:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)x_{t-1}^2 - (1-\lambda)\lambda^n x_{t-n-1}^2 \approx \lambda\sigma_{t-1}^2 + (1-\lambda)x_{t-1}^2 \tag{5.8}$$

This formula tells us that the estimate of volatility for day t, σ_t, made at the end of day $t-1$, is calculated from the previous day's volatility estimate, σ_{t-1}, and the previous day's return, x_{t-1}. The EWMA rule (Equation (5.8)) can therefore be interpreted as a simple updating rule that allows us to update our daily volatility estimate each day based on the most recent daily return. A high λ means that the weight declines slowly, and a low λ means it declines quickly. Its value would need to be determined from the data at hand, but the *RiskMetrics – Technical Document* suggests that we can often get away with taking λ to be about 0.94 for daily return data.[5]

Some EWMA volatility estimates are plotted in Figure 5.2 using the same data set as the last figure, with values of λ equal to 0.90 and 0.95. When the shock occurs at $t = 150$, both EWMA volatility estimates spike upwards, and then fall back down as time passes. The low-λ volatility rises the most, but declines more rapidly afterwards; the high-λ volatility rises less sharply, but also falls at a slower rate in later periods. It is clear from this figure that the EWMA tends to produce less prominent ghost effects than equal-weighted moving average schemes.

We can also use Equation (5.8) to make forecasts of future volatility. We begin by leading Equation (5.8) by one period:

$$\sigma_{t+1}^2 \approx \lambda\sigma_t^2 + (1-\lambda)x_t^2 \tag{5.9}$$

Taking expectations as of t, and noting that $E(x_t^2) = \sigma_t^2$, we get:

$$E(\sigma_{t+1}^2) \approx \lambda\sigma_t^2 + (1-\lambda)\sigma_t^2 = \sigma_t^2 \tag{5.10}$$

so the one-period ahead forecast of our volatility is approximately equal to our current volatility estimate, σ_t^2. It is easy to show, by similar reasoning, that our k-period ahead volatility forecast is the same:

$$E\left(\sigma_{t+k}^2\right) = \sigma_t^2; \quad k = 1, 2, 3, \ldots \tag{5.11}$$

[5] Morgan Guaranty Trust Company (1996), p. 97.

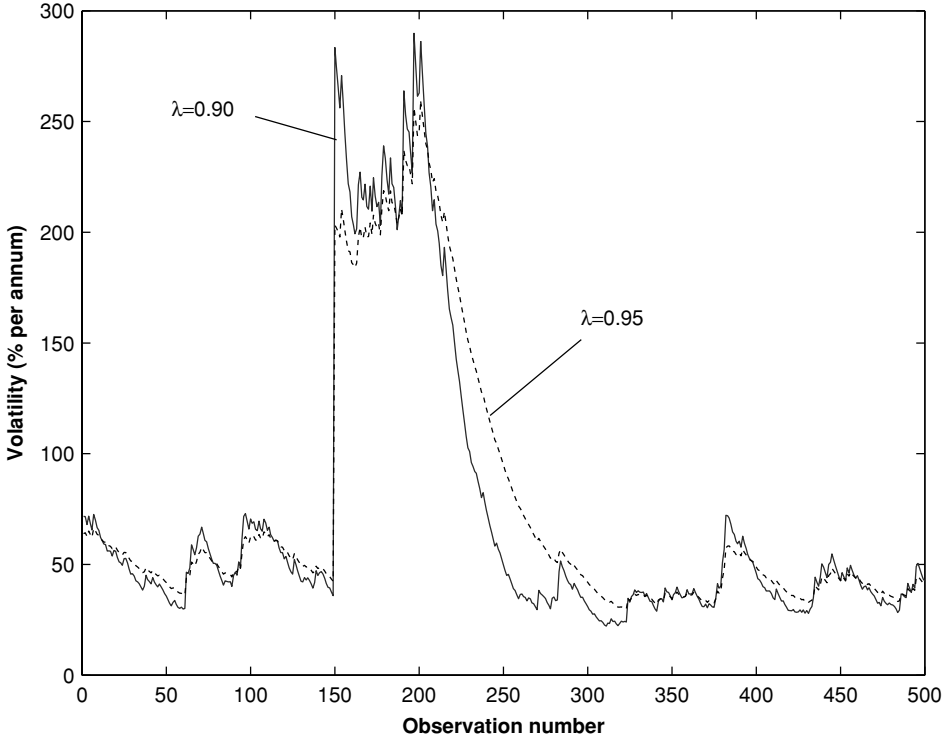

Figure 5.2 EWMA volatilities

The EWMA model therefore implies that our current volatility estimate also gives us our best forecast of volatility for any period in the future.

However, this 'flat' volatility forecast is not appealing, because it ignores any recent dynamics in our data: for example, even if volatility has been rising strongly in the recent past, the EWMA predicts – not too plausibly – that future volatility will immediately level off and remain at its current level.

5.1.4 GARCH Models

EWMA models are also implausible in that they take the λ parameter to be constant, and therefore unresponsive to market conditions, and a solution to these problems is provided by GARCH (generalised autoregressive conditional heteroscedasticity) models, which are a popular, and in some respects superior, set of alternatives to the EWMA model.[6] Two of the most important stylised facts with return data are that they show volatility clustering (i.e., they go through alternating periods of high and low volatility) and leptokurtosis (i.e., fatter than normal tails). GARCH models can accommodate both these stylised facts very easily. Indeed, they are tailor-made for volatility clustering, and this clustering produces returns with fatter than normal tails

[6] The basic ARCH model was first suggested by Engle (1982), and the GARCH generalisation was suggested by Bollerslev (1986). These subsequently led to a large family of GARCH-type models, including the AGARCH, EGARCH, IGARCH, MARCH, NARCH, QTARCH, SPARCH, STARCH, SWARCH and TARCH models, to name only the most pronounceable. Some of the more relevant ones for our purposes are discussed below.

even if the innovations – the random shocks – are themselves normally distributed. The basic GARCH(p,q) model postulates that volatility depends on q past volatilities and p past returns:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \omega > 0,$$
$$\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0 \tag{5.12}$$

where the constraints on parameter values are necessary to ensure that the conditional variance is always positive. GARCH models thus postulate that current volatility depends on one or more past volatilities and past returns, but in a more general way than EWMA models do. GARCH models vary in the number of past terms used, and we should choose these terms according to the principle of parsimony (i.e., choose the minimum that fit the data acceptably). GARCH models also differ in the distribution governing the error term ε_t. This distribution will typically be conditionally normal, and conditional normality will produce leptokurtosis – greater than normal kurtosis – in our returns, consistent with the stylised fact that observed returns usually show excess kurtosis. We can also further fatten our tails if we replace the assumption of conditional normality by a conditional t- or conditional mixture-of-normals distribution.[7]

5.1.4.1 The GARCH(1,1) model

The most popular GARCH model is the GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2; \quad \omega \geq 0, \alpha, \beta \geq 0, \alpha + \beta < 1 \tag{5.13}$$

This model is easy to apply, uses a small number of parameters, and often fits the data fairly well. A high value of β means that volatility is 'persistent' and takes a long time to change; and a high value of α means volatility is 'spiky' and quick to react to market movements. It is also common to get estimates of β of over 0.7, but α is usually less than 0.25.

To give an indication of what a typical GARCH volatility looks like, Figure 5.3 gives a plot of a GARCH(1,1) process against our earlier hypothetical data set, estimated with $\omega = 0.025$, $\alpha = 0.1$ and $\beta = 0.85$. In this particular case, the GARCH volatility estimate is quite responsive to the data, and the shock at $t = 150$ dissipates from the volatility estimate a little more quickly than was the case for the EWMA volatility estimate.

The GARCH(1,1) volatility depends on the same variables as the EWMA, but there are now three parameters instead of one, and we can regard the EWMA as a special case of the GARCH(1,1) process that occurs when $\omega = 0$, $\alpha = 1 - \lambda$ and $\beta = \lambda$.

The GARCH(1,1) with positive intercept ω also has the attraction that it allows us to model the volatility as mean-reverting – so that if the volatility is relatively high, it will tend to fall over time; and if volatility is relatively low, it will tend to rise over time. The long-run variance – the value to which the variance will tend to revert in the long run – is $\omega/(1 - \alpha - \beta)$.

GARCH models also give straightforward volatility forecasts. If we denote the long-run variance by $V = \omega/(1 - \alpha - \beta)$, then Equation (5.13) implies

$$\sigma_t^2 = (1 - \alpha - \beta)V + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2$$
$$\Rightarrow \sigma_t^2 - V = \alpha \left(x_{t-1}^2 - V\right) + \beta \left(\sigma_{t-1}^2 - V\right) \tag{5.14}$$

[7] There are many pre-programmed procedures for carrying out GARCH estimation available in standard econometric and statistical packages. The idea is to choose parameter estimates to maximise the likelihood under an assumed error density function. In applying GARCH models, we also have to choose the number of parameters, and there are a variety of standard tests to do so (e.g., the Box–Pierce and Ljung–Box tests). For more on these issues, see, e.g., Alexander (1998) or Giannopoulos (2000).

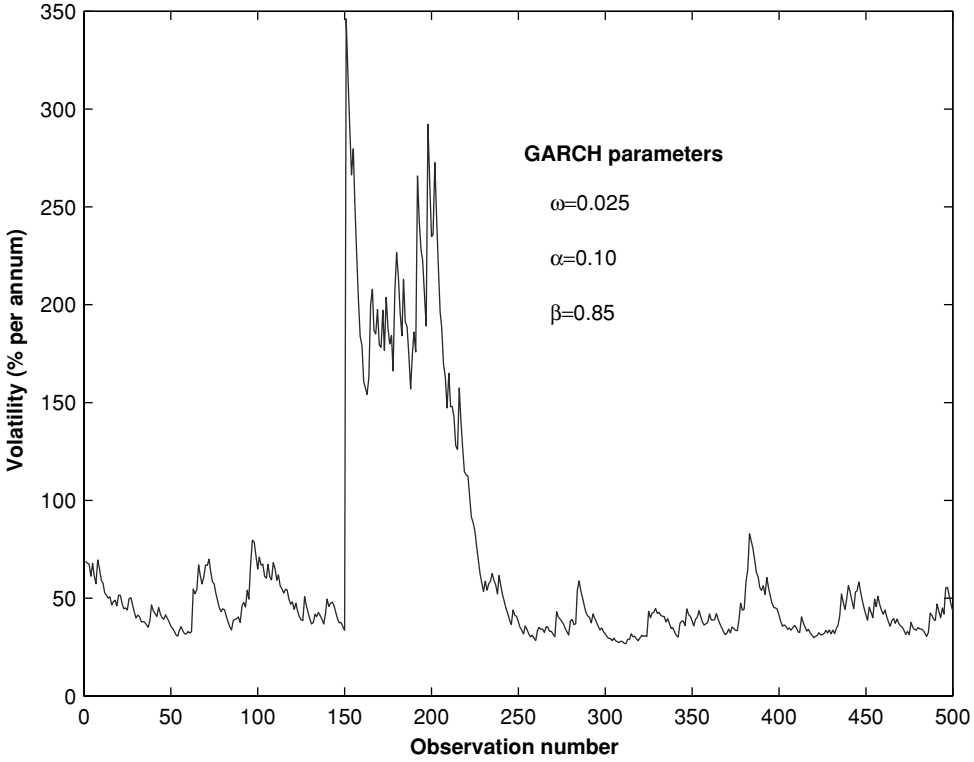

Figure 5.3 Plot of GARCH(1,1) volatility

We now lead Equation (5.14) by k periods:

$$\sigma_{t+k}^2 - V = \alpha\left(x_{t+k-1}^2 - V\right) + \beta\left(\sigma_{t+k-1}^2 - V\right) \tag{5.15}$$

Given that the expected value of $x_{t+k-1}^2 = \sigma_{n+k-1}^2$, it follows that

$$E\left[\sigma_{t+k}^2 - V\right] = (\alpha + \beta)\left(x_{t+k-1}^2 - V\right) \tag{5.16}$$

and so our k-period-ahead forecast is:

$$E\left(\sigma_{t+k}^2\right) = V + (\alpha + \beta)^k\left(\sigma_t^2 - V\right) \tag{5.17}$$

Since $\alpha + \beta < 1$, the second term in Equation (5.17) falls as k gets larger, so the forecasted variance converges to V as we look further into the future, which justifies the earlier claim that V can be interpreted as the long-run variance. If $\sigma_t^2 > V$, the expected k-period ahead variance is larger than V, and if $\sigma_t^2 < V$, the expected k-period ahead variance is smaller than V. The GARCH forecasts therefore tend to revert back to the long-run variance V.

These forecasts can then be used to derive estimates of the volatility term structure. Assuming that we are working with geometric returns, we know that the return at time t over the next n periods is:

$$r_{t,n} = \sum_{j=1}^{n} r_{t+j} \tag{5.18}$$

This implies that

$$\text{var}_t(r_{t,n}) = \sum_{i=1}^{n} \text{var}_t(r_{t+i}) + \sum_{i=1}^{n}\sum_{j=1}^{n} \text{corr}_t(r_{t+i}, r_{t+j}) \tag{5.19}$$

We can therefore derive estimates of the volatility term structure from forecasts of future one-period volatilities and covariances. However, in many cases the correlation terms in Equation (5.19) will be small relative to the volatility terms, so we can often ignore them and treat the volatility of the n-period return as the sum of n one-period volatilities:

$$\text{var}_t(r_{t,n}) \approx \sum_{i=1}^{n} \text{var}_t\,(r_{t+i}) \tag{5.20}$$

The basic GARCH model therefore produces a volatility term structure that eventually converges on nV, because each of the $\text{var}_t\,(r_{t+i})$ terms on the right-hand side of Equation (5.20) eventually converges on V. This is reasonable enough for some purposes, but can sometimes be rather restrictive.

Box 5.1 Estimating GARCH Models

The estimation of GARCH models involves a number of steps. The first, preliminary, step is to run our return through a simple filter (such as an autoregressive moving average (ARMA) model) to remove any serial correlation in the data.[8] The ARMA model should be as parsimonious as possible, and should produce a serially uncorrelated residual series with a mean close to zero. We then square the residuals and test the squared residuals for conditional heteroscedasticity (or changing variance), and there are a number of standard tests (Box–Pierce tests, Ljung–Box tests, etc.) that can be used for the purpose. A significant value for the test statistic indicates the presence of conditional heteroscedasticity and suggests that a GARCH model might be appropriate. We then look at partial autocorrelation coefficients or other evidence to select a particular GARCH specification. This includes the choice of a particular noise (or innovation) process, which would typically be a normal or t process. We then estimate the parameters of the model using an appropriate maximum likelihood method. The last stage is to check the adequacy of our GARCH model by testing that the standardised innovations (i.e., the innovations divided by the contemporary volatility estimates) are iid and follow the assumed noise distribution. We can check the iid prediction by standard tests such as Ljung–Box, and we can test the distribution prediction by checking that the moments of the standardised noise process are compatible with the assumed process, and by using suitable diagnostics such as QQ plots.[9] The GARCH model is then considered adequate if its standardised residuals pass these tests.

[8] In estimating GARCH models, it is often a good idea first to check the order of integration of the series (i.e., we should test for a unit root). If there is a unit root, then we should work with first differences of our data or, equivalently, work with autoregressive integrated moving average (ARIMA) and integrated GARCH models. For more on these issues, see any good econometrics textbook. The procedure set out in Box 5.1 thus presupposes that our data do not have a unit root (i.e., are stationary).

[9] The estimation of GARCH models is covered in all the good econometrics (and especially financial econometrics) textbooks, and there are many good software packages available with which GARCH models can be estimated (e.g., EViews, RATS, the MATLAB Garch Toolbox, etc.).

5.1.4.2 *Integrated GARCH*

Another popular GARCH model is the integrated GARCH or IGARCH model. This is applicable when our return series is not stationary – as is commonly the case – and so the long-term variance V does not exist. In the three-parameter case, we can think of $\alpha + \beta$ in Equation (5.14) becoming 1, and the GARCH(1,1) model becoming:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + (1-\beta)x_{t-1}^2 \tag{5.21}$$

This model is often used in currency markets and includes the EWMA as a special case when $\omega = 0$.

5.1.4.3 *Components GARCH*

The standard GARCH model has the volatility converge to a long-term or baseline level that depends on the GARCH parameters, but is constant over time. This is somewhat restrictive, but can be relaxed using the components GARCH model. For example, if we are using a GARCH(1,1) model, we would replace the constant V by a baseline volatility equation:

$$V_t = \bar{\omega} + \rho(V_{t-1} - \bar{\omega}) + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right) \tag{5.22}$$

which allows the baseline volatility to vary in response to market conditions. Equations (5.13) and (5.22) together define the components GARCH(1,1) model.

5.1.4.4 *Factor GARCH*

Another handy member of the GARCH family is the factor GARCH model.[10] This model allows a number of volatilities (and correlations) to be estimated from a single volatility estimate. In the standard case, we wish to estimate a number of volatilities from a single market volatility index M_t. For example, we might have n different assets, whose returns are linked to a market return by a CAPM-type equation:

$$r_{it} = \alpha_i + \beta_i M_t + \varepsilon_{it}; \quad i = 1, \ldots, n \tag{5.23}$$

The variance of asset r_{it}, σ_{it}, is then given by

$$\sigma_{it}^2 = \beta_i^2\sigma_{Mt}^2 + \sigma_{\varepsilon_{it}}^2 \tag{5.24}$$

To apply the factor GARCH model, we estimate the set of Equations (5.23) and recover estimates of the β_i and $\sigma_{\varepsilon_{it}}^2$ terms. We then apply a standard univariate GARCH procedure to estimate the market volatility, σ_{Mt}^2, and input our parameter estimates into Equation (5.24) to generate volatility estimates for our individual assets.[11]

[10] See Engle and Mezrich (1996).

[11] Market reactions often exhibit asymmetries, and these can be modelled using asymmetric GARCH models. The most prominent of these are the asymmetric GARCH (or AGARCH), exponential GARCH (or EGARCH) and threshold GARCH models. For more on these, see, respectively, Engle and Ng (1993), Nelson (1990) and Glosten *et al.* (1993).

5.1.5 Implied Volatilities

A very different approach to the estimation of volatilities is to use implied volatilities from options prices. The idea is that where options exist on underlying assets and we know the values of the other variables involved, we can use established option-pricing models to 'back out' the volatilities consistent with the known prices of these options. For example, suppose we have data on the prices of standard European Black–Scholes call options. Assuming the various Black–Scholes conditions hold (i.e., the underlying process follows a geometric Brownian motion, the underlying pays no dividend, etc.), the basic Black–Scholes theory tells us that the price of this option, c, should be:

$$c = SN(d_1) - Xe^{-rt}N(d_1 - \sigma\sqrt{t}) \tag{5.25}$$

where

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)t}{\sigma\sqrt{t}}$$

and S is the current stock price, X is the strike price, r is the risk-free rate of interest, t is the option's remaining term to maturity in years, σ is the volatility of the underlying asset, and $N(.)$ is the value of the cumulative standard normal distribution. We know all of these variables except the volatility, so if the model is correct, we ought to be able to use Equation (5.25) to derive the volatility it implies. These implied volatilities rarely have 'closed-form' solutions, but are very easy to derive numerically on a spreadsheet (e.g., using bisection or Newton–Raphson methods). For example, if we are given that $S = X = 1$, $r = 0.05$, $t = 1/2$, and $c = 0.0826$, then sigma must be 0.25, and this value can be verified by inserting it into Equation (5.25) and checking that that produces the observed market price.

It is important to appreciate that implied volatility is not some backward-looking econometric volatility estimate, but a forward-looking forecast of volatility over the option's term to maturity. This gives implied volatilities two big advantages over other, historically based, estimates of volatility: implied volatilities incorporate information that other approaches will ignore unless it is present in the historical data (e.g., expectations of the imminent devaluation of a hitherto stable currency), and they provide estimates of volatility on which experienced market operators have enough confidence to bet real money. There is therefore a general belief that implied volatility forecasts are better than historically based ones.

However, the implied volatility method of estimating volatilities is contingent on the accuracy of the model used: if the model is biased or otherwise flawed, then it may produce biased or flawed estimates of implied volatility. This is a major problem, because standard option-pricing models have well known limitations. These include the failure of most of them to allow for transactions costs, bid–ask spreads, the effects of market illiquidity, and other 'imperfections'. They also include more fundamental problems with some key assumptions on which the models are based, including the famous 'holes in Black–Scholes' – the assumptions that the underlying follows a geometric Brownian motion, that the risk-free interest rate is constant, and so on. The failure of these assumptions to hold produces phenomena such as volatility smiles and skews that seriously complicate the derivation of implied volatilities. For example, the volatility smile means that we get a number of different implied volatility estimates from options with different strike prices, and this leaves us with the problem of working out which of the implied volatilities on offer best corresponds to the one we are looking for. However, these sorts of problems are well known, and options practitioners have developed sophisticated ways of dealing with them.

Implied volatilities also suffer from another limitation: they only exist for assets on which options have been written. This means that they are available only for a small subset of the assets for which we might seek volatility estimates. However, as time passes and options are written on more assets, we should expect more implied volatility estimates to become available. Nonetheless, these methods only go so far, and implied volatility estimates are always likely to be in sort supply.

Box 5.2 Realised Volatility Estimators

One trend in the volatility literature is towards the use of ever higher-frequency data, and a promising volatility estimator to emerge from this work is the so-called realised volatility estimator. The realised volatility is the average of intra-period high-frequency squared returns (e.g., the period might be a day, say, and the intra-period frequency every five minutes), and work by Andersen and Bollerslev (1997, 1998b), Andersen *et al.* (2001), and Corsi *et al.* (2001) shows that the realised volatility estimator is more or less error-free. However, intra-day volatilities show pronounced biases resulting from market micro-structure effects, and care has to be taken to remove these biases when deriving daily volatilities from intra-day volatilities.

Andersen *et al.* point out that realised variances tend to be lognormally distributed, and that asset returns standardised by realised standard deviations tend to be normally distributed. Since realised returns are effectively observable, they can be handled directly by standard forecasting methods, and Andersen *et al.* suggest that we assume that the log-volatility process (which is normal) be estimated using Gaussian autoregressive moving average (ARMA) models. However, since realised volatility also has a long memory, we can allow for this by using a fractional order of integration in the ARMA process. This requires us to estimate the fractional order of integration – which is around 0.04 – and then apply an ARMA model to the fractionally integrated series, $y_t = (1 - L)^{0.4} \log \sigma_t$. As these authors observe, 'The striking feature of this approach is that it builds directly on observed time series and utilises only standard linear Gaussian modelling and forecasting techniques. Hence, it is simple to assess in-sample performance and evaluate model forecasts through well-established out-of-sample procedures.'

Moosa and Bollen (2001) use this realised volatility estimator to assess the bias in standard volatility estimators, and they find that this bias depends on both the methodology used and the length of the sample period. They also find that the best overall estimator – which is also unbiased – is an exponentially weighted moving average with a large decay factor. This suggests that the EWMA estimator might be best in practice, after all, despite the theoretical superiority of GARCH.

5.2 FORECASTING COVARIANCES AND CORRELATIONS

5.2.1 Defining Covariances and Correlations

We turn now to covariance and correlation forecasting. The covariance between two series x and y is given by:

$$\text{cov}(x, y) = E[xy] - E[x]E[y] \tag{5.26}$$

and is approximately $E[xy]$ if x and y have negligible means. The correlation between x and y is then the covariance standardised by the square root of the product of the variances of x and y:

$$\text{corr}(x,y) = \frac{\text{cov}(x,y)}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{\text{cov}(x,y)}{\sigma_x\sigma_y} \tag{5.27}$$

Equations (5.26) and (5.27) mean that we can obtain correlation estimates from covariance estimates, and vice versa.

Provided the series concerned satisfy suitable (i.e., elliptical) distributions, the correlation coefficient gives a useful indication of the extent to which they move together. In such circumstances, the correlation will lie between -1 and $+1$, and take the value -1 if they are perfectly negatively correlated, and $+1$ if they are perfectly positively correlated.

We should also keep in mind that the covariances and correlations of returns, like the volatilities of returns, are predicated on the period of time over which the return is measured, so each covariance or correlation refers to the returns of two assets over a specified period. For any set of assets, there is therefore a family of covariances and correlations, each predicated on returns measured over different periods.

However, when using correlations we must take care to ensure that our two variables, x and y, satisfy appropriate distributions (see the appendix to this chapter). For correlation even to exist, it is necessary that x and y have finite variances and that they be jointly stationary.

Unfortunately, many empirical returns are not jointly stationary. In such cases, correlations often do not exist, and attempts to estimate them would be fundamentally misconceived. This leads to a clear practical danger: we might be dealing with a case where correlation is not defined, and yet we might not realise that the correlation does not exist and proceed to estimate the correlation coefficient anyway. Our results would then be meaningless and we would not know it. We must therefore take care to ensure that correlations actually exist before we attempt to estimate them, and where we do estimate them, we should be on the look out for instability in our estimates: unstable correlation estimates are a classic sign of non-joint stationarity in our returns.

5.2.2 Historical Covariances and Correlations

The estimation of covariances and correlations directly parallels the estimation of volatilities. The most basic estimation method is therefore a historical (or equal-weighted) moving average, which yields the following correlation estimator:

$$\text{corr}(x,y)_t = \frac{\sum_{i=1}^{n} x_{t-i}\,y_{t-i}}{\sqrt{\sum_{i=1}^{n} x_{t-i}^2 \sum_{i=1}^{n} y_{t-i}^2}} \tag{5.28}$$

which is a direct analogue of the earlier volatility estimator (Equation (5.4)). Traditionally, practitioners have sought to choose n so that it is large enough to produce reasonable correlation estimates, but small enough to be responsive to market 'news'. However, we need to keep in

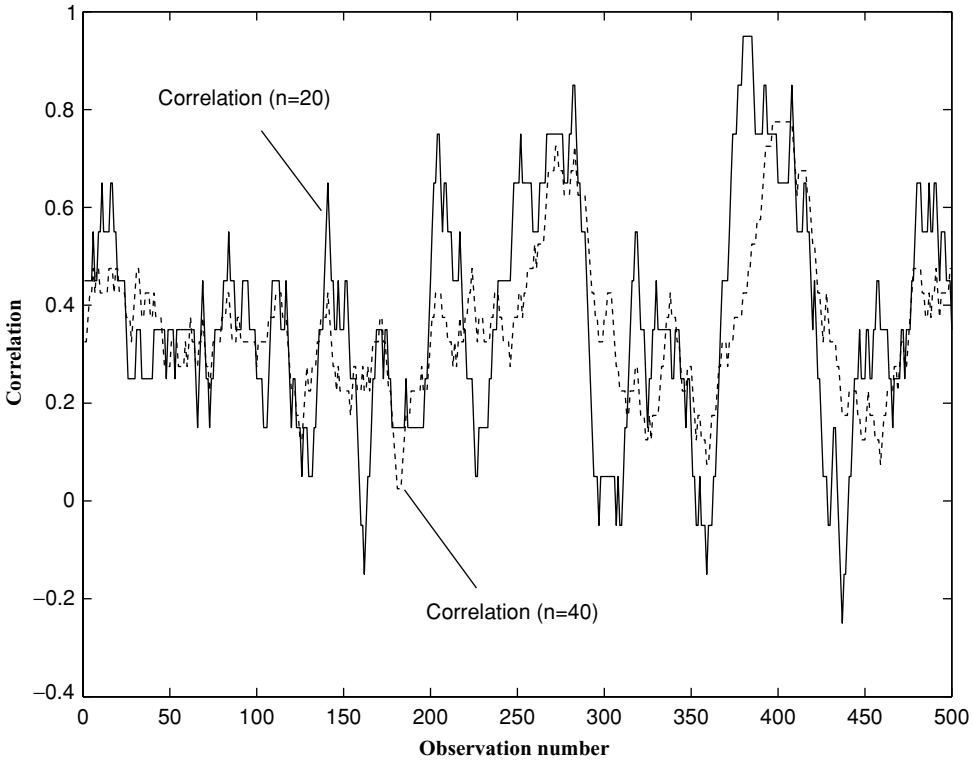

Figure 5.4 Historical correlation estimates

mind that estimated correlation coefficients will tend to become more stable as our sample period rises, regardless of whether our returns are jointly stationary or not. Consequently, we must be wary of the possibility that correlations might appear to be stable only because we use long sample periods to estimate them.

Some typical historical moving average correlations are shown in Figure 5.4. These are derived with rolling sample sizes of 20 and 40 respectively, and are estimated on simulated multivariate normal return data with a 'true' correlation of 0.5. The fact that both correlations move around fairly wildly, despite the fact that the data are drawn from the most well-behaved distribution imaginable, is ample illustration of how unreliable these correlation estimates can be. As we would expect, the most volatile estimate is the one based on the shorter sample period, but the even more stable estimate is itself fairly volatile: it varies between 0.025 at one extreme and 0.775 at the other, has a mean of 0.359 (which compares to the 'true' mean of 0.5), and has a standard deviation of 0.153. These correlations are very volatile indeed, and give relatively little indication of the 'true' correlation.

It is also interesting to note that correlation estimates are often immune to one-off shocks such as the ones considered earlier: such shocks would have no discernible impact on any of the backward-looking correlation forecasts considered here. On the other hand, market experience suggests that markets do tend to react to big shocks by radicalising correlations. In fact, radicalised correlations are often a key feature of market crises.

However, even where ghost effects might not be important for correlations, they *are* likely to be important for covariances. As Equation (5.27) implies, a covariance can be considered as a correlation multiplied by the product of the standard deviations (or volatilities) of x and y:

$$\text{cov}(x,y) = \sigma_x \sigma_y \text{corr}(x,y) \tag{5.29}$$

and we already know that volatilities can be subject to ghost effects. We should therefore expect our covariance estimates to exhibit ghost effects that reflect those in our volatility estimates: in effect, the ghost effects of our volatility estimates carry over to our covariance estimates.

5.2.3 Exponentially Weighted Moving Average Covariances

We can also estimate covariances using an EWMA rule, and the EWMA covariance is:

$$\text{cov}(x,y)_t = \lambda \text{cov}(x,y)_{t-1} + (1-\lambda) x_{t-1} y_{t-1} \tag{5.30}$$

This behaves in exactly the same way as the EWMA volatility model: the lower the value of λ, the more rapidly the weights on observations decline as time passes. EWMA covariances are similar to the EWMA volatilities in Figure 5.3.

EWMA correlations can then be found by applying Equation (5.27) to the EWMA covariance (Equation (5.30)). EWMA correlation estimates are a little more spiky than the equal-weighted averages in the Figure 5.4, but are otherwise much the same, and certainly not appreciably more accurate.

5.2.4 GARCH Covariances

Besides exhibiting ghost effects, EWMA covariance estimates are also fairly restrictive. The natural solution to these problems is to estimate covariances with GARCH models. GARCH covariance models are directly analogous to GARCH volatility models. Thus, the GARCH(1,1) covariance model is:

$$\text{cov}(x,y)_t = \omega_{xy} + \alpha_{xy} x_{t-1} y_{t-1} + \beta_{xy} \text{cov}(x,y)_{t-1} \tag{5.31}$$

which corresponds to the GARCH(1,1) volatility model (Equation (5.14)):

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{5.14}$$

We can also estimate covariances (or correlations) with any of the other GARCH-family models – IGARCH, components GARCH, factor GARCH, etc. – in an analogous way. For example, if we use the factor GARCH approach, the covariance of asset r_{it} with asset r_{jt}, $\text{cov}(i, j)_t$, is:

$$\text{cov}(i,j)_t = \beta_i \beta_j \sigma_{Mt}^2 + \sigma_{\varepsilon_{it}} \sigma_{\varepsilon_{jt}} \tag{5.32}$$

We then apply the factor GARCH model to covariance estimation in exactly the same way as we would apply it to volatility estimation (i.e., we estimate the parameters by regression, apply GARCH to the market return, and use Equation (5.32) to estimate the covariance).[12]

[12] As with other correlation estimates, GARCH correlation estimates can be found by applying Equation (5.27) to the relevant covariance forecast.

5.2.5 Implied Covariances and Correlations

We can also estimate covariances and correlations by deriving implied covariances and correlations analogous to the implied volatilities discussed earlier. To understand how, first note that we can write out the variance of the difference between x and y as:

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \tag{5.33}$$

where ρ is their correlation coefficient. We then rearrange Equation (5.33) to put the correlation on the left-hand side:

$$\rho = \frac{\sigma_x^2 + \sigma_y^2 - 2\sigma_{x-y}^2}{2\sigma_x\sigma_y} \tag{5.34}$$

This tells us how we can derive a correlation estimate from estimates of the three volatilities σ_x^2, σ_y^2, and σ_{x-y}^2. However, we can only derive these implied correlations if the relevant options exist from which we can obtain the necessary implied volatilities. In this particular case, this means that we need options on x and on y, and also, more problematically, an option on the difference between x and y (e.g., a spread, quanto or diff option). These latter options are obviously fairly rare, so we will rarely have the opportunity to work with implied correlations. But even where we can derive them, we need to treat them carefully, as they have all the limitations and more of implied volatility estimates, and can also be very unstable.

5.2.6 Some Pitfalls with Correlation Estimation

Finally, market experience suggests that when estimating correlations, practitioners should keep in mind some important reservations:

- Correlation estimates are often very volatile, so practitioners should interpret (and rely on) estimated correlations with great care.
- Precisely because of this volatility, it often takes a considerable number of observations to detect any substantial changes in correlations, and this means that changes in correlations can often be identified only after it is too late to do anything about them. Practitioners should be careful to protect themselves against the possibility that they might have estimated correlations incorrectly.
- Correlations can often appear to be fairly stable in 'normal' market conditions, and then jump to very high or very low values in stressful situations: in other words, correlations can break down just at the moment they are needed most. Correlation-based risk estimates should take this possibility into account, and should certainly not be reliant on assumptions that markets will behave 'normally' in a crisis.

There is also a vital methodological point: as mentioned already, the correlation is only defined (i.e., only exists) under limited conditions, so we should check to satisfy ourselves that these conditions hold before proceeding to estimate any correlations (or volatilities). If we fail to do this, there is a real danger that we will produce estimates that are meaningless, because the parameters being estimated do not exist. Just because we (think we) can estimate something does not mean that it exists!

5.3 FORECASTING COVARIANCE MATRICES

5.3.1 Positive Definiteness and Positive Semi-definiteness

We now turn to the forecasting of covariance (and correlation matrices). This is generally more difficult than the estimation of individual volatilities and covariances or correlations. This is because we want a covariance matrix that will give us a portfolio variance that is always positive or (if we allow for a zero variance) always non-negative, regardless of the relative sizes of the individual positions. In the former case, we require that the covariance matrix $\boldsymbol{\Sigma}$ be positive definite, which means that $\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathbf{T}} > 0$ for any $1 \times m$ position size vector $\mathbf{w} \neq 0$ and corresponding $m \times 1$ transpose vector $\mathbf{w}^{\mathbf{T}}$. In the latter case, we require the slightly weaker condition that $\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathbf{T}} \geq 0$. Whichever of these conditions we choose – positive definiteness is more restrictive but sometimes more convenient to handle – we also require that the estimated covariance matrix (or correlation matrix) satisfies this condition as well. We therefore need systemic approaches that estimate all parameters in the matrix, but in a way that ensures that our estimated variance–covariance matrix will be positive definite (or positive semi-definite).

The earlier discussion then suggests three possible ways to proceed: historical variance–covariance estimation, multivariate EWMA, and multivariate GARCH.

5.3.2 Historical Variance–Covariance Estimation

This is the most straightforward approach to the estimation of variance–covariance matrices: we choose our window size n and estimate our volatilities and covariances simultaneously. However, this approach has the same drawbacks as historical estimation applied to individual volatilities and correlations: it is strictly accurate only if the 'true' variance–covariance (or correlation) matrix is constant over time, which is a condition that will never be satisfied in practice, and it suffers from pronounced ghost effects.

5.3.3 Multivariate EWMA

We can avoid these drawbacks by estimating our variance–covariance matrix using multivariate EWMA. This is more flexible (i.e., it accommodates changing volatilities and covariances over time) and has less pronounced ghost effects. However, in applying multivariate EWMA, we have to choose the number of separate λ (or decay) terms in our system, and this leads to a dilemma: ideally, we would want each volatility and covariance to have its own specific decay factor, so that we get the best fit for each of these estimates; however, a large number of different λ values can be difficult to handle, and there is no guarantee that they will produce a positive definite estimate of our variance–covariance matrix. These considerations led RiskMetrics to choose a single decay factor – $\lambda = 0.94$ – when estimating the variance–covariance matrix of daily returns in their model.

5.3.4 Multivariate GARCH

Since GARCH approaches are generally better than EWMA ones, theory suggests that we should, ideally, prefer to estimate our variance–covariance matrices using a multivariate GARCH approach.[13] However, this is easier said than done, as multivariate GARCH models

[13] But whether we would in practice is another matter. Leaving aside the difficulties of estimating variance–covariance matrices using GARCH methods, evidence also suggests that simple EWMA matrices are sometimes best when the matrices are used for

need a lot of parameters, and the need to estimate these parameters restricts the size of the systems we can handle. Multivariate GARCH systems are also prone to convergence-in-estimation problems, making it difficult to obtain reliable estimates of all our parameters. As a result of these problems, unrestricted multivariate GARCH systems are only practically feasible if we have a relatively small number of different return series – say, no more than 10.

To give a flavour of the issues involved, consider one of the standard multivariate GARCH models – the BEKK model, named after its authors, Baba, Engle, Kraft and Kroner. This model takes the following matrix form:

$$\boldsymbol{\Sigma}_t = \mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{B}^{\mathbf{T}}\mathbf{x}_{t-1}^{\mathbf{T}}\mathbf{x}_{t-1}\mathbf{B} + \mathbf{C}^{\mathbf{T}}\boldsymbol{\Sigma}_{t-1}\mathbf{C} \tag{5.35}$$

where there are n different returns, $\boldsymbol{\Sigma}_t$ is the $n(n+1)/2$ matrix of (distinct) conditional variance and covariance terms at t, $\mathbf{x}_t$ is the $1 \times n$ vector of errors, and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $n \times n$ matrices. This model imposes no (questionable) cross-equation restrictions, and ensures that our variance–covariance matrix will be positive definite.

However, a major problem with this model is that it has a lot of parameters. For example, with only two different factors (i.e., $n = 2$), the BEKK model involves 11 different parameters, and the number of parameters rapidly rises as n gets larger. This model therefore requires far too many parameters to be used for large dimensional problems. Of course, we can reduce the number of parameters by imposing restrictions on the parameters, but these only help us so much, and the restrictions themselves can create problems of their own.

Various solutions have been suggested to these problems. One simple suggestion is to use GARCH to forecast the volatilities, but assume that correlations are constant. A more sophisticated response is to apply principal components to the risk factors, and then carry out a GARCH analysis on the principal components. An example of this is the orthogonal GARCH model first suggested by Alexander and Chibumba (1997).[14] To implement this approach, we divide our risk factors into groups of highly correlated risk categories and use principal components analysis to orthogonalise each subsystem of risk factors. We then apply univariate GARCH to each of the principal components of each risk category, and 'splice' the results together to produce the large covariance matrix we are really seeking. In principle, this method can be applied to any large-dimensional problem, but care needs to be taken with the initial calibration of the model.[15]

5.3.5 Computational Problems with Covariance and Correlation Matrices

As mentioned earlier, practitioners also face the problem of how to ensure that their estimated covariance (or correlation) matrix will be positive definite (or positive semi-definite). Recall that a symmetric matrix will be positive definite if all its eigenvalues are real and positive, and it will be positive semi-definite if all its eigenvalues are real and non-negative. Consequently, a covariance matrix will be positive definite (positive semi-definite) if all its eigenvalues are real and positive (non-negative), and we can think of the positive definiteness (or otherwise) of a covariance matrix as determined by its eigenvalues. However, even if the true eigenvalues

VaR forecasting (see, e.g., Lopez and Walter (2001), p. 22). Taken at face value, these results might suggest that it is pointless using sophisticated variance–covariance approaches for VaR purposes. However, this conclusion is controversial, and the next footnote gives at least one instance where GARCH models have led to better forecasts.

[14] Other solutions of a similar nature are discussed in the survey by Bauwens *et al.* (2004).

[15] Orthogonal GARCH models appear to be more promising than traditional simple- and weighted-average procedures in forecasting crisis volatilities: Byström (2000) presents evidence that orthogonal GARCH models performed much better than traditional models in dealing with the very high volatility of 1997–98.

satisfy the required properties, the eigenvalues of an estimated covariance matrix might not. For example, risk factors might be very closely correlated, leading to an estimated covariance matrix with eigenvalues that are negative or too close to zero. The estimated covariance matrix would then be singular or nearly singular, and produce pathological underestimates of risk as a result.[16] These problems can be aggravated if estimated covariance matrices are used to choose the portfolio (e.g., as in portfolio optimisation routines). Instead of pathological risk estimates being an occasional problem, we could then find that portfolios are being selected precisely because of their low measured risk – so pathological underestimation of portfolio risks can become the norm rather than the exception.[17]

These are tricky problems. A partial answer is to choose risk factors that are not too closely correlated with each other. We can also try to avoid singularity problems by keeping the number of risk factors down, and one way to do this is to map our assets onto a limited number of underlying risk factors (e.g., using principal components analysis; see also Chapter 12).[18]

We can also check and if necessary adjust the eigenvalues of the covariance matrix, and then recover an adjusted covariance matrix from the adjusted eigenvalues. To see how this works, suppose we have an $m \times 1$ return vector $\mathbf{x}$ with an unknown 'true' covariance matrix $\boldsymbol{\Sigma}$. We have to work with an estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ that may or may not be positive definite (or positive semi-definite). We know from principal components theory that the $m \times 1$ principal components $\mathbf{p}$ have a diagonal covariance matrix $\boldsymbol{\Lambda}$ and a matrix of eigenvectors $\mathbf{A}$, such that $\boldsymbol{\Sigma} = \mathbf{A}^{\mathbf{T}}\boldsymbol{\Lambda}\mathbf{A}$. We begin by estimating the eigenvalues for the estimated covariance $\hat{\boldsymbol{\Sigma}}$. Call the estimated diagonal eigenvalues matrix $\hat{\boldsymbol{\Lambda}}$. We now check that each diagonal element of $\hat{\boldsymbol{\Lambda}}$, $\hat{\lambda}_i$, is positive (if we want a positive definite covariance matrix) or non-negative (if we want a positive semi-definite one). If this is the case, all is well and good and we know that our estimated covariance matrix meets the required positive definiteness or positive semi-definiteness condition. If any of the eigenvalues fail the required condition (i.e., are non-positive, or negative, respectively), we then adjust them so that they meet the condition. So, for example, if a particular eigenvalue is negative and we want positive definiteness, we might add a small positive amount ε to it, so the adjusted eigenvalue is positive as required. Having made the relevant adjustments, we now have an adjusted eigenvalue matrix, $\hat{\boldsymbol{\Lambda}}^*$ say, that meets the required condition. We then recover an adjusted covariance matrix, $\hat{\boldsymbol{\Sigma}}^*$, using $\hat{\boldsymbol{\Sigma}}^* = \mathbf{A}^{\mathbf{T}}\hat{\boldsymbol{\Lambda}}^*\mathbf{A}$, and we can be confident that $\hat{\boldsymbol{\Sigma}}^*$ is positive definite (or positive semi-definite) because the eigenvalues are all positive (or non-negative). This checking and adjustment process can also be automated to ensure that we always have a covariance matrix to work with that satisfies the desired positive definiteness (or semi-definiteness) properties.

[16] We can also get more blatant rank defects: if the number of observations in our data set is not at least as large as the number of risk factors, the covariance matrix will have a rank defect and be singular, and any results we might get will be worthless.

[17] This is the awkward problem of endogenous model risk covered in Box 16.2.

[18] The best approach also depends on the context (e.g., large matrices are considerably more difficult to estimate than small ones, etc.). For more on these sorts of issues and other methods of cleaning data and/or generating positive definite or semi-definite covariance matrices, see, e.g., Davé and Stahl (1997), Kreinin and Levin (2000), Alexander (2000) or Jäckel (2002), Chapter 6.

Appendix
Modelling Dependence: Correlations and Copulas

This appendix deals with the important but often misunderstood issue of dependence between risky variables. If we have two risky variables X and Y, how do we model the relationship or dependence, if any, between them? The standard answer is to measure this dependence by means of their (linear) correlation, and then use this correlation (or correlation matrix, if we have more than two variables) to derive the multivariate distribution $F(x,y)$ taking this dependence into account. Indeed, there is a tendency in risk measurement to use correlation as if it were an all-purpose dependence measure, with the result that correlation is often misused and applied to problems for which it is not suitable:

> as well as being one of the most ubiquitous concepts in modern finance and insurance, [correlation] is also one of the most misunderstood concepts. Some of the confusion may arise from the literary use of the word to cover any notion of dependence. To a mathematician correlation is only one particular measure of stochastic dependence among many. It is the canonical measure in the world of multivariate normal distribution functions, and more generally for spherical and elliptical distributions. However, empirical research in finance and insurance shows that the distributions of the real world are seldom in this class.[19]

Thus, the correlation is merely one dependence measure out of many, and we need to establish where it is a good dependence measure and where it is not. Where it is not, we need an alternative dependence measure that is reliable.

A5.1 CORRELATION AS A MEASURE OF DEPENDENCE

We begin by recalling some basic terms. Let us consider two random variables X and Y.[20] These have marginal distributions $F_x(x) = \Pr\{X \leq x\}$ and $F_y(y) = \Pr\{Y \leq y\}$, which give cumulative probabilities for each variable considered on its own. The vector of risky variables $[X,Y]$ has a joint distribution function $F(x,y) = \Pr\{X \leq x, Y \leq y\}$, which gives the probability that X is no more than x and that Y is no more than y. The marginal distribution function looks at each variable separately, and the joint distribution takes into account the dependence structure between them.

The most common way to measure this dependence is to use standard (i.e., linear) correlation, defined as:

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{A5.1}$$

where σ_X and σ_Y are the standard deviations of X and Y and $\text{cov}(X,Y)$ is their covariance. But when is linear correlation a reliable measure of dependence?

[19] Embrechts *et al.* (1999b), p. 2.
[20] We consider two random variables for convenience, but all our analysis generalises easily to the case of n random variables.

A5.1.1 Correlation Good with Elliptical Distributions

It turns out that (linear) correlation is a good measure of dependence when our random variables are distributed as multivariate elliptical. Elliptical distributions include the normal and t-distributions as special cases, and are attractive because they are less restrictive than the normal distribution, but retain much of its attractiveness and tractability. In particular, under elliptical conditions, the VaR can be determined from knowledge of variances, correlations and positions sizes alone: in other words, correlation tells us everything we need to model dependence.

That said, correlation still has its limitations even in an elliptical world. If risks are independent, we will have a correlation of zero, but the reverse does not necessarily hold except in the special case where we have a multivariate normal distribution: so zero correlation does not imply that risks are independent unless we have multivariate normality. Another drawback is that correlation is not invariant to transformations of the underlying variables: for instance, the correlation between X and Y will not in general be the same as the correlation between $\ln(X)$ and $\ln(Y)$. Hence, transformations of our data can affect our correlation estimates.[21]

A5.1.2 Correlation Bad with Non-elliptical Distributions

Unfortunately, the correlation measure runs into more serious problems once we go outside elliptical distributions. One problem is that correlation is not even defined unless variances are finite, and this is obviously a problem if we are dealing with a heavy-tailed distribution that has an infinite variance (e.g., a Lévy distribution with $\alpha < 2$), or if we are dealing with trended return series that are not cointegrated.

But even where correlations are defined, outside an elliptical world we cannot count on correlations varying all the way from –1 to 1. This means that perfectly positively (or negative) dependent variables do not necessarily have a correlation of 1 (or –1). For instance, suppose X is normally distributed with mean 0 and standard deviation 1. Now define $Y_1 = \exp(X)$, $Y_2 = \exp(2X)$, and $Y_3 = \exp(-X)$. Clearly, Y_1 and Y_2 move perfectly with each other (i.e., are comonotonic), while Y_1 and Y_3 move perfectly against each other (i.e., are countermonotonic). If we now estimate their correlations, we find that the correlation coefficient between Y_1 and Y_2 is 0.731, and that between Y_1 and Y_3 is –0.367. So in this case two variables that move perfectly with each other have a correlation of 0.731, and two that move perfectly against each other have a correlation of -0.367. The correlation therefore gives a misleading indication of dependence.

More seriously, the marginal distributions and correlations no longer suffice to determine the joint multivariate distribution: so correlation no longer tells us everything we need to know about dependence. Indeed, correlation sometimes tells us very little about dependence, *especially in the tails*.

[21] This drawback means that we have to be wary of transformation effects when using linear correlations. However, this drawback is peculiar to linear correlation, and can be avoided if we use rank correlation measures (e.g., Spearman's rank correlation). Rank correlations also have the attraction that they can always vary from -1 to $+1$, and are defined even if our variables have infinite variances. On the other hand, rank correlations are less useful than linear correlations (e.g., because they have little use in portfolio optimisation).

A5.2 COPULA THEORY

A5.2.1 Basics of Copula Theory

We need an alternative dependence measure, and the answer is to be found in the theory of copulas. The term 'copula' comes from the Latin. It refers to connecting or joining together, and is closely related to more familiar English words such as 'couple'. However, the 'copulas' we are speaking of here are statistical concepts that refer to the way in which random variables relate to each other: more precisely, a copula is a function that joins a multivariate distribution function to a collection of univariate marginal distribution functions. We take the marginal distributions – each of which describes the way in which a random variable moves 'on its own' – and the copula function tells us how they 'come together' to determine the multivariate distribution. Copulas enable us to extract the dependence structure from the joint distribution function, and so separate out the dependence structure from the marginal distribution functions.[22]

The key result is a theorem due to Sklar (1959). Again suppose for convenience that we are concerned with only two random variables, X and Y. If $F(x,y)$ is a joint distribution function with continuous marginals $F_x(x) = u$ and $F_y(y) = v$, then $F(x,y)$ can be written in terms of a unique function $C(u,v)$:

$$F(x,y) = C(u,v) \tag{A5.2}$$

where $C(u,v)$ is known as the copula of $F(x,y)$. The copula function describes how the multivariate function $F(x,y)$ is derived from or coupled with the marginal distribution functions $F_x(x)$ and $F_y(y)$, and we can interpret the copula as giving the dependence structure of $F(x,y)$.[23]

This result is important because it enables us to construct joint distribution functions from marginal distribution functions in a way that takes account of the dependence structure of our random variables. To model the joint distribution function, all we need to do is specify our marginal distributions, choose a copula to represent the dependence structure, estimate the parameters involved, and then apply the copula function to our marginals. Once we can model the joint distribution function, we can then in principle use it to estimate any risk measures (see Chapter 6, section 6.8).

A5.2.2 Common Copulas

There are many different copulas to choose from. Some of the simplest are the independence (or product) copula, the minimum (or comonotonicity) copula, and the maximum (or countermonotonicity) copula:

$$C_{\text{ind}}(u,v) = uv \qquad \text{(independence copula)} \tag{A5.3a}$$

$$C_{\min}(u,v) = \min[u,v] \qquad \text{(minimum copula)} \tag{A5.3b}$$

$$C_{\max}(u,v) = \max[u+v-1,0] \qquad \text{(maximum copula)} \tag{A5.3c}$$

[22] For more on the theory of copulas and/or their financial applications, see, e.g., Nelson (1999) and Cherubini *et al.* (2004).

[23] The copula has various attractive properties as a representation of dependence. One of these (and a property that linear correlation does not have) is scale invariance to (reasonable) transformations of the random variables and/or their distribution functions. This means, for example, that if we change our units of measurement, or convert from P/L to returns, then the copula itself will remain unaffected.

We might use the first where X and Y are independent, the second where they are positively dependent or comonotonic (i.e., rise or fall together), and the third where they are negatively dependent or countermonotonic (i.e., one rises as the other falls).

Other important copulas are the Gaussian (or normal) copula and t-copulas. The normal copula is:

$$C_{\rho}^{Ga}(x,y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{0.5}} \exp\left\{\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} dsdt \tag{A5.4}$$

where $-1 \leq \rho \leq 1$ and Φ is the univariate standard normal distribution function. Note that this copula depends only on the correlation coefficient, ρ, which confirms that ρ is sufficient to determine the whole dependence structure. Variables that have standard normal marginal distributions and this dependence structure will then be multivariate normally distributed. Unfortunately, the Gaussian copula does not have a closed-form solution, so the copula has to be estimated by numerical methods (e.g., fast Fourier transforms).

The t-copula for υ degrees of freedom is a straightforward generalization of the normal one:

$$C_{\upsilon,\rho}^{t}(x,y) = \int_{-\infty}^{t_{\upsilon}^{-1}(x)} \int_{-\infty}^{t_{\upsilon}^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{0.5}} \exp\left\{1 + \frac{(s^2 - 2\rho st + t^2)}{\upsilon(1-\rho^2)}\right\}^{-\frac{\upsilon+2}{\upsilon}} dsdt \tag{A5.5}$$

where $t_{\upsilon}^{-1}(x)$ is the inverse of the distribution function of the standard univariate t-distribution for υ degrees of freedom.

Another important copula is the Gumbel or logistic copula:

$$C_{\beta}^{Gu}(x,y) = \exp\left[-\left\{(-\log x)^{\frac{1}{\beta}} + (-\log y)^{\frac{1}{\beta}}\right\}^{\beta}\right] \tag{A5.6}$$

where β satisfies $0 < \beta \leq 1$ and determines the amount of dependence between our variables: $\beta = 1$ indicates that the variables are independent, $\beta > 0$ indicates limited dependence, and the limiting value of 0 indicates perfect dependence. This copula is consistent with EV theory – unlike, say, the Gaussian copula – and could reasonably be used to model multivariate extremes.

Copulas can also be grouped into distinctive families. These include the elliptical and Archimedean copulas. The latter copulas are particularly useful and can be written in the form:

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{A5.7}$$

where $\varphi(.)$ is strictly decreasing and convex, and $\varphi(1) = 0$. Archimedean copulas are easy to use and can fit a wide range of dependence behaviour. Another important family is the family of extreme-value (EV) copulas. These arise from a multivariate version of the Fisher–Tippett theorem (1928), the upshot of which is that if $G(x,y)$ is a multivariate generalised extreme value (GEV) distribution, then $G(x,y)$ must have a copula that satisfies the condition:

$$C(u^t, v^t) = [C(u,v)]^t \tag{A5.8}$$

If we are dealing with multivariate extremes, we should use a copula that satisfies this condition. The EV copulas include, among others, the product, minimum and Gumbel copulas, as well

as others such as the Gumbel II and Galambos copulas:

$$\text{Gumbel II copula} = uv \exp\left[\frac{\beta uv}{u+v}\right] \tag{A5.9a}$$

$$\text{Galambos copula} = uv \exp\left[\left(u^{-\beta} + v^{-\beta}\right)^{-\frac{1}{\beta}}\right] \tag{A5.9b}$$

which are defined over β ranges of [0,1] and [0,∞) respectively.

A5.2.3 Tail Dependence

Copulas can also be used to investigate tail dependence, which is an asymptotic measure of the dependence of extreme values. Tail dependence is an important issue because extreme events are often related (i.e., disasters often come in pairs or more), and models that fail to accommodate their dependence can literally lead a firm to disaster.

If marginal distributions are continuous, we can define a coefficient of (upper) tail dependence of X and Y as the limit, as $\alpha \to 1$ from below, of

$$\Pr[Y > F_y^{-1}(\alpha) | X > F_x^{-1}(\alpha)] = \lambda \tag{A5.10}$$

provided such a limit exists.[24] If $0 < \lambda \leq 1$, X and Y are asymptotically dependent in the upper tail; and if $\lambda = 1$, they are asymptotically independent.

Tail dependence is an asymptotic property of the copula. For example, if we have a Gumbel copula, we can show that

$$\lambda = 2 - 2^{\beta} \tag{A5.11}$$

so X and Y are asymptotically dependent provided that $\beta < 1$. So long as this condition holds, X and Y will appear to be dependent however far into the tail we go. We can also show that the Gaussian copula gives asymptotic independence provided that $\rho < 1$: if this condition holds, extreme events will appear to be independent if we go far enough out into the tail. By contrast, if X and Y are t-distributed, they will be asymptotically dependent provided that $\rho > -1$: X and Y will be asymptotically dependent *unless* they are perfectly negatively correlated. It follows, then, that if $-1 < \rho < 1$, a t-distribution will show asymptotic dependence while a normal distribution will not. This difference arises because of the heavier tails of the t-distribution.

A5.3 ESTIMATING COPULAS

Estimating copulas involves choosing the copula functional form and then estimating the parameter(s) involved. We can estimate these parameters using parametric or non-parametric methods. In both cases, the idea is to find estimators that best fit the data by some criterion. For example, parametric approaches find the estimators that maximise a likelihood function, and require us to solve maximum likelihood equations. However, such methods are quite difficult to implement. Non-parametric methods are easier to implement, and there are various non-parametric methods to choose from.[25]

[24] See, e.g., Embrechts *et al*. (1999b) p. 18.

[25] For more on the estimation of copulas, see, e.g., Bouyé *et al*. (2000), Scaillet (2000a), pp. 6–15, and the references referred to in note 22.

A5.4 CONCLUSIONS

This appendix has emphasised two key points. The first is that we should not confuse correlation and dependence: correlation is only one measure of dependence, and it is important for risk practitioners to be aware of its limitations. In particular, correlation can be used to measure dependence only when risks are elliptical. The second point follows from the first: with more general distributions, we need a more reliable dependence measure. The answer is to use copulas, but they are not particularly easy to use.

6

Parametric Approaches (I)

This chapter and the next look at parametric approaches to risk measurement. These approaches estimate risk by fitting probability curves to the data and then inferring the risk measure – such as the VaR or the ES – from the fitted curve. Parametric approaches are more powerful than non-parametric ones, because they make use of additional information contained in the assumed density or distribution function. Many parametric approaches also give rise to straightforward VaR (and sometimes ES) formulas, which makes them easy to use. However, parametric approaches are vulnerable to error if the assumed density function does not adequately fit the data. The quality of parametric estimates of risk measures also depends on the quality of the parameter estimates used, so we have to make sure we have enough data to estimate parameters reliably and that we use estimation methods appropriate to the distribution(s) we choose.

In applying any parametric approach, the objective is to make assumptions that are consistent with the features of the empirical processes we are trying to model. Empirical return processes are often characterised by heavy tails, or excess kurtosis, and they are sometimes skewed as well. We would therefore often want to fit a parametric model that can accommodate heavy tails, and possibly also skews. However, it does not follow that all we then do is unconditionally fit some distribution that has these features. Fitting a distribution unconditionally to the data ignores the important stylised fact that empirical return processes also exhibit volatility clustering, and volatility clustering can also lead to excess kurtosis. If we wish to take account of volatility clustering – which we should – then we will often want to fit a distribution to returns, conditional on an assumed volatility process that is itself consistent with volatility clustering. For example, we might fit a normal distribution to the data, conditional on a GARCH volatility process, and this would typically produce a return process that exhibits both tail heaviness and volatility clustering. Thus, we would often think in terms of fitting distributions conditionally rather than unconditionally, and simplistic assumptions such as normality often fit the data much better when applied in this way.

Getting down to basics, we must first choose the random variable to which we will apply a parametric distribution: do we wish to apply it to P/L, losses, returns, etc.? In making this decision, we must also specify any conditionality in the way in which the chosen distribution is to be fitted. We discuss these issues in section 6.1. We then move onto our core material. Broadly speaking, we discuss parametric estimation of risk measures at two different levels – at the portfolio level, where we are dealing with portfolio P/L (or losses or returns); and at the subportfolio or individual-position level, where we deal with the P/L (or losses, etc.) to individual positions. Beginning at the portfolio level, sections 6.2–6.5 discuss alternative parametric approaches based on the normal, *t*, lognormal, and various other distributions. After that, we turn to the position level, and sections 6.6–6.7 deal with variance–covariance approaches in which asset returns are assumed to follow multivariate normal or related distributions, and section 6.8 deals with position-level approaches based on copulas. Some conclusions are offered in section 6.9. This chapter does not deal with parametric approaches based on extreme-value

theory (EVT); EVT approaches are a major subject in their own right and discussed in the next chapter.[1]

6.1 CONDITIONAL VS UNCONDITIONAL DISTRIBUTIONS

Let us suppose that we wish to apply a particular parametric approach to, say, daily portfolio returns R_t. This implies that daily returns will be 'driven off' some random noise process ε_t that obeys a chosen distribution, but we need to specify how returns depend on this process. In the simplest case, we might specify that the two are the same:

$$R_t = \varepsilon_t \tag{6.1}$$

An example is where $\varepsilon_t \sim N(\mu_t, \sigma_t^2)$. To estimate any quantile of R_t, $q(R_t)$, we estimate the parameters μ_t and σ_t, and $q\,(R_t)$, is equal to its ε_t equivalent $q(\varepsilon_t)$, i.e.,

$$q\,(R_t) = q\,(\varepsilon_t) \tag{6.2}$$

In practice μ_t is often set to 0 in very short-horizon problems, and σ_t would be estimated by some appropriate (e.g., EWMA or GARCH) method and would typically vary from one day to the next. In such cases we can say that R_t is conditionally normal, with the conditionality referring to dependence on σ_t. We can also say that standardised returns (R_t/σ_t) are distributed as standard normal:

$$\frac{R_t}{\sigma_t} \sim N(0, 1) \Rightarrow q(R_t) = \sigma_t q(z_t) \tag{6.3}$$

where z_t is standard normal. The R_t quantile, $q(R_t)$, is the product of σ_t and a standard normal quantile.

More generally, we often have situations where R_t depends on a deterministic component d_t as well as a random component ε_t. In such cases R_t is a function of both components:

$$R_t = f\,(d_t, \varepsilon_t) \tag{6.4}$$

R_t is now a function of ε_t conditional on a given d_t. Typically, this function will be a simple linear one:

$$R_t = d_t + \varepsilon_t \tag{6.5}$$

The deterministic component d_t will depend on the context. It might, for example, reflect seasonal or holiday effects, and it might be determined judgementally or by regression analysis. $q(R_t)$ will now be the sum of d_t and the noise quantile $q(\varepsilon_t)$:

$$q(R_t) = d_t + q(\varepsilon_t) \tag{6.6}$$

For convenience we will largely ignore d_t in our later discussions, but it could sometimes matter in practice.

There is also a related issue. Consider for the sake of argument a simple case such as Equation (6.3). This tells us that R_t is conditionally normal. To appreciate what this means, note that Equation (6.3) tells us that R_t is based on a normal noise process ε_t. However, except in the special case where σ_t is constant over time, R_t itself will not be normal: R_t is only normal conditional on σ_t. In general, the distribution of R_t will also depend on the behaviour

[1] As with Chapter 4, we will focus on the estimation of VaR and ES in order to keep the discussion straightforward, but the methods discussed here extend easily to the estimation of coherent risk measures. These measures and their standard errors or confidence intervals can be estimated using the methods discussed in Chapter 3, sections 3.4 and 3.5.2.

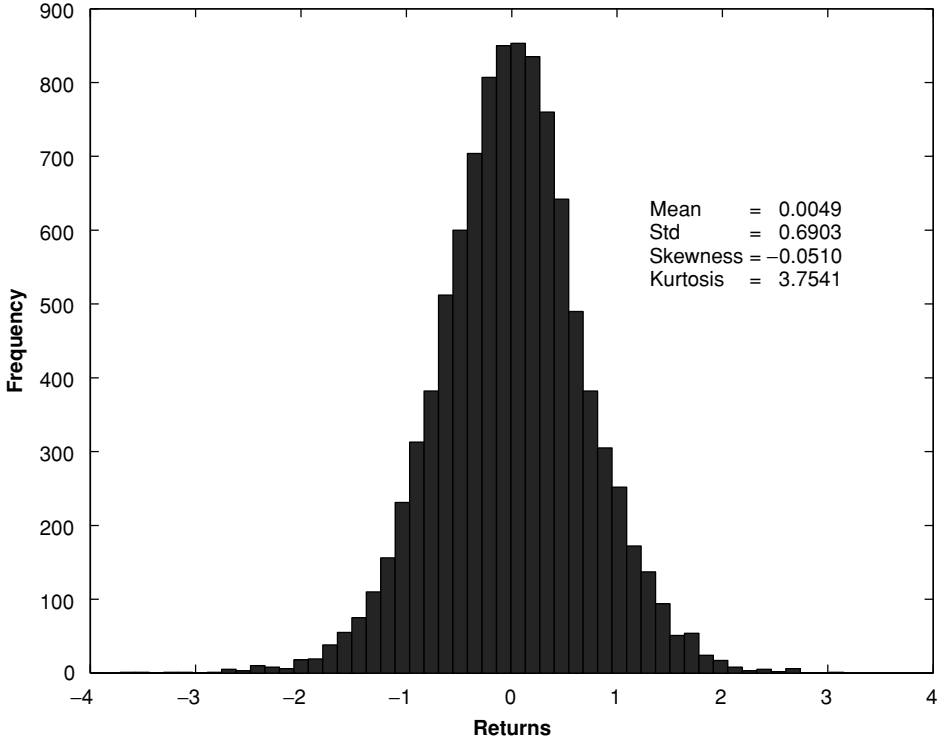

Figure 6.1 Returns under stochastic volatility
Note: Based on a simulated sample of 10 000 observations drawn from a GARCH(1, 1) process with $\omega = 0.025$, $\alpha = 0.1$ and $\beta = 0.85$, and starting values set equal to long-run values.

of σ_t, and σ_t will usually be random. In more complex cases, the distribution of returns will depend on the processes driving other random parameters as well.

To illustrate the impact of the stochastic volatility process, suppose that σ_t follows a GARCH(1, 1) process with $\omega = 0.025$, $\alpha = 0.1$ and $\beta = 0.85$:

$$\sigma_t^2 = 0.025 + 0.1R_{t-1}^2 + 0.85\sigma_{t-1}^2 \tag{6.7}$$

This means that R_t is actually drawn from a normal with a stochastic volatility, and the stochastic volatility is important because it can make the returns both skewed and heavy-tailed. To show this, Figure 6.1 gives a histogram of 10 000 drawings of R_t from a single run of this process. The stochastic volatility makes quite a difference: the mean return is a little positive (0.0049), even though standardised returns are drawn from a distribution with a zero mean, and the standard deviation of returns (0.6903) is a little below the 'long-run' value 0.7071. Most significantly, returns have a negative skew (-0.051) and greater than normal kurtosis (3.7541). Stochastic volatility can therefore create both skewness and excess kurtosis.

The general point is that there is a difference between conditional and actual return processes: even if returns *conditionally* follow a particular distribution, this does not mean that *actual* returns will follow the same distribution. Indeed, the distributions governing conditional and actual returns will usually be quite different.

Having made this point, we now focus on the choice of distribution.

6.2 NORMAL VaR AND ES

The normal (or Gaussian) distribution was briefly introduced in Chapter 2. It is very widely used, and has plausibility in many contexts because of the central limit theorem. It is also attractive because it requires only two independent parameters – a mean μ and standard deviation σ (or its square, the variance σ^2). The normal distribution is also convenient because it produces straightforward formulas for both VaR and ES. If we apply a normal distribution to P/L,[2] then the VaR and ES are:

$$VaR = -\mu_{P/L} + \sigma_{P/L} z_\alpha \tag{6.8a}$$

$$ES = -\mu_{P/L} + \sigma_{P/L} \frac{\phi(z_\alpha)}{1-\alpha} \tag{6.8b}$$

where $\mu_{P/L}$ and $\sigma_{P/L}$ have their obvious meanings, and z_α is the standard normal variate corresponding to our chosen confidence level (e.g., $z_\alpha = 1.645$ if we have a confidence level $\alpha = 0.95$), and $\phi(.)$ is the value of the standard normal density function.[3] However, in most cases the mean and standard deviation are not known, so we have to work with estimates of them, m and s. Our estimates of VaR and ES, VaR^e and ES^e, are then:[4]

$$VaR^e = -m_{P/L} + s_{P/L} z_\alpha \tag{6.9a}$$

$$ES^e = -m_{P/L} + s_{P/L} \frac{\phi(z_\alpha)}{1-\alpha} \tag{6.9b}$$

Figure 6.2 shows the standard normal L/P pdf curve and normal 95% VaR and 95% ES. The normal pdf has a distinctive bell-shaped curve, and the VaR cuts off the top 5% tail while the ES is the probability-weighted average of the tail quantiles.

Example 6.1 (Normal VaR and ES)

Suppose P/L is normally distributed with mean 0 and standard deviation 1. Then Equation (6.8a) tells us that the 95% VaR is 1.645, and Equation (6.8b) tells us that the 95% ES is $\phi(1.645)/(1-0.95) = 0.1031/0.05 = 2.063$.

One of the nice features of parametric approaches is that the formulas they provide for VaR (and, where they exist, ES) also allow us to estimate these risk measures at any confidence level or holding period we like. In the normal case, we should first note that Equation (6.8) gives the VaR and ES formulas for a confidence level reflected in the value of z_α, and for a holding period equal to the period over which P/L is measured (e.g., a day). If we change the confidence level α from 0.95 to 0.99, z_α changes from 1.645 to 2.326.

[2] As discussed already in Chapter 3, we can apply a normality assumption to portfolio P/L, losses or returns. The formulas then vary accordingly: if we assume that losses are normal, our VaR and ES formulas are the same as in Equation (6.8), except for the mean terms having a reversed sign; if we assume that arithmetic returns are normal, then the μ and σ terms refer to returns, rather than P/L, and we need to multiply our VaR and ES formulas by the current value of our portfolio. The case of normality applied to geometric returns is considered in Section 6.4.

[3] The formula for ES can also be generalised to the broader class of elliptical distributions. For more on elliptical ES, see, e.g., Landsman and Valdez (2003).

[4] The normality assumption has the additional attraction of making it easy for us to get good estimators of the parameters. As any econometrics text will explain, under normality, least squares (LS) regression will give us best linear unbiased estimators of our parameters, and these are also the same as those we would get using a maximum likelihood approach. See also Box 6.1.

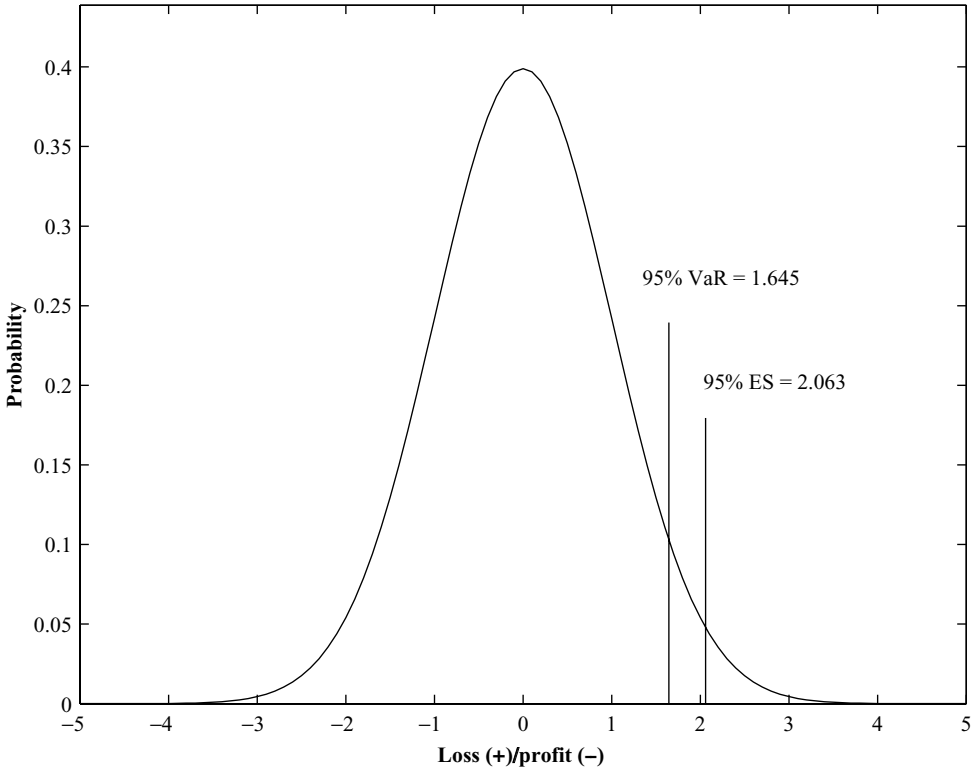

Figure 6.2 Normal VaR and ES
Note: Based on a 1-day holding period with L/P having a mean 0 and standard deviation 1. The figure is obtained using the 'normalesfigure' function.

To take account of a change in the holding period, we need formulas for the mean and standard deviation of P/L over arbitrary periods. If we now define $\mu_{P/L}$ and $\sigma_{P/L}$ as the mean and standard deviation of P/L over a given period (e.g., a day), then the mean and standard deviation of P/L over h such periods are:

$$\mu_{P/L}(h) = h\mu_{P/L} \tag{6.10a}$$

$$\sigma^2_{P/L}(h) = h\sigma^2_{P/L} \Rightarrow \sigma_{P/L}{}^{(h)} = \sqrt{h}\sigma_{P/L} \tag{6.10b}$$

We now substitute these into Equation (6.8) to get the formulas for VaR and ES over an arbitrary holding period h and confidence level α:

$$VaR(h, \alpha) = -h\mu_{P/L} + \sqrt{h}\sigma_{P/L} z_\alpha \tag{6.11a}$$

$$ES(h, \alpha) = -h\mu_{P/L} + \sqrt{h}\sigma_{P/L} \frac{\phi(z_\alpha)}{1 - \alpha} \tag{6.11b}$$

These formulas make it very easy to measure VaR and ES once we have values (or estimates) of $\mu_{P/L}$ and $\sigma_{P/L}$ to work with. These formulas tell us that VaR and ES will rise with the confidence level, as shown earlier in Figure 2.6. However, the effects of a rising holding period are ambiguous, as the first (i.e., μ) terms in each formula fall with h, but the second (i.e., σ) terms rise. Since the μ terms are proportional to h, while the σ terms are proportional to the

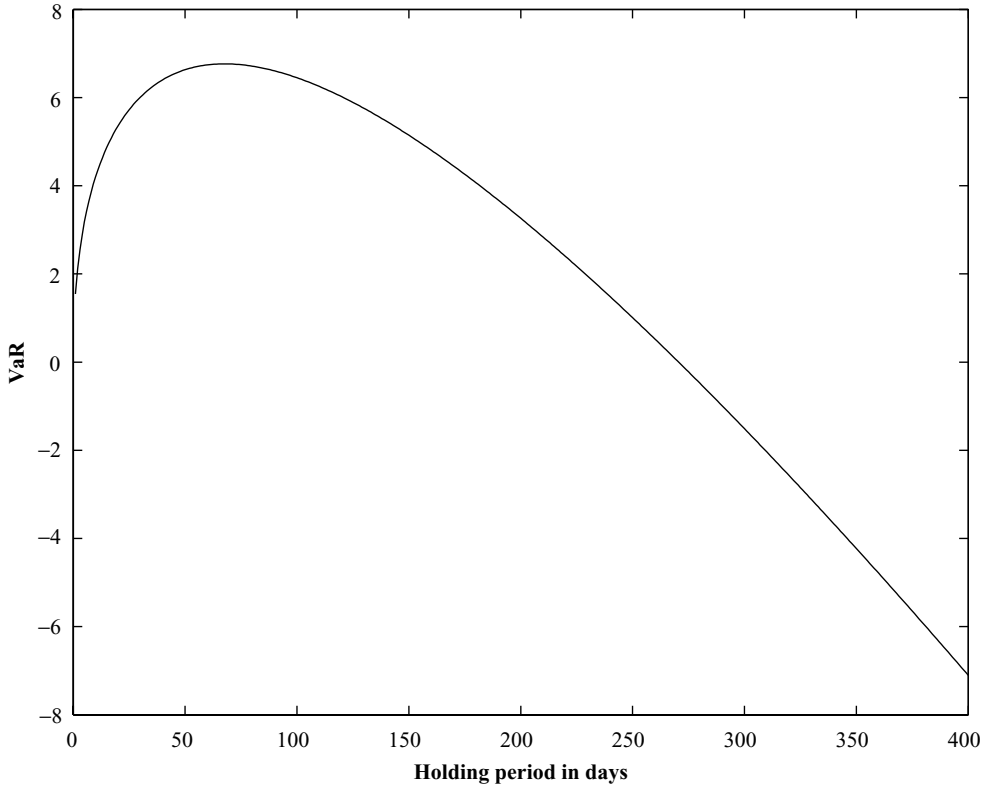

Figure 6.3 Normal VaR and holding period
Note: This figure is obtained using the 'normalvarplot2D_hp' function and shows the 95% VaR for a normal P/L distribution with mean 0.1 and standard deviation 1.

square root of h, the μ terms will become more prominent as h gets larger. If we assume plausible parameter values (e.g., $\alpha = 0.95$, and μ positive but 'small' relative to σ), we then get the following:

- When h is very low, the σ terms dominate the μ terms, so the VaR and ES are positive.
- As h gets bigger, the μ terms grow at a faster rate, so VaR and ES will rise but at a diminishing rate.
- As h continues to rise, VaR and ES will turn down, and eventually become negative.
- Thereafter, they will continue to fall as h continues to rise.

Figure 6.3 gives an example of how VaR behaves under these conditions. (We get a similar-looking chart for the ES.) In this particular case (with daily parameters of $\mu = 0.1$ and $\sigma = 1$), the VaR peaks at a holding period of around 70 days, and becomes negative at a holding period of around 250 days. VaRs beyond that holding period continue to fall as the holding period continues to rise.

It is also worth emphasising that the VaR (and ES) will always have the same basic shape provided that $\mu > 0$, and the value of μ merely affects the timing of the peak and the rate of subsequent decline: the bigger is μ, the quicker it takes for the VaR to peak and the more rapid its subsequent rate of decline.

Had the mean been zero, on the other hand, our VaR and ES equations (6.11) would have reduced to:

$$VaR(h, \alpha) = \sqrt{h}\sigma_{P/L} z_\alpha \tag{6.12a}$$

$$ES(h, \alpha) = \sqrt{h}\sigma_{P/L} \frac{\phi(z_\alpha)}{1 - \alpha} \tag{6.12b}$$

and it is immediately clear that both VaR and ES would rise indefinitely with the square root of the holding period h.

Example 6.2 ('True' vs square-root VaR)

Suppose our P/L is distributed as normal, with mean 10 and standard deviation 25 per day. Then Equation (6.8a) tells us that the one-day 95% VaR is $-10 + 25 \times 1.645 = 31.125$. Using Equation (6.11a), the 'true' VaR over a 5-day horizon is $-10 \times \sqrt{5} + 25 \times \sqrt{5} \times 1.645 = 41.958$, and over a 10-day horizon $-10 \times 10 + 25 \times \sqrt{10} \times 1.645 = 30.049$. However, the square-root formula (Equation (6.12)) gives us 5-day and 10-day VaRs of $\sqrt{5} \times 25 \times 1.645 = 91.958$ and $\sqrt{10} \times 25 \times 1.645 = 130.049$. These results illustrate the unreliability of the square-root rule, particularly over longer horizons.

We should also keep in mind the disadvantages of the normality assumption. One disadvantage is that it allows P/L to take any value, which implies the possibility of arbitrarily large losses. However, it is usually the case (e.g., due to limited liability and similar constraints) that our losses are bounded. The failure of the normality assumption to respect constraints on the maximum loss can then lead to major overestimates of what we really stand to lose.

A second potential problem is one of statistical plausibility. As mentioned already, the normality assumption is often justified by reference to the central limit theorem, but this only applies to quantities and probabilities in the central mass of the density function, rather than the tails. When dealing with extremes, we should refer to extreme value theory, which tells us very clearly *not* to use normality to model extremes.

A third problem is that most financial returns have excess kurtosis, or fatter than normal tails, and a failure to allow for excess kurtosis can lead to a major underestimation of the VaR or ES. The implications of excess kurtosis are illustrated in Figure 6.4. This figure shows both the standard normal pdf and a particular type of heavy-tailed pdf, a t pdf with zero mean, unit standard deviation and 5 degrees of freedom. The impact of excess kurtosis is seen very clearly in the tails: excess kurtosis implies that tails are heavier than normal, and this means that VaRs (at the relatively high confidence levels we are usually interested in) will be bigger. For example, if we take the 99% confidence level, the standard normal VaR is 2.326, but the t VaR is 2.610 – which is 12% bigger. It is also obvious from the figure that the difference between the two VaRs gets bigger as the confidence level continues to rise.

We should also keep in mind the difference between conditional and actual distributions highlighted in the previous section. One of the most important stylised features of empirical return processes is that they exhibit volatility clustering, and we can model such clustering by assuming that volatility follows a process such as a GARCH one. As we saw earlier, if we apply such a process to conditionally normal returns, then the combination of volatility clustering and

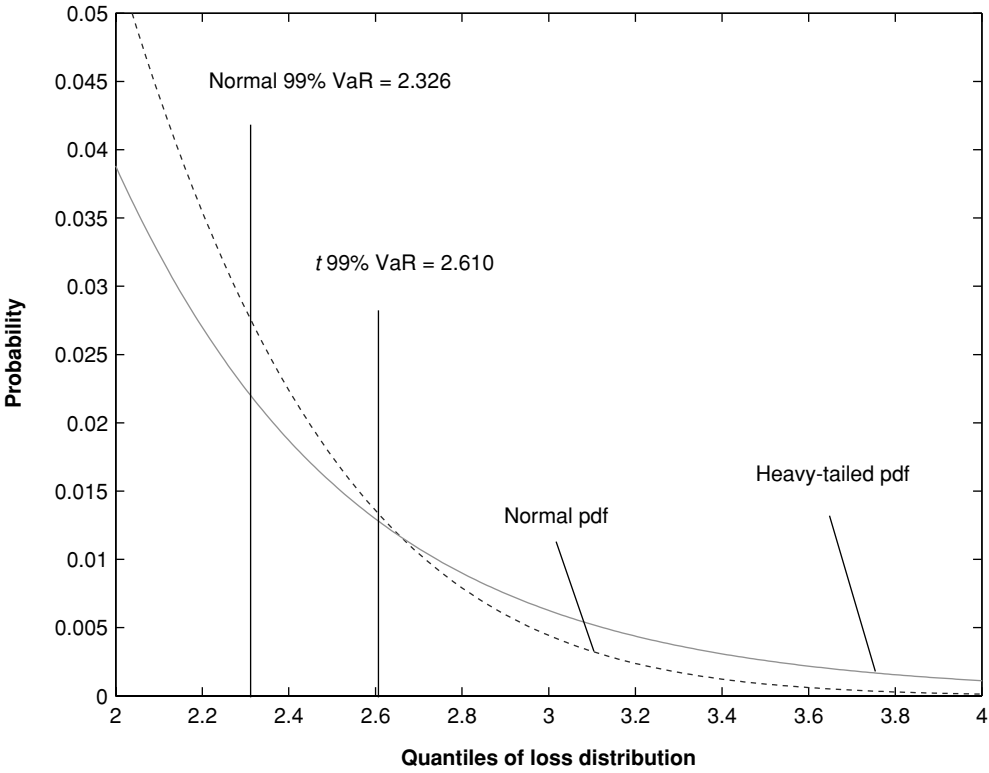

Figure 6.4 Normal VaR vs heavy-tailed VaR
Note: Both distributions have zero mean and unit standard deviation, and the t has 5 degrees of freedom.

conditional normality can (and typically will) make actual returns heavy tailed (and possibly skewed as well). This implies that conditional normality with stochastic volatility will usually provide a much better fit to empirical return processes than an unconditionally normal process. This said, it is often the case that even conditional normality is unable to fully accommodate the heavy tails of empirical return processes: the best fits usually come when we conditionally apply a heavy-tailed distribution (such as a t) with a stochastic volatility process.

Box 6.1 Parameter Estimation

To apply any parametric approach, we need to estimate the parameters involved, and it is important to choose an estimation method that is suitable for the distribution we are dealing with. There are various approaches to parameter estimation, including least squares (LS), maximum likelihood (ML), semi-parametric, robust estimation methods, and methods of moments approaches, and these methods are explained in most econometrics textbooks.

If we are working with a normal distribution, then LS and ML methods coincide, and LS/ML methods are theoretically ideal and straightforward to implement. LS/ML methods would therefore be our preferred choice if we are working with normal distributions or distributions (such as the lognormal) that can be transformed into the normal. If we are

working with a t-distribution, then LS and ML differ: ML is theoretically superior if we have chosen the right t-distribution, but LS is more reliable if we are concerned that we might have misspecified the distribution. And if we are working with other distributions, parameter estimation can become more problematic: ML methods are still theoretically best, but are often harder to implement and less reliable in practice (e.g., because they may involve the solution of non-linear equations, the likelihood function might have multiple maxima or be relatively flat in the neighbourhood of its maximum). In such circumstances, we might use semi-parametric or robust methods instead: a good instance is where we might use semi-parametric estimators such as the Hill estimator to estimate a tail index in an EV model.

When estimating risk measures we are usually more concerned with robustness than is usually the case in conventional econometric practice. Whereas the econometrician is typically concerned with using the *best* method for any given set of parametric assumptions that are taken to be true, the risk manager is more concerned with getting a *good* method for the purpose at hand, given that the assumptions he or she is working with might be wrong: in risk management it is better to have an estimate that is approximately right than a theoretically 'perfect' answer that could be wrong because it is based on an empirically false premise.

6.3 THE *t*-DISTRIBUTION

As we have just seen, one way to accommodate excess kurtosis is to use a t-distribution instead of a normal one. A t-distribution with υ degrees of freedom has a kurtosis of $3(\upsilon - 2)/(\upsilon - 4)$, provided $\upsilon \geq 5$, so we can approximate an observed kurtosis of up to 9 by a suitable choice of υ: if we want a relatively high kurtosis, we would choose a relatively low value for υ, and if we want a relatively low kurtosis, we would choose a high value for υ. This gives us considerable scope to accommodate excess kurtosis, so long as the kurtosis is not too extreme. For risk measurement purposes, we would work with a generalised t-distribution that allows us to specify the mean and standard deviation (or variance) of our P/L or return distribution, as well as the number of degrees of freedom. Using the same notation as before, our VaR is then:

$$VaR(h, \alpha) = -h\mu_{P/L} + \sqrt{h}\sqrt{\frac{\upsilon - 2}{\upsilon}}\sigma_{P/L}t_{\alpha,\upsilon} \tag{6.13}$$

The derivation of this formula is explained in Box 6.2. This t-VaR formula differs from the earlier normal VaR formula, Equation (6.11a), in that the confidence level term, $t_{\alpha,\upsilon}$, now refers to a Student-t distribution instead of a normal one and so depends on υ as well as α, and also includes the additional multiplier term $\sqrt{(\upsilon - 2)/\upsilon}$.

Since the t-distribution converges to the normal distribution as υ gets large, we can regard the t as a generalisation of the normal that produces higher than normal kurtosis when υ is finite. However, as υ gets large, $t_{\alpha,\upsilon}$ approaches its normal equivalent z_α, $\sqrt{(\upsilon - 2)/\upsilon}$ approaches 1, and the t-VaR, Equation (6.13), approaches the normal VaR, Equation (6.11a).

The t-VaR is closely related to the normal VaR, and has many of the same properties. In particular, it behaves in much the same way as normal VaR in the face of changes in α or h: it rises with α; unless $\mu_{P/L} = 0$, it tends to rise initially with h, and then peak and fall; and so on. Consequently, it produces curves and surfaces that are similar to the normal ones we have already seen.

The great advantage of the t over the normal is its ability to handle reasonable amounts of excess kurtosis. However, the t also has its problems. Like the normal, it fails to respect constraints on maximum possible losses, and can produce misleadingly high risk estimates as a result. When used at very high or very low confidence levels, it also has the drawback, like the normal, of not being consistent with extreme-value theory: we should therefore avoid using a t-distribution at extreme confidence levels. The t also suffers from an additional problem that does not affect the normal: the t-distribution is not stable (i.e., the sum of two or more t-distributed random variables is not necessarily distributed as a t variable itself).[5]

Example 6.3 (t-VaR)

Suppose P/L is distributed as a generalised t with mean 10, standard deviation 25, and $\upsilon = 5$ degrees of freedom. Then applying Equation (6.6) with $h = 1$, the VaR at the 99% confidence level is $-10 + \sqrt{(\upsilon - 2)/\upsilon} \times 25 \times t_{\alpha,\upsilon} = -10 + 0.775 \times 25 \times 3.365 = 55.197$. The corresponding normal estimate would be $-10 + 2.326 \times 25 = 48.150$, which is a little less.

Box 6.2 A Quantile Formula for a t-Distribution

The Student-t distribution is defined as a one-parameter distribution. If $t(\upsilon)$ is the variate of a Student-t with υ degrees of freedom, where υ is a positive integer, then $t(\upsilon)$ is distributed as $N(0, 1)/\sqrt{\chi^2_\upsilon/\upsilon}$, where $N(0, 1)$ is a standard normal distribution and χ^2_υ is a chi-squared distribution with υ degrees of freedom. This distribution will be symmetric around zero and have: a mean of zero, provided $\upsilon > 1$, which is necessary for the mean to exist; a variance of $\upsilon/(\upsilon - 2)$ provided $\upsilon > 2$, which is necessary for the variance to exist; a zero skew provided $\upsilon > 3$, which is necessary for the skew to exist; and a kurtosis of $3(\upsilon - 2)/(\upsilon - 4)$ provided $\upsilon > 4$, which is necessary for the kurtosis to exist.[6]

In risk measurement we usually prefer to deal with a generalised t-distribution that allows the values of the mean and standard deviation to be set by the user. If a and b are location and scale parameters, the generalised t variate, $t(a, b, \upsilon)$, is related to our original Student-t by the equation $t(a, b, \upsilon) = a + bt(\upsilon)$. Assuming they exist, this generalised t has mean a, variance $b^2\upsilon/(\upsilon - 2)$, skew 0, and kurtosis $3(\upsilon - 2)/(\upsilon - 4)$. If $t_{\alpha,\upsilon}$ is the quantile of the original Student-t for confidence level α and υ degrees of freedom, then the inverse function or quantile of the generalised t is $a + bt_{\alpha,\upsilon}$. We substitute out b for the standard deviation σ, and our quantile becomes $a + \sigma\sqrt{(\upsilon - 2)/\upsilon}t_{\alpha,\upsilon}$. This gives us the quantile for a t-distribution with mean a, standard deviation σ, and kurtosis $3(\upsilon - 2)/(\upsilon - 4)$.

To make use of it in practice, we would choose a and σ to match the observed mean and standard deviation, and choose υ to approximate the observed kurtosis. As υ gets large, the kurtosis approaches 3; and as υ falls to 5, which is the minimum number of degrees of freedom needed for the kurtosis to exist, then the kurtosis rises to 9. (Of course, for $\upsilon \leq 4$, we get even heavier tailed t-distributions, but their kurtosis does not exist.) We can approximately match υ to any empirical kurtosis in the range [3, 9] by setting υ as the integer closest to $(4k - 6)/(k - 3)$.

[5] Strictly speaking, the t is only stable in two special cases: when υ is 1 and when υ is infinite. These special cases correspond to the Cauchy and normal distributions. However, the t is not stable in the more general case considered in the text, where υ is greater than one but finite.

[6] See, e.g., Evans *et al.* (2000), pp. 179–180.

6.4 THE LOGNORMAL DISTRIBUTION

Another popular alternative is to assume that geometric returns are normally distributed. As explained in Chapter 3, this is tantamount to assuming that the value of our portfolio at the end of our holding period is lognormally distributed. Hence, this case is often referred to as lognormal. The lognormal VaR is given by the following formula:

$$VaR = P(1 - \text{Exp}[h\mu_R - \sqrt{h}\sigma_R z_\alpha]) \tag{6.14}$$

where P here is the current value of our portfolio, and μ_R and σ_R are the mean and standard deviation of geometric returns. Equation (14) generalises the earlier lognormal VaR equation (i.e., Equation (3.16) in Chapter 3) by allowing for an arbitrary holding period h. The lognormal assumption has the attraction of ruling out the possibility of a positive-value portfolio becoming a negative-value one, and this implies that the VaR can never exceed the value of our portfolio.[7]

A lognormal asset price is also implied by a geometric Brownian motion process or lognormal random walk, which is the archetypal process used in mathematical options valuation. A stock price S obeys such a process when it evolves randomly according to:

$$dS = S\mu dt + S\sigma dz \tag{6.15}$$

where dz is a Wiener process.[8] Applying Itô's lemma (see Box 6.3), then

$$d\log(S) = \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma dz \tag{6.16}$$

This implies, in turn, that

$$\begin{aligned} \log(S_t) - \log(S_0) &\sim N((\mu - \sigma^2/2)t, \sigma^2 t) \\ \Rightarrow \log(S_t) &\sim N(\log(S_0) + (\mu - \sigma^2/2)t, \sigma^2 t) \end{aligned} \tag{6.17}$$

Thus, a geometric Brownian motion implies that the asset price S_t is lognormal, conditional on its initial value S_0. If we wanted, we could also obtain a lognormal VaR using Equation (6.17) instead of Equation (6.14): we would use Equation (6.17) to infer the critical stock price S_t^* associated with a loss equal to the VaR, and then obtain the VaR from the difference between S_0 and S_t^*.[9]

One point to note about a lognormal distribution is its asymmetry, and an implication of this is that long and short positions have asymmetric risk exposures. A long position loses if the market goes down, and a short position loses if the market goes up, but with any symmetric distribution the VaR on a long position and the VaR on a short position are mirror images of each other, reflecting the symmetry of the lower and upper tails of the distribution. However, the situation can be very different with asymmetric distributions. With the lognormal, the worst the long position can do is lose the value of its investment – any risk measures are bounded above in a natural way – but a short position can make much larger losses.

---

[7] The lognormal VaR was illustrated earlier in Figure 3.7, and we have more to say on it in the appendix to this chapter, which deals with VaR over longer-term horizons.

[8] A Wiener process is the continuous-time limit, as the time interval δt goes to zero, of the process $\delta z = \varepsilon\sqrt{\delta t}$, where ε is a random drawing from a standard normal. Hence, a Wiener process is closely related to a normal; indeed, it is the Wiener process that implies that $\log(S)$ is normal.

[9] Assuming a GBM process with constant parameters, μ and σ are easily estimated from the sample mean m and sample standard deviation s measured from data taken at intervals of Δt. Rearranging the expressions given at the end of Box 6.3, $\hat{\sigma}$, the estimator of σ, is given by $\hat{\sigma} = s/\sqrt{\Delta t}$, and $\hat{\mu}$, the estimator of μ, is $\hat{\mu} = m/\Delta t + s^2/(2\Delta t)$.

Box 6.3 Geometric Brownian Motion, Geometric Returns and Lognormality

Suppose we have a time-dependent stochastic process:

$$dS = a(S,t)dt + b(S,t)dz$$

where dz is a Wiener process. This process has drift rate $a(S,t)$ and variance rate $b(S,t)$. If G is an arbitrary function of S, then Itô's lemma tells us that:[10]

$$dG = \left(a\frac{\partial G}{\partial S} + \frac{\partial G}{\partial t} + \frac{1}{2}b^2\frac{\partial^2 G}{\partial S^2}\right)dt + b\frac{\partial G}{\partial S}dz$$

This Itô process has the following drift and variance rates:

$$\textit{drift rate} = a\frac{\partial G}{\partial S} + \frac{\partial G}{\partial t} + \frac{1}{2}b^2\frac{\partial^2 G}{\partial S^2}$$

$$\textit{variance rate} = b^2\left(\frac{\partial G}{\partial S}\right)^2$$

If our S process is a geometric Brownian motion given by

$$dS = S\mu dt + S\sigma dz$$

then $a = \mu S$ and $b = \sigma$. We can now apply Itô's lemma to the log transformation $G = \log(S)$, in which case the Itô process is:

$$d\log(S) = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dz$$

This tells us that for a geometric Brownian motion process, the change in $\log(S)$ between time 0 and some period t is normally distributed with mean $(\mu - \sigma^2/2)t$ and variance $\sigma^2 t$, i.e.,

$$\log(S_t) - \log(S_0) \sim N\left[\left(\mu - \frac{\sigma}{2}\right)t, \sigma^2 t\right]$$

The change in $\log(S)$ is the geometric return, so a GBM process implies that geometric returns are normal. And this, in turn, implies that the stock price is lognormally distributed, conditional on S_0:

$$\log(S_t) \sim N\left[\log(S_0) + \left(\mu - \frac{\sigma}{2}\right)t, \sigma^2 t\right]$$

To show this, Figure 6.5 shows VaRs for long and short positions in a lognormally distributed asset. The long position loses when the market falls, but the loss is limited by the value of the initial investment, and the 95% VaR of a \$1 position – a long position – is 0.807. However, the corresponding short position – a short position worth −\$1 – has a VaR of 4.180. The short side has a potentially unlimited VaR, and its VaRs will be particularly high because it gets hit very hard by the long right-hand tail in the asset price, shown earlier in Figure 3.4 – a tail that translates into very high profits for the long position, and very high losses for the short position.

In short, the lognormal approach has the attraction of taking account of maximum-loss constraints on long positions. It also has the advantage of being consistent with a geometric

[10] Proofs of Itô's lemma can be found in all the good textbooks (e.g., Hull (2003), pp. 232–233).

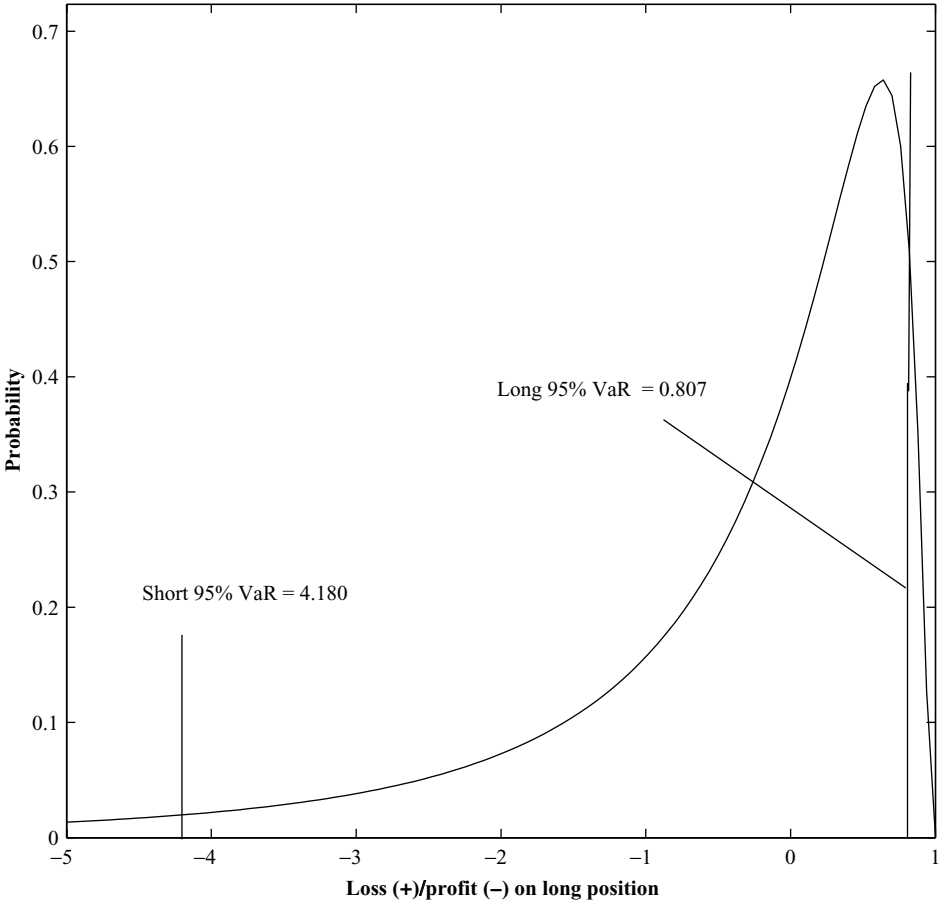

Figure 6.5 Lognormal VaRs for long and short positions
Note: Estimated with the 'lognormalvarfigure' function assuming the mean and standard deviation of geometric returns are 0 and 1, and with initial positions worth plus or minus $1.

Brownian motion process for the underlying asset price – a process that has a number of nice features and is widely used in the derivatives industry. Of course, if we use a lognormal approach, we would usually wish to use it in conjunction with a stochastic volatility (e.g., GARCH) model to accommodate the kurtosis and volatility clustering of most empirical return processes.[11]

Example 6.4 (Alternative estimates of lognormal VaR)

Suppose that geometric returns are normally distributed with mean 0.05 and standard deviation 0.25, and we have a portfolio currently worth 1. Using the lognormal VaR formula Equation (6.14), the 95% VaR over the 1-year holding period is $1 - \text{Exp}(0.05 - 0.25 \times 1.645) = 1 - \text{Exp}(-0.361) = 1 - 0.697 = 0.303$.

[11] A natural alternative is to replace the assumption that geometric returns are normally distributed with the assumption that they are distributed as a t-distribution. This 'log-t' approach combines the benefits of the lognormal with the heavier geometric-return tails of the t.

We can also estimate the same VaR using the log-price formula Equation (6.17). Given the same information, and bearing in mind that the mean return equals $\mu - \sigma^2/2$, where μ is the drift term in Equation (6.15), then $\sigma = 0.25$ and $\mu = 0.05 + 0.25^2/2 = 0.0813$. Plugging these into Equation (6.17) tells us that $\log(S_1)$ is distributed as normal with mean $0.0813 - 0.025^2/2 = 0.050$ and standard deviation 0.25. The critical value of the log price, $\log(S^*)$, is $0.050 - 0.25 \times 1.645 = -0.361$, so the critical asset price, S^*, is $\exp(-0.361) = 0.697$. The VaR is therefore $1 - 0.697 = 0.303$.

Example 6.5 (Normal vs lognormal VaRs on long and short positions)

Suppose that arithmetic returns are normal with mean 0.1 and standard deviation 0.25, and we have a portfolio currently worth 1. Using equation Equation (3.14) from Chapter 3, the 95% VaR on a long position is therefore $-0.10 + 0.25 \times 1.645 = 0.311$, and the corresponding VaR on a short position is exactly the same.

Now assume lognormality, and assume that geometric returns are normal with mean 0.1 and standard deviation 0.25. Applying Equation (6.14), the VaR on a long position is $1 - \text{Exp}(0.10 - 0.25 \times 1.645) = 0.268$, and the corresponding VaR on a short position is $-(1 - \text{Exp}(0.10 + 0.25 \times 1.645)) = 0.752$. With an asymmetric distribution such as the lognormal, the VaR on a short position is no longer a mirror image of the long-position VaR.

Box 6.4 Estimating Confidence Intervals for Parametric Risk Measures

There are various ways we can estimate confidence intervals for parametric VaR. Leaving aside special cases where we can derive confidence intervals analytically,[12] we can obtain confidence intervals using any of the following approaches:

- We can apply textbook confidence-interval formulas using estimates of the quantile standard errors (see Chapter 3, section 3.5.1). These are easy to apply, but can be inaccurate and are not recommended for VaR at high confidence levels.
- We can obtain VaR confidence intervals using order statistics theory (see Chapter 3, Appendix 2).
- We can obtain confidence intervals by bootstrapping from the assumed parametric distribution. (The bootstrap is discussed further in Chapter 4, and in Chapter 4, Appendix 2.)
- We can obtain confidence intervals using Monte Carlo simulation. One way to do so is to specify the distributions from which relevant parameters (e.g., μ, σ, etc.), are drawn. We then run m simulation trials, each involving the estimation of a complete parameter set, and use each parameter set to obtain an estimated VaR. We then obtain our confidence interval from the simulated distribution of m VaR estimates. We have more to say on simulation methods in Chapter 8.

[12] For examples of such special cases, see, e.g., Huschens (1997) or Chappell and Dowd (1999).

- We can obtain confidence intervals using profile likelihood approaches: we use estimates of the likelihood function of the assumed distribution function to determine the range of VaR values consistent with a given data set at a chosen probability level.[13]

Some of these methods can also be adapted to estimate confidence intervals for ES or other coherent risk measures. For example, the OS approach is also useful for estimating ES confidence intervals, and bootstrap and Monte Carlo methods can always be adapted to obtain confidence intervals for any quantile-based risk measure, including ES and other coherent risk measures.

6.5 MISCELLANEOUS PARAMETRIC APPROACHES

We now consider some of the other parametric approaches sometimes suggested for VaR and/or ES estimation.

6.5.1 Lévy Approaches

One promising avenue is to use Lévy processes, sometimes also known as α-stable or stable Paretian processes. These processes were first suggested as plausible representations of financial return processes by Mandelbrot and Fama in the 1960s as a means of accommodating the fat tails of financial return data. Lévy processes also have a certain theoretical plausibility because they arise from a generalised version of the central limit theorem: if we have a large number of independent identically distributed random variables, their sum is a process belonging to the Lévy family. Lévy processes generally lack conventional 'closed-form' solutions for their density functions, but can be represented instead by their characteristic function:

$$\ln\phi(x) = \begin{cases} i\mu x - \gamma|x|^{\alpha}[1 - i\beta(x/|x|)\tan(\pi\alpha/2)] & \text{if} \quad \alpha \neq 1 \\ i\mu x - \gamma|x|[1 + i\beta(x/|x|)(2/\pi)\log|x|] & \text{if} \quad \alpha = 1 \end{cases} \tag{6.18}$$

This function has four parameters: a stability index, α, lying in the range $0 < \alpha \leq 2$; a skewness parameter, β, lying in the range $-1 \leq \beta \leq 1$, and taking the value 0 if the distribution is symmetric; a location parameter, μ, that is a real number; and a positive scale parameter, γ.[14]

The Lévy process has some interesting special cases. If the stability index, α, is 2, the Lévy process is equivalent to a normal (or Gaussian) process, which indicates that the Lévy process is a generalisation of the normal. However, if $\alpha < 2$, the Lévy process has heavier than normal tails and an infinite variance, and if $\alpha = 1$ and the Lévy process is symmetric ($\beta = 0$), then the Lévy is equivalent to a Lorentzian or Cauchy distribution. These parameters can be estimated using ML methods.

We discussed these distributions earlier in Chapter 2, Appendix 1. Lévy processes have the attractions of stability, self-similarity and domains of attraction. However, except for the normal special case (i.e., provided $\alpha < 2$), all Lévy processes have probability density functions that converge to power-law tails and so produce heavy tails and infinite variance. These heavy tails are illustrated in Figure 6.6, which also shows that the tails of Lévy distributions can decline at a considerably slower rate than the tails of normal distributions. This suggests that non-normal

[13] See, e.g., McNeil (1998).
[14] See, e.g., Mantegna and Stanley (2000), p. 25.

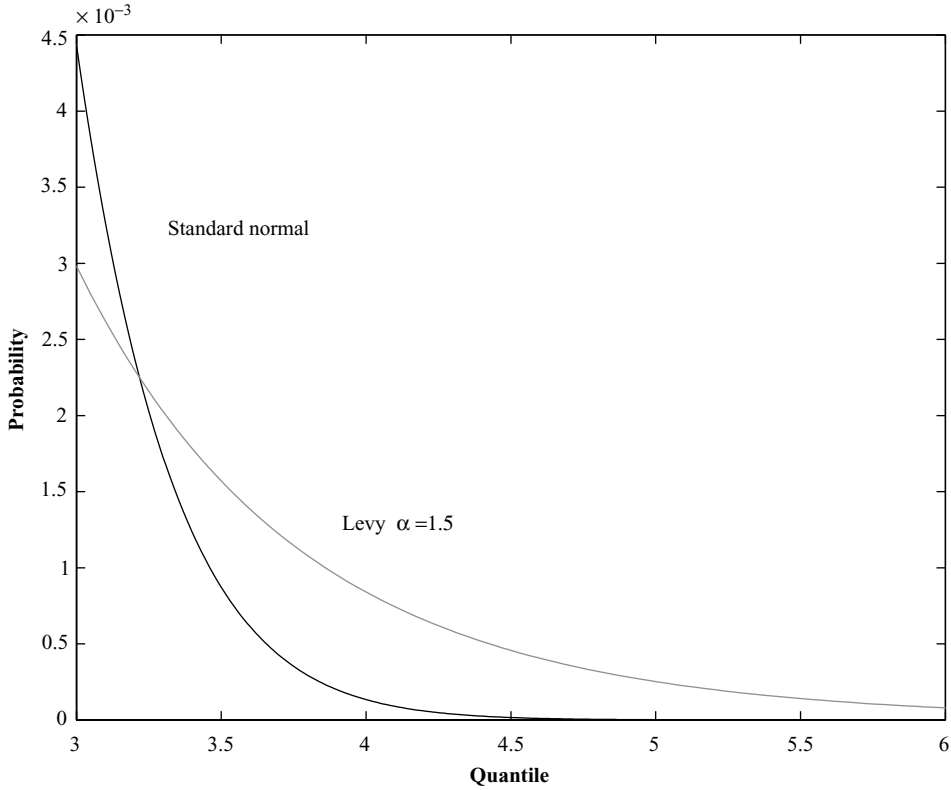

Figure 6.6 Lévy and normal tails

Lévy processes might be suitable for modelling heavy tails, *provided* we are comfortable with the implication that the variance is infinite.[15]

Mittnik, Paolella and Rachev have provided considerable evidence to support the applicability of Lévy processes to financial return data.[16] Their work suggests that Lévy processes can also be regarded as (partial) alternatives to GARCH processes for modelling financial returns, because GARCH models also lead to heavy-tailed return distributions. However, Lévy processes also have their drawbacks. Perhaps the most obvious is that the plausibility of non-normal Lévy processes is undermined by widespread evidence that the variances of financial returns are finite (although this evidence has been challenged[17]). There is also some evidence that financial return processes are not always scale invariant, and that the fat tails produced by

[15] If we are unhappy about the infinite variance, we can also resort to a related process known as truncated Lévy flight (TLF), which is a Lévy process with truncated extremes. A TLF has some Lévy-like features, and behaves like a Lévy within a certain range; however, it differs from a 'pure' Lévy insofar as it has a finite variance (which is presumably 'good') but lacks stability (which is presumably 'bad'). Since its variance is finite, a TLF will eventually converge to a normal process. Hence, a TLF is a stochastic process that behaves like a Lévy for relatively 'short' periods, but eventually converges to a normal in the long run. For more on this process, see Mantegna and Stanley, p. 64).

[16] See, e.g., Mittnik *et al*. (1998, 2000).

[17] Mittnik *et al*. (2000), p. 391.

non-normal Lévy processes might be too heavy for the data.[18] In addition, there is the practical problem that Lévy processes are not particularly easy to model.[19]

6.5.2 Elliptical and Hyperbolic Approaches

A promising alternative is to use elliptical and hyperbolic distributions, as suggested in recent work by Eberlein, Bauer, and others.[20] The elliptical distributions usually used in financial risk management are symmetric (i.e., zero skew) distributions with a more general than normal kurtosis. They have the attraction of providing a straightforward – and very normal-like – VaR formula. If our P/L distribution is elliptical with location and scale parameters equal to μ and δ^2, say, then the VaR is:

$$VaR = -\mu + \delta z_\alpha \tag{6.19}$$

where z_α is the percentile of the standard form (i.e., location zero, scale 1) elliptical distribution (e.g., in the case of the normal, z_α will be the percentile of the standard normal distribution). Elliptical distributions also provide formulas for the ES (see note 3 above) and have the advantage that we can easily work with them at the individual position level, as well as at the aggregate portfolio level.

To estimate the elliptical VaR, we need to estimate these parameters and plug their values into Equation (6.19). We can estimate μ by taking it to be the mean, which we can estimate by simple (e.g., LS) methods, and we can estimate the other parameters using a maximum-likelihood procedure. Once we have the algorithms in place, they involve (relatively) little extra work, and the evidence in these papers suggests that they can provide significant improvements over normal approaches.

A further generalisation is the family of generalised hyperbolic distributions. The generalised hyperbolic adds an additional parameter to the hyperbolic and therefore provides a more flexible fit. These distributions include the hyperbolic, elliptical, normal inverse Gaussian and of course normal distributions as special cases. However, they have forbidding density functions and, special cases aside, do not yield closed-form solutions for VaR.

6.5.3 Normal Mixture Approaches

Another alternative is to model returns using a mixture-of-normals approach.[21] This process assumes that returns are 'usually' drawn from one particular normal process, but are occasionally drawn from another normal process with a higher variance. A typical normal mixture process is:

$$\textit{normal mixture process} = (1 - \delta_t)\varepsilon_{1,t} + \delta_t \varepsilon_{2,t} \tag{6.20}$$

where $\varepsilon_{1,t}$ is distributed as a normal with mean μ and variance σ_1^2, $\varepsilon_{2,t}$ is distributed as normal with mean μ and variance σ_2^2, and $\sigma_2^2 > \sigma_1^2$. The variable δ_t is a binary random variable that

[18] Cont *et al.* (1997), p. 5.

[19] The parameter α can be difficult to estimate, and the Lévy process itself should be estimated with fast Fourier transform methods. However, these obstacles are comparatively minor, and we might expect to see Lévy methods become more widely used as software becomes more user-friendly.

[20] See, e.g., Eberlein *et al.* (1998), Eberlein (1999), Bauer (2000), or Breckling *et al.* (2000).

[21] See Zangari (1996b) and Venkataraman (1997).

Table 6.1 Moments of a normal mixture distribution

Mean	μ
Variance	$[p\sigma_{1,t}^2 + (1-p)\sigma_{2,t}^2]$
Skewness	0
Kurtosis	$3[p\sigma_{1,t}^4 + (1-p)\sigma_{2,t}^4]/[p\sigma_{1,t}^2 + (1-p)\sigma_{2,t}^2]^2$

Source: Lagnado *et al.* (2000), p. 4.

takes the value 0 most of the time, but occasionally takes the value 1, which it does with some small probability p. Most of the time, the normal mixture variate is therefore $\varepsilon_{1,t}$, but on a small proportion of occasions, it takes the value of 1 and the mixture process is more volatile. Very large or very low values are then more frequent – and our tails are heavier – than under an unadjusted normal distribution.

The moments of the mixture model are shown in Table 6.1, and indicate that the mixture that has a zero skew as well as excess kurtosis. We can match the mixture variance and kurtosis to the empirical variance and kurtosis by selecting appropriate values for p, σ_1 and σ_2.

Normal mixture approaches have a number of merits: they are conceptually (fairly) simple; they can accommodate any reasonable degree of kurtosis; they make use of the standard linear normal estimates of variances and correlations, and so retain much of the tractability of normal approaches; and (at least for portfolios with a small number of different assets) they require relatively few additional parameters. However, if we are to implement a normal mixture approach we need to obtain estimates of the parameters involved, and this is trickier than it looks. The most obvious approach to parameter estimation is maximum likelihood, but unfortunately, the normal mixture likelihood function does not have a global maximum, and this makes a standard ML approach unusable. Zangari and Venkataraman suggest alternative solutions to this problem, the former involving a Gibbs sampling tool, and the latter involving quasi-Bayesian maximum likelihood, and the details of these solutions can be found in their respective articles. When applying a normal mixture approach at the position level in a multi-position portfolio, we also have to address the (non-trivial) issue of whether/how the individual risk factors, including the δ_t binary terms, are correlated with each other.

6.5.4 Jump Diffusion

A related way of obtaining fat tails is to assume again that the distribution consists of ordinary and extraordinary components, but in this case the extraordinary component is a 'jump' variable reflecting a large market move. This process is known as a jump-diffusion process, and was first suggested by Merton (1976).[22] The jump-diffusion process decomposes the change in the log price, say, into the following two components:

$$\Delta \log p_t = e_t + e_t^J \tag{6.21}$$

where e_t is an 'ordinary' normal random variable with mean 0 and variance σ_t^2, and e_t^J is the sum of N_t random variables, J_n, where $n = 1, \ldots, N_t$, and each J_n is normally distributed with mean μ_J and variance σ_J^2. N_t is a Poisson random variable with mean λ, where λ is the mean number of jumps per day.

[22] For more on the jump-diffusion process and its applications to VaR, see, e.g., Duffie and Pan (1997, pp. 12–13), Zangari (1997) or Gibson (2001).

Table 6.2 Moments of a jump-diffusion distribution

Mean	$\lambda\mu_J$
Variance	$\sigma^2 + \lambda(\mu_J^2 + \sigma_J^2)$
Skewness	$\lambda\mu_J(\mu_J^2 + 3\sigma_J^2)$
Kurtosis	$3 + \lambda(\mu_J^4 + 6\mu_J^2\sigma_J^2 + 3\sigma_J^4)$

Source: Zangari (1997, pp. 25–26).

The moments of the jump model are shown in Table 6.2, and indicate that we can use a jump process to obtain skewness as well as excess kurtosis in our distribution. The sign of the average jump size, λ, determines the sign of the skew, and the size of the jump volatility parameter, σ_J, determines the excess kurtosis.

The jump-diffusion process is a useful way of obtaining heavy-tailed (and skewed, if desired) distributions over short-run horizons,[23] but models involving jump diffusion are more difficult to estimate, and do not necessarily outperform more basic models that omit jump diffusion.

Perhaps the easiest way to estimate risk measures with a jump-diffusion model is to use Monte Carlo simulation: we simulate a large number of paths of a jump-diffusion process to the end of the holding period, and estimate our risks measures in the usual way from the histogram of simulated terminal prices.

6.5.5 Stochastic Volatility Approaches

A natural alternative is to model volatility as a persistent stochastic random process (e.g., so volatility is random, but if it is high today, it is likely to be high tomorrow, and so forth). Typically, volatility today is taken to be a random function of one or more past volatilities, and there are many ways of specifying this function. For example, we might model volatility σ_t by a GARCH process such as a GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta\sigma_{t-1}^2; \quad \omega \geq 0, \alpha, \beta \geq 0, \alpha + \beta < 1 \tag{6.22}$$

where x_{t-1} is the most recent observation of a filtered return process. A high value of β means that volatility is 'persistent' and takes a long time to change; and a high value of α means volatility is 'spiky' and quick to react to market movements. We could also model return volatility by any other GARCH family process. Alternatively, we might model volatility by a regime-switching Markov chain: volatility might take a finite number of possible values, and its random transition over time can be modelled by a transition matrix that gives the probabilities of it switching from any one state to any other.

Whether we use a GARCH process or a Markov chain, we would implement the stochastic volatility model using Monte Carlo simulation: we would simulate many paths for the random evolution of volatility out to the end of the holding period, obtain the asset prices paths associated with each volatility path, and obtain the estimated risk measure from the terminal asset prices associated with each asset price path.[24]

[23] However, we should note that as the horizon period rises, the jump-diffusion process converges to normality. There is therefore no point using a jump-diffusion process to estimate risk measures over a long holding period.

[24] For more on the use of SV models for VaR estimation, see, e.g., Duffie and Pan (1997, pp. 14–17), Billio and Pelizzon (1997), and Eberlein *et al.* (2001).

Box 6.5 Other Parametric Approaches to Estimating Risk Measures

Box–Cox

There are also other parametric alternatives. One of these is a Box–Cox transformation, which is a simple and widely used method of dealing with non-normal data. The Box–Cox approach could be quite useful in situations where we suspect non-normality but are uncertain about the 'true' P/L or return distribution. If we have a set of observations X, the Box–Cox transformation

$$X^{(\lambda)} = \begin{cases} (X^{\lambda} - 1)/\lambda & \text{if} \quad \lambda \neq 0 \\ \ln(X) & \quad\quad\; \lambda = 0 \end{cases}$$

defines a transformed variable $X^{(\lambda)}$ that is approximately normal.[25] λ can be found by maximising a normal likelihood function.

To estimate VaR, we apply the transformation to our data and estimate normal VaR on the transformed data. We then obtain our estimate of Box–Cox VaR by applying a reverse Box–Cox transformation on our normal VaR estimate. We can estimate our other risk measures in the usual way by taking them as suitably weighted VaR averages.

The generalised error distribution

Another approach is to use a generalised error distribution (GED, sometimes also known as an exponential power distribution). The GED is a family of distributions which takes a variety of particular forms depending on the value of a parameter υ, and which includes the normal as a special case. The GED can therefore be regarded as a more flexible generalisation of the normal distribution. This approach takes standardised returns, r, say, to have a pdf given by

$$f(r) = \frac{\upsilon \exp\left[-\left(\frac{1}{2}\right)\left|\frac{r}{\lambda}\right|^{\upsilon}\right]}{\lambda 2^{(1+1/\upsilon)}\Gamma(1/\upsilon)}$$

where $\Gamma(1/\upsilon)$is the standard gamma function and $\lambda = [2^{-(2/\upsilon)}\Gamma(1/\upsilon)/\,\Gamma(3/\upsilon)]^{1/2}$. This process becomes normal when $\upsilon = 2$, but has heavier than normal tails when $\upsilon < 2$. We can therefore select the value of υ to accommodate our desired kurtosis. His results suggest that the GED approach is a considerable improvement over the normal one.[26]

Johnson family,[27] *Pearson family,*[28] *skew-t*[29] *and generalised lambda*[30]

We can also fit empirical moments to Johnson family, Pearson family, skew-*t* or generalised lambda distributions. These distributions are very flexible in that they can accommodate more or less arbitrary values for their first four moments. Hence, if we have estimates

[25] The original Box–Cox transformation is limited to positive data, but the transformation is easily extended to include zero and negative observations by using the alternative transformation $(|X|^{\lambda}\text{sign}(Y) - 1)/\lambda$ suggested by Bicksel and Docksum (1981).

[26] For more on fitting a GED distribution, see Zangari (1996d) and Lin and Chien (2002).

[27] See, e.g., Johnson (1949) and Zangari (1996c).

[28] A good example of how a Pearson family distribution might be fitted to the data is provided by Aston and Tippett (2004).

[29] For more on the skew-*t*, see, e.g., Jones and Faddy (2003) and Giot and Laurent (2003).

[30] See, e.g., Lakhany and Mausser (2000).

of the first four moments of an empirical distribution, we can fit these moments to one of these distributions, and then infer estimates of risk measures from the latter. However, the flexibility of these distributions comes at the cost of tractability, and none of them is particularly easy to use.[31]

6.5.6 The Cornish–Fisher Approximation

A very different approach is to use a Cornish–Fisher approximation. This is based on the Cornish–Fisher expansion, which is used to determine the percentiles of distributions that are near-normal. The actual expansion provides an adjustment factor that can be used to adjust estimated percentiles for non-normality, and the adjustment is reliable provided departures from normality are 'small'.

Suppose that z_α is a standard normal variate for a confidence level α (i.e., so $z_\alpha = 1.645$, etc.). Then the Cornish–Fisher expansion is

$$z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\rho_3 + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\rho_4 - \frac{1}{36}\left(2z_\alpha^3 - 5z_\alpha\right)\rho_3^2 + \textit{higher-order terms} \tag{6.23}$$

where ρ_3 is the distribution's skewness coefficient and ρ_4 is its kurtosis.[32] If we treat the higher-order terms as negligible – which is tantamount to assuming that departures from normality are 'small' – the expansion becomes

$$z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\rho_3 + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\rho_4 - \frac{1}{36}\left(2z_\alpha^3 - 5z_\alpha\right)\rho_3^2 \tag{6.24}$$

To use the expansion, we simply replace z_α by Equation (6.24) as our estimated percentile. This is equivalent to adjusting the normal variate z_α for non-normal skewness and/or kurtosis.

When using the Cornish–Fisher approximation, we should keep in mind that it will only provide a 'good' approximation if our distribution is 'close' to being normal, and we cannot expect it to be much use if we have a distribution that is too non-normal. To illustrate the latter point, suppose we have P/L data drawn from a zero-mean, unit-standard deviation t-distribution with 5 degrees of freedom. This distribution has a kurtosis of 9, which is clearly well away from the normal kurtosis of 3. The true (i.e., t) VaR, the normal VaR, and the Cornish–Fisher VaR are shown in Figure 6.7, and we can clearly see that the Cornish–Fisher approximation is generally quite poor. Except around a confidence level of 95%, the Cornish–Fisher approximation generally produces a VaR estimate that is even worse than the (poor) estimate we would get if we assumed normality. We should therefore use the Cornish–Fisher approximation with some care, because it can worsen our estimates as well as improve them.

Example 6.6 (Cornish–Fisher approximation)

Suppose our P/L has mean 0, standard deviation 1, skewness 0, and kurtosis 9. We wish to estimate the 95% VaR using a Cornish–Fisher expansion. Using Equation (6.25), the adjusted percentile is $1.645 + (1/24)(1.645^3 - 3 \times 1.645) \times 9 = 1.464$. The Cornish–Fisher adjusted VaR is 1.464.

---

[31] Nor is this treatment of parametric alternatives by any means exhaustive. There are also some promising approaches based on very different foundations, such as fuzzy logic (as in Cherubini and della Lunga (2001)) and gauge theory (e.g., Illinski, 2001), which is a recent import from physics.

[32] See Lee and Lin (1992), p. 9.

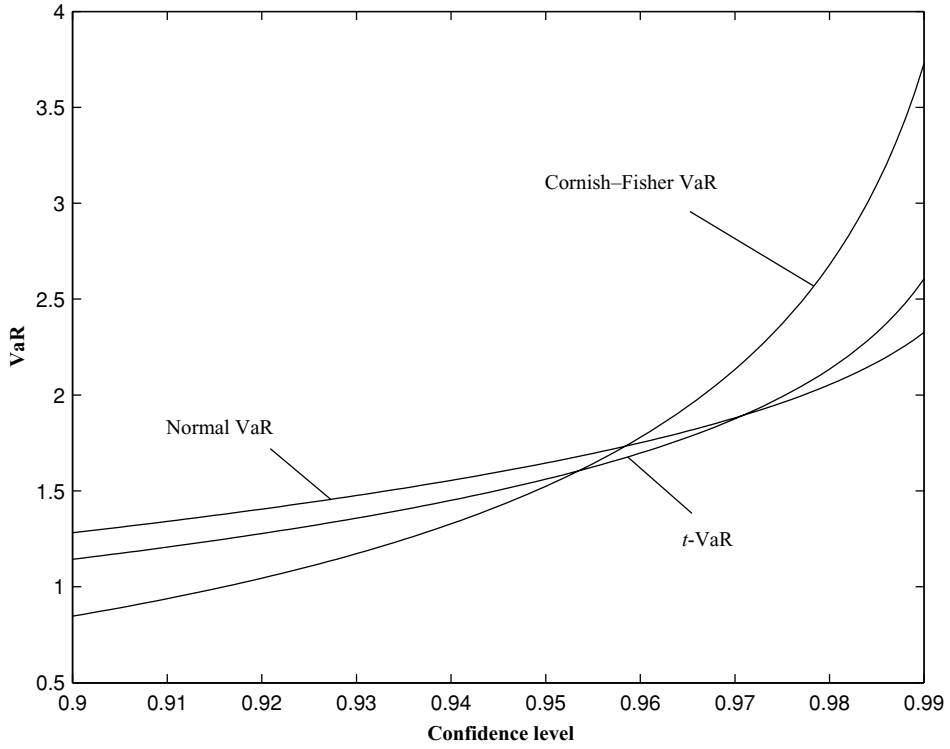

Figure 6.7 Cornish–Fisher VaR with moderately high kurtosis
Note: Estimated with the mean and standard deviation of P/L equal to zero and 1, and with 5 degrees of freedom for the *t*-distribution.

Note, however, that we can also fit a t-distribution with $\upsilon = 5$ to these skewness and kurtosis parameters, and the 'true' t-VaR in this case is $\sqrt{\upsilon - 2)/\upsilon}t_{\alpha,\upsilon} = \sqrt{(5-2)/5} \times 2.015 = 1.561$.

Thus, an initial normal VaR estimate (of 1.645) is adjusted using the Cornish–Fisher approach to 1.464, while the 'true' VaR is 1.561 – in this case at least, the Cornish–Fisher overadjusts so much that it is useless.

Box 6.6 Quantile Regression Approaches to VaR

A number of recent studies have suggested a very different approach to VaR estimation: instead of modelling a distribution and inferring the VaR from the relevant quantile, these papers model the quantile itself using quantile-regression methods.[33] A good example is Engle and Manganelli's Conditional Autoregressive Value at Risk (CAViaR) model, which specifies how the quantile evolves over time in response to news such as new P/L

[33] Quantile regression was first proposed by Koenker and Bassett (1978), and quantile regression methods have been applied to VaR by Engle and Manganelli (1999), Taylor (2000) and Chernozhukov and Umantsev (2001).

observations. A simple instance is the first-order CAViaR model:

$$VaR_t = \beta_0 + \beta_1 VaR_{t-1} + \beta_2[I(x_{t-1} > VaR_{t-1}) - p]$$

where x_t is the loss (or profit, if negative) realised in t, and $I(.)$ is the indicator function that takes a value 1 if $x_{t-1} > VaR_{t-1}$ and a value 0 otherwise. The autoregressive term $\beta_1 VaR_{t-1}$ makes the VaR_t change smoothly over time and the second term indicates how VaR_t responds to the 'news' embedded in the latest P/L figure. This term also gives a much larger impact to a loss in excess of VaR than to a more 'normal' observation beneath the VaR threshold. Estimation of the parameters of this model is not straightforward because the regression quantile objective function is not differentiable and has many local optima, so Engle and Manganelli use a genetic algorithm to do the numerical optimisation. Their results suggest that versions of the CAViaR model perform well, and are good at adapting to changes in the risk environment.

A related alternative is suggested by Granger and Sin (2002). They model the critical return quantile r^* in terms of an autoregression of absolute returns:

$$r* = \alpha + \sum_{i-1}^{p} \beta_i |r_{t-i}|$$

and the VaR is easily obtained from this using $VaR = P(1 + r^*)$. The VaR is thus, in effect, a simple autoregression of past absolute returns. This approach is fairly easy to model, and their results indicate that it performs well in practice.

6.6 THE MULTIVARIATE NORMAL VARIANCE–COVARIANCE APPROACH

We now consider parametric VaR and ES from the individual position level, where we make parametric assumptions about individual asset returns rather than assumptions about portfolio returns. In many respects the obvious assumption to make – or at least the assumption we would like to be able to make – is that individual asset returns are distributed as multivariate normal. This assumption is the counterpart to the earlier assumption that portfolio returns are normal – and, indeed, the assumption that individual asset returns are multivariate normal implies that portfolio returns will be normal.

Suppose we have a portfolio consisting of n different assets, the (arithmetic) returns to which are distributed as multivariate normal with mean $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}$ is an $n \times 1$ vector, and $\boldsymbol{\Sigma}$ is an $n \times n$ matrix with variance terms down its main diagonal and covariances elsewhere. The $1 \times n$ vector $\mathbf{w}$ gives the proportion of our portfolio invested in each asset (i.e., the first element w_1 gives the proportion of the portfolio invested in asset 1, and so on, and the sum of the w_i terms is 1). Our portfolio return therefore has an expected value of $\mathbf{w}\boldsymbol{\mu}$, a variance of $\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}$, where $\mathbf{w}^{\mathrm{T}}$ is the $n \times 1$ transpose vector of $\mathbf{w}$, and a standard deviation of $\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}$. If the current value of our portfolio is P, then its VaR over holding period h and confidence level cl is:

$$VaR(h, \alpha) = [-h\mathbf{w}\boldsymbol{\mu} + \sqrt{h}\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}} z_\alpha]P \tag{6.25}$$

Equation (6.25) is the multivariate normal equivalent of our earlier normal VaR, Equation (6.11a), and the z_α term again refers to the relevant quantile from a standard normal distribution.

The 'P' term arises because we are applying multivariate normality to returns, rather than P/L. It follows, too, that our ES is:

$$ES(h, \alpha) = [-h\mathbf{w}\boldsymbol{\mu} + \sqrt{h}\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}\frac{\phi(z_\alpha)}{1-\alpha}]P \tag{6.26}$$

which is the multivariate normal equivalent of the normal ES, Equation (6.11b). Note that in Equation (6.26), as in Equation (6.11b), the $\phi(z_\alpha)$ term is from the univariate standard normal distribution.

The covariance matrix $\boldsymbol{\Sigma}$ captures the interactions between the returns to the different assets. It is also closely related to the $n \times n$ correlation matrix. If we define $\boldsymbol{\sigma}$ as an $n \times n$ matrix consisting of standard deviations along its main diagonal and zeros elsewhere, then $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\mathbf{C}\boldsymbol{\sigma}$, where $\mathbf{C}$ is the $n \times n$ correlation matrix. This matrix has 1s down its main diagonal – these can be interpreted as each return's correlation with itself – and the correlations between different asset returns elsewhere.

The values of these covariance or correlation terms have a critical bearing on the VaR. To see these effects in their simplest form, suppose our portfolio consists of two assets, 1 and 2, with a relative amount $w_1 = 0.5$ held in asset one, and relative amount $w_2 = 0.5$ held in asset 2. If each asset has a return with mean 0 and standard deviation 1, the standard deviation of the portfolio return σ_p is:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2} = \sqrt{\frac{1+\rho}{2}} \tag{6.27}$$

where ρ is the correlation between the returns to the two assets. If we now take the value of our portfolio to be 1, and we set the holding period to be 1 as well, then our VaR is:

$$VaR = \sigma_p z_\alpha = \sqrt{\frac{1+\rho}{2}} z_\alpha \tag{6.28}$$

which shows very clearly the dependence of the VaR on the value of the correlation coefficient.

The correlation coefficient can vary over the range $[-1, +1]$, and there are three important special cases:

- If ρ takes its minimum value of -1, the portfolio standard deviation is zero, and so, too, is the VaR. The explanation is that the shocks to the two returns perfectly offset each other, the portfolio return is certain, and so the standard deviation of the portfolio return and the VaR are both zero.
- If ρ is 0, the portfolio standard deviation is $\sqrt{1/2} = 0.707$, and the VaR is 1.163. In this case the returns are independent of each other, and the portfolio VaR is less than the VaR we would get (i.e., 1.645) if we invested our portfolio entirely into one asset or the other. This result reflects the well-known idea that if returns are independent, a portfolio of two assets will be less risky than a portfolio consisting of either asset on its own.
- If ρ reaches its maximum, 1, the portfolio standard deviation is 1 and the VaR is 1.645, which is the same as the VaR we would have obtained investing entirely in either asset. If asset returns are perfectly correlated, the portfolio variance is the sum of the variances of the individual returns, and there is no risk diversification.

More generally, the main points are that the portfolio VaR falls as the correlation coefficient falls and, except for the special case where the correlation coefficient is 1, the VaR of the portfolio is less than the sum of the VaRs of the individual assets.

These insights extend naturally to portfolios with more than two assets. If we have n assets in our portfolio and invest equally in each, and if we assume for convenience that all correlations take the same value ρ and that all asset returns have the same standard deviation σ, then the standard deviation of the portfolio return is:

$$\sigma_p = \sigma\sqrt{\frac{1}{n} + \frac{(n-1)\rho}{n}} \tag{6.29}$$

Except for the special case where $\rho = 1$, in which case σ_p is always σ, the portfolio standard deviation falls as n gets bigger. The behaviour of the portfolio standard deviation as n changes is illustrated in Figure 6.8. The portfolio standard deviation falls rapidly at first – which indicates substantial reductions in portfolio risk as n initially rises – but then levels out and thereafter falls much more slowly.

This figure also has other interpretations. The curves in the figure are the portfolio standard deviations taking account of diversification effects, divided by the portfolio standard deviation if we take no account of diversification effects (i.e., and invest entirely in one asset). The curves can therefore be regarded as measuring the extent to which diversification reduces

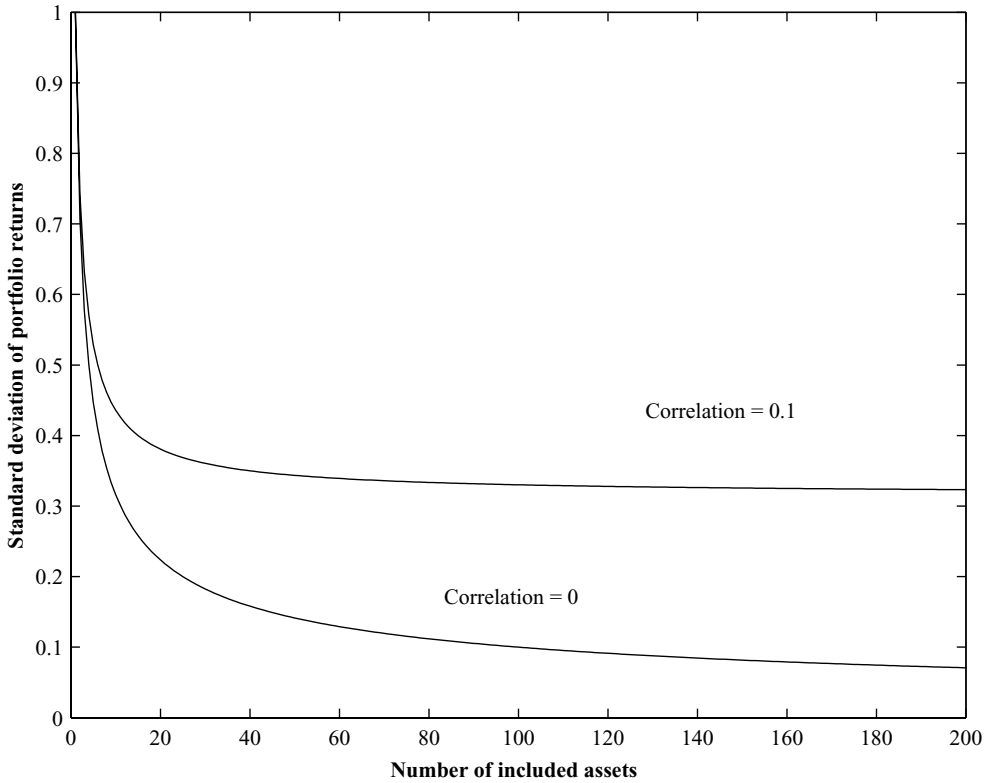

Figure 6.8 Portfolio standard deviation and the number of included assets
Note: Estimated assuming that individual assets have the same standard deviation σ, equal to 1, and all correlations take the same value.

overall risk, dependent on the number of included assets and the correlation. The curves tell us that diversification reduces risk more, the more assets are included and the lower the correlation ρ.

Furthermore, given the parameters assumed in Figure 6.8, these same curves also give us the ratio of diversified to undiversified VaR – they give the VaR taking account of diversification, divided by the VaR taking no account of diversification (i.e., the VaR we would get if we invested only in one asset). Again, we see that the extent to which VaR is reduced by diversification depends on the number of assets and the correlation. However, we can also see that even a small amount of diversification can have big effects on the VaR, and these effects are particularly strong if the correlation coefficient is relatively low.

Example 6.7 (Multivariate normal variance–covariance VaR and ES)

Suppose we have two assets in our portfolio, which is currently worth 1 and divided equally between the two assets. Suppose also that the returns on these assets are distributed as multivariate normal with mean 0, standard deviation 1, and correlation 0.5. Then applying Equation (6.25), our 95% VaR is

$$\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}z_{\alpha} = \sqrt{[0.5, 0.5]\begin{bmatrix}1, 0.5\\ 0.5, 1\end{bmatrix}\begin{bmatrix}0.5\\ 0.5\end{bmatrix}} \times 1.645 = 1.425$$

The corresponding ES is

$$\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}\frac{\phi(z_{\alpha})}{1-\alpha} = \sqrt{[0.5, 0.5]\begin{bmatrix}1, 0.5\\ 0.5, 1\end{bmatrix}\begin{bmatrix}0.5\\ 0.5\end{bmatrix}}\frac{\phi(1.645)}{0.05} = 1.786$$

As an aside, note that if we had all our portfolio invested in the first asset only, say, our VaR would have been

$$\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}z_{\alpha} = \sqrt{[1, 0]\begin{bmatrix}1, 0.5\\ 0.5, 1\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}} \times 1.645 = 1.645$$

and our ES would have been

$$\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathrm{T}}}\frac{\phi(z_{\alpha})}{1-\alpha} = \sqrt{[1, 0]\begin{bmatrix}1, 0.5\\ 0.5, 1\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}}\frac{\phi(1.645)}{0.05} = 2.063$$

These are the same as the univariate normal VaR and ES – which is as it should be, since the portfolio now depends only on a single asset return.

6.7 NON-NORMAL VARIANCE–COVARIANCE APPROACHES

6.7.1 Multivariate *t*-Distributions

If our returns are not multivariate normal, it may still be possible to apply a variance–covariance approach to related distributions. An obvious choice is a multivariate t, which yields the

following VaR and ES formulas:[34]

$$VaR(h,\alpha) = [-h\mathbf{w}\boldsymbol{\mu} + \sqrt{h}\sqrt{\frac{\upsilon-2}{\upsilon}}\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathbf{T}}}t_{\upsilon,\alpha}]P \tag{6.30a}$$

$$ES(h,\alpha) = [-h\mathbf{w}\boldsymbol{\mu} + \sqrt{h}\sqrt{\frac{\upsilon-2}{\upsilon}}\sqrt{\mathbf{w}\boldsymbol{\Sigma}\mathbf{w}^{\mathbf{T}}}\frac{f(t_{\upsilon,\alpha})}{1-\alpha}]P \tag{6.30b}$$

The multivariate t retains the advantages of the multivariate normal (e.g., it uses the same parameters, it is very tractable, etc.), but (as with the univariate t) also allows for moderate degrees of excess kurtosis. The $f(t_{\upsilon,\alpha})$ term is the pdf from the univariate Student t-distribution with υ degrees of freedom.

6.7.2 Multivariate Elliptical Distributions

More generally, we can also apply a variance–covariance approach provided returns are multivariate elliptically distributed.[35] These yield VaR and ES formulas similar to those we get under multivariate normality, but with (typically) different sets of location and scale parameters. The problem in applying these approaches is to estimate the parameters. ML parameter estimators often lack closed-form solutions, and numerical methods are difficult to apply in high-dimension problems. However, we can often break the estimation problem down into two more manageable steps. In the first, we estimate the mean and covariance matrix using standard methods. The estimated mean gives us our location parameter, and the estimated covariance matrix can be combined with ML algorithms to estimate the other parameters.[36]

6.7.3 The Hull–White Transformation-into-normality Approach

A different approach is suggested by Hull and White (1998). If returns are not multivariate normal, they suggest that we can still apply a variance–covariance approach if we transform our returns to make them multivariate normal. We then apply the variance–covariance analysis to our transformed returns, and unravel the results to derive our 'true' VaR estimates. We therefore replace the assumption that our returns are multivariate normal with the much weaker assumption that they can be transformed into multivariate normality.[37] This allows us to retain the benefits of the variance–covariance approach – its tractability, etc. – while allowing us to estimate VaR and ES in a way that takes account of skewness, kurtosis, and the higher moments of the multivariate return distribution.

Assume there are m different instruments in our portfolio. Let e_{it} be the returns on asset i in period t, and let G_{it} be an assumed distribution function for e_{it}. This function will, in general, be time dependent reflecting factors such as GARCH volatility and correlation processes. We

[34] The estimation of multivariate-t VaR is discussed at length in Glasserman *et al.* (2000a). For more on ES under the multivariate-t, see Landsman and Valdez (2003).

[35] In principle, it can also be applied to multivariate hyperbolic or generalised hyperbolic distributions. We could estimate their parameters and then infer the VaR from the estimated distribution functions (as in HS), but the practical problem is getting suitable numerical algorithms. All we currently know is that the approach in the text would also work for a multivariate normal inverse Gaussian (or NIG) distribution, but other forms of multivariate hyperbolic or generalised hyperbolic distribution can run into major numerical problems. See Eberlein and Prause (2000).

[36] Alternative ML methods are suggested by Bauer (2000, p. 460) and Eberlein and Prause (2000, p. 14).

[37] The assumption that returns can be transformed into multivariate normality is very reasonable. Malevergne and Sornette (2001, pp. 4–6) show that if returns come from some arbitrary multivariate distribution, then we can map them to multivariate normality under very weak conditions, and the transformed variables will retain their dependence structure.

now transform e_{it} into a new variable f_{it} using the transformation:

$$f_{it} = N^{-1}[G_{it}(e_{it})] \tag{6.31}$$

where N is the standard normal distribution function. The term in square brackets, $G(e_{it})$, is the zth percentile of the assumed distribution function G_{it}, and f_{it} is the same percentile of the standard normal distribution. Hence, Equation (6.31) transforms the return e_{it} into its standard normal equivalent, f_{it}. We can also invert Equation (6.31) to map the f_{it} back to the original returns, e_{it}, using the reverse mapping:

$$e_{it} = G_{ij}^{-1}[N(f_{it})] \tag{6.32}$$

Equation (6.31) thus allows us to transform our returns and Equation (6.32) allows us to untransform them. The function G_{it} can take any form we like: we can take it to be some particular heavy-tailed distribution, for instance, or we can take it to be the empirical distribution function drawn from our original data.

Next, we assume that our transformed returns – the f_{it} – are distributed as multivariate normal, and we estimate their mean vector and covariance matrix. Hull and White now suggest that we use a Monte Carlo method to simulate values of the transformed returns f_{it}, based on the estimated mean and variance–covariance parameters. We then use Equation (6.32) to map our simulated f_{it} values to their untransformed counterparts, the e_{it}, to give a set of simulated non-normal returns, and we can estimate the desired risk measure using a standard method (e.g., a non-parametric method based on the simulated series).

This approach is easy to implement and can be applied to a much wider set of non-multivariate-normal return distributions than the elliptical approaches just considered. It therefore gives us a way of handling departures from multivariate normality beyond the relatively limited range allowed by multivariate elliptical distributions, and the results reported by Hull and White also suggest that it works quite well in practice.

6.8 HANDLING MULTIVARIATE RETURN DISTRIBUTIONS WITH COPULAS

6.8.1 Motivation

The previous position-level approaches all seek to fit our multivariate distribution into a form where it could be tackled within a variance–covariance framework – either by working with elliptical distributions that are compatible with such a framework or, as in Hull–White, by transforming our data to fit a variance–covariance framework. However, from a statistical point of view, the most natural way to handle multivariate distributions is to use copulas (see the appendix to Chapter 5). A copula enables us to extract the dependence structure from the joint distribution function, and so separate out the dependence structure from the marginal distribution functions. More specifically, a copula enables us to construct a multivariate distribution function from the marginal distribution functions, but in a way that allows for very general dependence structure. Thus, unlike variance–covariance approaches, which are only valid in circumstances where we can use correlation as a dependence measure, copula approaches are *universally valid*, and can be used in circumstances where correlation-based approaches cannot. From a statistical point of view, copulas offer *the* universally correct way to estimate risk measures from the position level.

Implicit in this universality is another advantage, which is extremely important for risk integration and enterprise-wide risk management (ERM). As one attempts to estimate risk measures across diverse positions, it becomes increasingly difficult to sustain any assumption that diverse positions have similar risks. For example, suppose we have built models to estimate the risk measures of, say, two very diverse types of market risk. One risk model might be based on a normal distribution and the other on, say, a heavy-tailed Lévy. We then find ourselves wondering how to estimate an integrated risk measure across both positions. However, it then comes as a shock to realise that none of the earlier multivariate methods is applicable to this problem. This is because these multivariate methods presuppose a *single* multivariate density function, whereas what we actually have are two *quite different* univariate density functions. Fortunately, copulas can solve this problem with ease. This example is of course a very basic one, but it is important because we are increasingly likely to encounter the same problem as we attempt to integrate risks across increasingly diverse positions or business units across the firm. Even if we manage to fit the same distribution to all our market risk positions (which will often not be the case), it will almost certainly be the case that we cannot fit the same distribution across both market and credit portfolios, or across market, credit and operational risk portfolios. Thus, any attempt to integrate risks across increasingly diverse positions within the firm will almost certainly run into this problem, *unless* we resort to copulas. The bottom line is the perhaps surprising and certainly provocative conclusion that copulas are essential to successful ERM.

6.8.2 Estimating VaR with Copulas

There are various ways we can estimate VaR with copulas, but one fairly natural approach is the following, based on first principles.[38] Suppose we have a portfolio with positions in two assets whose random losses are X and Y, and we wish to estimate the portfolio VaR using a copula. The portfolio loss is $X + Y$. If $X + Y > VaR$, we get a 'bad' outcome and a loss exceeding VaR, and this occurs with probability $1 - \alpha$. Hence

$$\begin{aligned} 1 - \alpha = \Pr\{X + Y > VaR\} &\Rightarrow \alpha = 1 - \Pr\{X + Y > VaR\} \\ &\Rightarrow \alpha = \Pr\{X + Y \leq VaR\} \end{aligned} \tag{6.33}$$

This tells us the confidence level α *given* the VaR, and the confidence level is equal to the probability that the sum of X and Y is less than the VaR.

However, the copula gives us the multivariate cumulative density function $\Pr\{X \leq x \text{ and } Y \leq y\}$, which can also be written as $F(x, y)$. We therefore need to infer $\Pr\{X + Y \leq x + y = VaR\}$, say, from the multivariate cdf $F(x, y)$.

Noting now that $F(x, y) = C(\phi(x), \phi(y))$, where $C(\phi(x), \phi(y))$ is the copula function predicated on the marginals $\phi(x)$ and $\phi(y)$, we can show that

$$\Pr\{X + Y \leq VaR\} = \int_{-\infty}^{\infty} \partial_1 C(\phi(z), \phi(VaR - z))\phi'(z)dz \tag{6.34}$$

[38] For more on this method and several further examples, see Dowd and Fackler (2004). I am indebted to Peter Oliver and Paul Fackler for their inputs into the development of this method.

where ∂_1 refers to the partial derivative with respect to the first argument in the copula function following it. Now transform z using $w = \phi(z)$. It then follows that $dw = \phi'(z)dz$ and Equation (6.34) simplifies to

$$\Pr\{X + Y \leq VaR\} = \int_0^1 \partial_1 C(\phi(w), \phi(VaR - \phi^{-1}(w)))dw \tag{6.35}$$

If we can evaluate the partial differential analytically, then we have a straightforward solution for $\Pr\{X + Y \leq VaR\}$, but we cannot always do so. However, we can always obtain an approximation for $\Pr\{X + Y \leq VaR\}$ if we use a discrete-sum approximation for the right-hand side of Equation (6.35):

$$\begin{aligned}\Pr\{X + Y \leq VaR\} &= \int_0^1 \partial_1 C(\phi(w), \phi(VaR - \phi^{-1}(w)))dw \\ &\approx \sum_{w\in(0,1)} [C(\phi(w + \Delta w), \phi(VaR - \phi^{-1}(w))) - C(\phi(w), \phi(VaR - \phi^{-1}(w)))]\Delta w \end{aligned} \tag{6.36}$$

where w is taken in increments of Δw.

This suggests the following procedure to obtain a numerical estimate of the confidence level associated with any given VaR:

- We choose a copula function.
- We write a software function to estimate the value of the copula, which is straightforward.
- We use the copula software function to write a function to approximate the right-hand side of Equation (6.36), and so obtain an estimate of $\Pr\{X + Y \leq VaR\}$. This requires that we choose a specific increment size Δw. The smaller the increment, the more accurate the results, but the greater the number of calculations involved.

The answer provided by this last step is the confidence level associated with a given VaR.

However, in practice, we usually want to estimate the VaR *given* the confidence level, so how do we obtain the one from the other? One answer is to embed our confidence level function in a suitable root finding algorithm, such as a bisection algorithm. To show how we might do so, suppose that X and Y are each standard normal *and* that their dependence structure can be described by a product copula (see Equation (A5.3a) in the appendix to Chapter 5). We then start by assuming that the VaR lies within some very wide range, and (if the VaR does lie within this range) the root-find algorithm will gradually work its way towards the true VaR.

The algorithm and the calculations involved are given in Table 6.3. We assume that we wish to estimate the 95% VaR, and we initially presume that the VaR lies somewhere in the region [0,3]. This gives us a midpoint value of 1.5, and we estimate the confidence level of this midpoint to be 0.8558. Since this is *below* the target confidence level ($\alpha = 0.95$), this tells us that the VaR must lie in the *upper* half of the initial region, i.e., in the region [1.5,3]. We then take the next midpoint, 2.25 and estimate its confidence level, which turns out to 0.9431. Since this is also below 0.95, this tells us that the VaR lies in the upper half of our second region (i.e., it lies in [2.25,3]). We proceed to take the next midpoint (2.625) and estimate its confidence level value (0.9666). As this lies above 0.95, the VaR must lie in the lower half of the remaining region (i.e., in [2.25,2.625]). We then carry on in the same manner again and again, and with each iteration we half the length of the region within which the VaR is known to be. As the region narrows, the confidence levels converge towards α and the midpoints converge to the

Table 6.3 A bisection-copula algorithm for the estimation of VaR

Iteration No.	Range of uncertainty	Midpoint of range	Midpoint Confidence level	Comment and response
1	[0,3]	1.5	0.8558	Below $\alpha \Rightarrow$ in upper half of range
2	[1.5,3]	2.25	0.9431	Below $\alpha \Rightarrow$ in upper half of range
3	[2.25,3]	2.625	0.9666	Above $\alpha \Rightarrow$ in lower half of range
4	[2.25,2.625]	2.4375	0.9562	Above $\alpha \Rightarrow$ in lower half of range
5	[2.25,2.4375]	2.3438	0.9516	Above $\alpha \Rightarrow$ in lower half of range
6	[2.25,2.3438]	2.2969	0.9483	Below $\alpha \Rightarrow$ in upper half of range
7	[2.2969,2.3438]	2.3203	0.9499	Close to target $\alpha \Rightarrow VaR \approx 2.32$

VaR, and a point will eventually come where we get a confidence level that is sufficiently close to α that we can take the midpoint as our estimated VaR. In this particular case, the process converges in about seven steps to an answer (about 2.32) that is accurate to a couple of decimal points. If we wish to, we can achieve greater accuracy by carrying out more iterations or by suitable refinements to the numerical approximations of Equation (6.35).

The correctness of this estimate is easily demonstrated. A product copula with Gaussian (or normal) marginals is the same as a multivariate standard normal distribution with a zero correlation. In such a case, the portfolio P/L is distributed as normal with zero mean and variance 2, and applying the usual textbook VaR formula (i.e., $VaR = \sigma z_{0.95} = \sqrt{2} \times 1.645$, where σ is the portfolio standard deviation) gives us a value of 2.326.

The nice thing about this approach is that we can plug any copula we like into it, and the algorithm will give us the VaR for the assumed copula. This means that we can now estimate VaRs in multivariate contexts without being restricted to correlation-based measures of dependence.

There are many other ways we can apply copulas to estimate risk measures: we can combine copulas with Monte Carlo simulation engines, we can sometimes derive implicit formulas for VaR using particular types of copula, and so on. An example of this latter type of approach is suggested by Malevergne and Sornette (2001). They start off by showing that we can map any arbitrary multivariate return distribution to a multivariate normal distribution by making use of a Gaussian copula. (*En passant*, in so doing, they also provide the theoretical underpinning for Hull and White's 'transformation to multivariate normality' approach. This also tells us that the HW approach is, in effect, a disguised copula approach.) They then apply this result to a class of heavy-tailed distributions: the so-called stretched exponential or modified Weibull distributions. These postulate that returns have the following marginal distribution:

$$p(x) = \frac{1}{2\sqrt{\pi}} \frac{c}{\chi^{c/2}} |x|^{\frac{c}{2}-1} \exp\left[-\left(\frac{x}{2}\right)^2\right] \tag{6.37}$$

with a shape or tail parameter c and a scale parameter χ. If $c = 2$, the pdf is normal; and if $c < 1$, it is a stretched exponential or subexponential. If all assets have the same shape parameter c, we can derive implied solutions for asymptotic VaR, dependent on the value of c. The simplest case is where $c < 1$ and our returns are independent over time, in which case our VaR is given implicitly by:

$$1 - \alpha \sim \Gamma\left(\frac{1}{c}, VaR\right) \tag{6.38}$$

where Γ is a gamma function and α as usual is the VaR confidence level. This equation can be solved numerically on a spreadsheet. The other cases (i.e., where returns are temporally dependent, $c > 1$, etc.) lead to their own implied VaR equations, but these tend to be more difficult to handle. The Malevergne–Sornette approach thus yields powerful results, but only works if we are prepared to make certain assumptions and – in common with many copula applications – requires a high level of statistical competence from the user.

Example 6.8 (Simple Malevergne–Sornette VaR)

If we have temporally independent returns and all assets have the same shape parameter $c = 0.9$, then the asymptotic 95% VaR is given implicitly by $0.05 = \Gamma(1/0.9, \mathit{VaR})$. Using an elementary search (e.g., a bisection method) the asymptotic VaR turns out to be about 3.550.

6.9 CONCLUSIONS

This chapter has covered a lot of ground. Parametric approaches are very diverse, and can be applied at both the portfolio level and the position level. Their main attraction is that they give us a lot of information about risk measures on the basis of relatively limited assumptions. Moreover, because they use the additional information embodied in the parametric assumptions on which they are based, parametric approaches will generally give better estimates of risk measures than their non-parametric counterparts, *provided* those parametric assumptions are correct. Their main weakness is the flipside of this last point: parametric approaches depend on those assumptions and can lead to serious errors if those assumptions are false. In using parametric approaches, it is therefore important to get the parametric assumptions right. To do so, we need to consider a number of points:

- We should think carefully over the range of permissible values that returns or P/L can take (e.g., do we want losses or returns to be bounded?).
- We should be attentive to possible skewness and kurtosis in our data, and this information should be critical in our choice of distribution.
- We should be conscious of possible time dependence (e.g., volatility clustering) and/or conditionality.
- If we are just interested in a risk measure at a particular confidence level, or over a relatively narrow range of confidence levels, then there is no point trying to fit a curve to the whole distribution and we should focus on choosing a distribution that fits the confidence level or range of confidence levels we are interested in, i.e., we should seek to fit a distribution locally rather than universally.
- We should try to get a feel for model and parameter risk. If we think we have heavy tails, for example, we should try out various heavy-tailed alternatives, and try to get a sense for how our results might vary with alternative specifications, including alternative calibrations of key parameters.

In short, the key to success in parametric risk estimation is to make reasonable assumptions, and take account of our uncertainty as well as what (we think) we know.

If we are using parametric approaches at the position level, we should also pay particular attention to the dependence structure in our data. We should be wary of assuming multivariate

normality unless we believe that our data are fairly close to multivariate normality, and we should not rely too much on the central limit theorem to justify treating the data 'as if' they were multivariate normal. If we wish to get away from the restrictive confines of multivariate normality – as we generally should – we can sometimes use alternatives such as multivariate elliptical approaches or the Hull–White 'transformation-to-normality' approach. However, if we want a framework that is generally valid (i.e., valid for any reasonable assumptions about dependence structure), then we have no choice but to grasp the statistical nettle and use copula approaches. Copula approaches are extremely flexible as well as very general, and also allow us to integrate diverse types of risk within the same risk estimation engine. So if we are interested in serious risk aggregation, we have to use copulas.

Appendix
Forecasting Longer-term Risk Measures

Most of the VaR literature is concerned with forecasting VaR over very short horizons, typically horizons of 1 or maybe 10 trading days. This reflects the fact that VaR evolved in a securities market context where horizons are short. However, there are many cases where we might be interested in forecasting financial risks over longer – possibly much longer – horizons. For example, banks are often concerned with horizons of quarters or years, and insurance companies and pension funds can be concerned with horizons that might span decades. So how do we handle longer-term risk forecasting problems?

One answer sometimes offered in the VaR literature is the square-root VaR rule. If $VaR(h)$ is the VaR over a horizon of h days, and $VaR(1)$ is the VaR over 1 day, this rule tells us that we can obtain the former from the latter by multiplying it by the square root of h:

$$VaR(h) = \sqrt{h}.VaR(1) \tag{A6.1}$$

The square-root VaR rule is enshrined prominently in the 1996 Market Risk Amendment to the Basel Accord, and was reiterated in the recent revision (2004). However, it is suggested there only as a means of extrapolating from a 1-day to a 10-day VaR, and not as a means of obtaining a VaR over a genuinely long horizon. This is also just as well, because the square-root rule can be very unreliable over longer horizons.

A6.1 PARAMETRIC LONG-TERM VaR

A better answer is to start by making some parametric assumption about the distribution of returns over the long horizon we are concerned about.[39] In making such an assumption, we also have to take account of compounding, so it makes sense to work with assumptions about geometric rather than arithmetic returns. An obvious choice is therefore to assume lognormality – that asset prices are lognormal or (equivalently) that geometric returns are normal. This leads to the following formula for the h-period (e.g., h-year) VaR:

$$VaR(h) = P[1 - \exp(\mu_h h - \sqrt{h}\sigma_h z_\alpha)] \tag{A6.2}$$

where P is the current value of our portfolio and μ and σ are the mean and volatility of anticipated h-year returns expressed in annualised terms. Assuming that we have the parameter values μ_h and σ_h, we can now estimate VaR over any horizon.

A typical example is shown in Figure A6.1, based on assumed annualised parameter values of $\mu_h = 0.075$ and $\sigma_h = 0.25$ and an initial portfolio value of \$1. This illustrates how VaR changes with both the horizon h and the confidence level. For any given confidence level, as the time horizon increases, the VaR rises initially but then peaks and turns down; after that

[39] A non-parametric alternative would also be possible, but this would require a long data set and we would have to be comfortable with the underlying assumption that the future would be similar enough to the past to justify a non-parametric approach applied to a long forecast horizon. Alternatively, we could use a semi-parametric approach and 'correct' historical data to reflect our best estimates of future parameters such as means and volatilities.

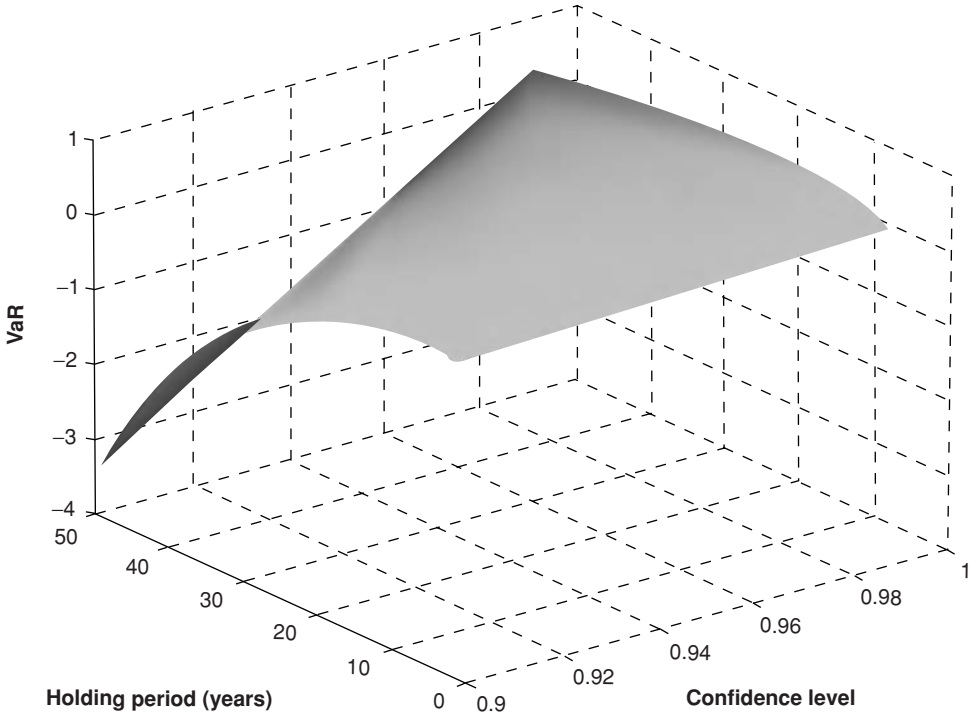

Figure A6.1 Longer-term VaR
Note: Based on assumed parameter values of $\mu_h = 0.075$ and $\sigma_h = 0.25$, and an assumed initial investment of \$1.

it keeps falling, becomes negative at some point, and thereafter remains negative and moves further and further away from zero. The behaviour of the VaR also depends on the confidence level: for relatively low confidence levels, the VaR peaks quickly and then rapidly falls; but for relatively high confidence levels, the VaR peaks slowly and stays at or near its maximum value – which is bounded above by, and sometimes close to, the value of the investment itself – for a long time. Note, too, that while the VaR has this natural upper bound, it has no corresponding lower bound, and will fall indefinitely as the horizon continues to rise.

The explanation is fairly obvious. In the earlier stages, the VaR will generally be dominated by the volatility parameter, because this parameter will (usually) be large relative to the mean parameter, and this will cause the VaR to rise, and often to its maximum possible value. However, in the longer term, the VaR is dominated by the mean parameter, because this parameter increases with h and the volatility parameter only increases with the square-root of h. The long-term dominance of the mean will eventually bring the VaR down, and make it continue to fall thereafter. The long-term VaR is dominated by the compounding of mean returns.

It is then immediately obvious why the square-root VaR rule is very unreliable in long-run contexts: the square-root rule has the VaR continuing to rise indefinitely, whereas the true VaR can only ever rise to a maximum given by the value of the portfolio, and (given a positive mean) will eventually turn down and then fall indefinitely.

A6.2 OBTAINING PARAMETER FORECASTS

However, this analysis presupposes that we have the parameter forecasts needed, the μ_h and σ_h, and these parameters depend on the horizon h.

Obtaining the former parameter is fairly straightforward: we can resort to judgemental methods over the horizon concerned (i.e., we can 'take a view') or we can obtain forecasts of expected returns using statistical analysis and then extrapolate these over the horizon, and extrapolation would be straightforward because the mean compounds in a well-behaved way.

However, obtaining a volatility forecast can be more difficult. One approach to this latter problem is to think in terms of temporally aggregating short-term volatility. And, once again, an obvious but controversial way of doing so is to apply a square-root extrapolation rule to a short-term volatility. So if $\sigma_{1,t}$ is our forecast at time t of volatility over the next 1-day period, then our forecast of volatility over the next h days is $\sqrt{h}.\sigma_{1,t}$. This rule is known to be valid if geometric returns are independently and identically distributed (iid), but its validity in non-iid situations is questionable. For example, suppose we have a GARCH(1,1) process

$$\sigma_{1,t}^2 = \omega + \alpha x_{1,t-1}^2 + \beta\sigma_{1,t-1}^2; \quad \omega \geq 0, \alpha, \beta \geq 0, \alpha + \beta < 1 \tag{A6.3}$$

that gives 1-day ahead forecasts. Drost and Nijman (1993) show that the corresponding h-day process is:

$$\sigma_{h,t}^2 = \omega_h + \alpha_h x_{h,t-1}^2 + \beta_h \sigma_{h,t-1}^2 \tag{A6.4}$$

where $\omega_h = h\omega[1 - (\alpha + \beta)^h][1 - (\alpha + \beta)]^{-1}$, $\alpha_h = (\alpha + \beta)^h - \beta_h$ and $|\beta_h| < 1$ is the solution to the quadratic equation:

$$\frac{\beta_h}{1 + \beta_h^2} = \frac{a(\alpha + \beta)^h - b}{a[1 + (\alpha + \beta)^{2h}] - 2b} \tag{A6.5}$$

where

$$\begin{aligned} a &= h(1 - \beta)^2 + 2h(h - 1)\frac{(1 - \alpha - \beta)^2(1 - \beta^2 - 2\alpha\beta)}{(\kappa - 1)[1 - (\alpha + \beta)^2]} \\ &\quad + 4\frac{[h - 1 - h(\alpha + \beta) + (\alpha + \beta)^h][\alpha - \alpha\beta(\alpha + \beta)]}{1 - (\alpha + \beta)^2} \\ b &= [\alpha - \alpha\beta(\alpha + \beta)]\frac{1 - (\alpha + \beta)^{2h}}{1 - (\alpha + \beta)^2} \end{aligned}$$

where κ is the kurtosis of x_t. It should be fairly obvious that this (conditional) volatility forecast can be rather different from the (unconditional) volatility forecast we would get from a square-root extrapolation rule.

This suggests that we should not blindly apply the square-root rule to volatility forecasting, unless returns are iid. It also suggests that if we have to forecast volatility over longer horizons, then we should use a correct volatility-forecasting rule (i.e., one that correctly extrapolates the h-period volatility from parameters estimated over a 1-day period) or we should use an h-period model. Unfortunately, both routes are problematic: volatility-forecasting rules can be difficult to derive, and often do not exist; and in working with h-period data frequencies, we lose a lot of data points and therefore have relatively few observations to work with.

Yet there are also counterarguments. The available empirical evidence seems to suggest that the square-root rule does in fact work quite well for the horizons of up to a month that can be

reasonably considered given the length of data available.[40] Furthermore, and more helpfully, simulation results suggest that the square-root volatility rule works for vanilla and other *symmetric* GARCH processes over arbitrarily long horizons *provided* that we scale up from the long-run average (i.e., the square root of $\omega/(1-\alpha-\beta)$ in the case of the GARCH(1,1)) rather than the current-period daily volatility.[41] This suggests that the square-root volatility rule may not be so bad after all, provided it is used with care. Furthermore, where the square-root rule cannot reliably be used (i.e., in the case of asymmetric GARCH processes), the most natural solution is to use Monte Carlo simulation: to specify the GARCH process over some given horizon, and use that process to simulate the h-period volatilities.

Moreover, over very long-term horizons all these approaches become unnecessary. Over a very long horizon a lot of the randomness cancels out, and the best approach is simply to take a view about the long-run average value of the volatility over that horizon, and then carry out some robustness analysis to gauge the sensitivity (or otherwise) of our results to the assumed parameter value(s). However, estimates of long-term risk measures, like those of their short-term counterparts, are likely to be subject to considerable model and parameter risk.

[40] For example, see Andersen *et al.* (2001), p. 53.
[41] See Dowd (2003).

7
Parametric Approaches (II): Extreme Value

There are many problems in risk management that deal with extreme events – events that are unlikely to occur, but can be very costly when they do. These events are often referred to as low-probability, high-impact events, and they include large market falls, the failures of major institutions, the outbreak of financial crises and natural catastrophes. Given the importance of such events, the estimation of extreme risk measures is a key concern for risk managers.

However, to estimate such risks we have to confront a difficult problem: extreme events are rare by definition, so we have relatively few extreme observations on which to base our estimates. Estimates of extreme risks must therefore be very uncertain, and this uncertainty is especially pronounced if we are interested in extreme risks not only *within* the range of observed data, but *well beyond* it – as might be the case if we were interested in the risks associated with events more extreme than any in our historical data set (e.g., an unprecedented stock market fall).

Practitioners can only respond by relying on assumptions to make up for lack of data. Unfortunately, the assumptions they make are often questionable. Typically, a distribution is selected arbitrarily, and then fitted to the whole data set. However, this means that the fitted distribution will tend to accommodate the more central observations, because there are so many of them, rather than the extreme observations, which are much sparser. Hence, this type of approach is often good if we are interested in the central part of the distribution, but is ill-suited to handling extremes.

When dealing with extremes, we need an approach that comes to terms with the basic problem posed by extreme-value estimation: that the estimation of the risks associated with low-frequency events with limited data is inevitably problematic, and that these difficulties increase as the events concerned become rarer. Such problems are not unique to risk management, but also occur in other disciplines as well. The standard example is hydrology, where engineers have long struggled with the question of how high dikes, sea walls and similar barriers should be to contain the probabilities of floods within reasonable limits. They have had to do so with even less data than financial risk practitioners usually have, and their quantile estimates – the flood water levels they were contending with – were also typically well out of the range of their sample data. So they have had to grapple with comparable problems to those faced by insurers and risk managers, but have had to do so with even less data and potentially much more at stake.

The result of their efforts is extreme-value theory (EVT) – a branch of applied statistics that is tailor-made to these problems.[1] EVT focuses on the distinctiveness of extreme values and makes as much use as possible of what theory has to offer. Not surprisingly, EVT is quite

[1] The literature on EVT is vast. However, some standard book references on EVT and its finance applications are Embrechts *et al.* (1997), Reiss and Thomas (1997) and Beirlant *et al.* (2004). There is also a plethora of good articles on the subject, e.g., Bassi *et al.* (1998), Longin (1996, 1999), Danielsson and de Vries (1997a,b), McNeil (1998), McNeil and Saladin (1997), Cotter (2001, 2004), and many others.

different from the more familiar 'central tendency' statistics that most of us have grown up with. The underlying reason for this is that central tendency statistics are governed by central limit theorems, but central limit theorems do not apply to extremes. Instead, extremes are governed, appropriately enough, by extreme-value theorems. EVT uses these theorems to tell us what distributions we should (and should not!) fit to our extremes data, and also guides us on how we should estimate the parameters involved. These EV distributions are quite different from the more familiar distributions of central tendency statistics. Their parameters are also different, and the estimation of these parameters is more difficult.

This chapter provides an overview of EV theory, and of how it can be used to estimate measures of financial risk. As with earlier chapters, we will focus mainly on the VaR (and to a lesser extent, the ES) to keep the discussion brief, but the approaches considered here extend naturally to the estimation of other coherent risk measures as well.

The chapter itself is divided into four sections. Sections 7.1 and 7.2 discuss the two main branches of univariate EV theory, section 7.3 discusses some extensions to, including multivariate EVT, and section 7.4 concludes.

7.1 GENERALISED EXTREME-VALUE THEORY

7.1.1 Theory

Suppose we have a random loss variable X, and we assume to begin with that X is independent and identically distributed (iid) from some unknown distribution $F(x) = \text{Prob}(X \leq x)$. We wish to estimate the extreme risks (e.g., extreme VaR) associated with the distribution of X. Clearly, this poses a problem because we don't know what $F(x)$ actually is.

This is where EVT comes to our rescue. Consider a sample of size n drawn from $F(x)$, and let the maximum of this sample be M_n.[2] If n is large, we can regard M_n as an extreme value. Under relatively general conditions, the celebrated Fisher–Tippett theorem (1928) then tells us that as n gets large, the distribution of extremes (i.e., M_n) converges to the following generalised extreme-value (GEV) distribution:

$$H_{\xi,\mu,\sigma} = \begin{cases} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right] & \xi \neq 0 \\ \text{if} & \\ \exp\left[-\exp\left(\dfrac{x-\mu}{\sigma}\right)\right] & \xi = 0 \end{cases} \tag{7.1}$$

where x satisfies the condition $1 + \xi(x - \mu)/\sigma > 0$.[3] This distribution has three parameters. The first two are μ, the location parameter of the limiting distribution, which is a measure of the central tendency of M_n, and σ, the scale parameter of the limiting distribution, which is a measure of the dispersion of M_n. These are related to, but distinct from, the more familiar mean and standard deviation, and we will return to these presently. The third parameter, ξ, the tail index, gives an indication of the shape (or heaviness) of the tail of the limiting distribution.

[2] The same theory also works for extremes that are the minima rather than the maxima of a (large) sample: to apply the theory to minima extremes, we simply apply the maxima extremes results but multiply our data by -1.

[3] See, e.g., Embrechts *et al.* (1997), p. 316.

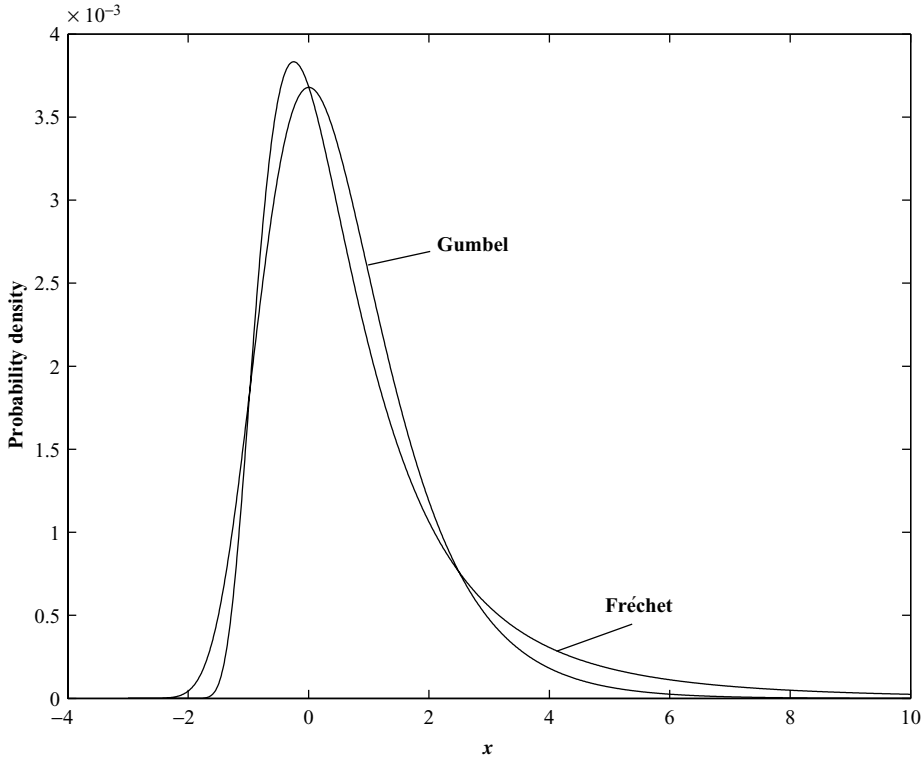

Figure 7.1 Standardised Gumbel and Fréchet probability density functions

The GEV Equation (7.1) has three special cases:

- If $\xi > 0$, the GEV becomes the Fréchet distribution. This case applies where the tail of $F(x)$ obeys a power function and is therefore heavy (e.g., as would be the case if $F(x)$ were a Lévy distribution, a t-distribution, a Pareto distribution, etc.). This case is particularly useful for financial returns because they are typically heavy-tailed, and we often find that estimates of ξ for financial return data are positive but less than 0.35.
- If $\xi = 0$, the GEV becomes the Gumbel distribution, corresponding to the case where $F(x)$ has exponential tails. These are relatively light tails such as those we would get with normal or lognormal distributions.
- If $\xi < 0$, the GEV becomes the Weibull distribution, corresponding to the case where $F(x)$ has lighter than normal tails. However, the Weibull distribution is not particularly useful for modelling financial returns, because few empirical financial returns series are so light-tailed.[4]

The standardised (i.e., $\mu = 0$, $\sigma = 1$) Fréchet and Gumbel probability density functions are illustrated in Figure 7.1. Both are skewed to the right, but the Fréchet is more skewed than the Gumbel and has a noticeably longer right-hand tail. This means that the Fréchet has considerably higher probabilities of producing very large X-values.

[4] We can also explain these three cases in terms of domains of attraction. Extremes drawn from Lévy or t-distributions fall in the domain of attraction of the Fréchet distribution, and so obey a Fréchet distribution as n gets large; extremes drawn from normal and lognormal distributions fall in the domain of attraction of the Gumbel, and obey the Gumbel as n gets large, and so on.

Observe that most of the probability mass is located between x values of –2 and +6. More generally, this means most of the probability mass will lie between x values of $\mu - 2\sigma$ and $\mu + 6\sigma$.

To obtain the quantiles associated with the GEV distribution, we set the left-hand side of Equation (7.1) to p, take logs of both sides of Equation (7.1) and rearrange to get:

$$\ln(p) = \begin{cases} -\left\{1 + \xi\left(\dfrac{x-\mu}{\sigma}\right)\right\}^{-1/\xi} & \xi \neq 0 \\ \text{if} & \\ -\exp\left\{-\left(\dfrac{x-\mu}{\sigma}\right)\right\} & \xi = 0 \end{cases} \tag{7.2}$$

We then unravel the x-values to get the quantiles associated with any chosen (cumulative) probability p:[5]

$$x = \mu - \frac{\sigma}{\xi}[1 - (-\ln(p))^{-\xi}] \quad \text{(Fréchet, } \xi > 0) \tag{7.3a}$$

$$x = \mu - \sigma \ln[-\ln(p)] \quad \text{(Gumbel, } \xi = 0) \tag{7.3b}$$

Example 7.1 (Gumbel quantiles)

For the standardised Gumbel, the 5% quantile is $-\ln[-\ln(0.05)] = -1.0972$ and the 95% quantile is $-\ln[-\ln(0.95)] = 2.9702$.

Example 7.2 (Fréchet quantiles)

For the standardised Fréchet with $\xi = 0.2$, the 5% quantile is $-(1/0.2)[1 - (-\ln(0.05))^{-0.2}] = -0.9851$ and the 95% quantile is $-(1/0.2)[1 - (-\ln(0.95))^{-0.2}] = 4.0564$. For $\xi = 0.3$, the 5% quantile is $-(1/0.3)[1 - (-\ln(0.05))^{-0.3}] = -0.9349$ and the 95% quantile is $-(1/0.3)[1 - (-\ln(0.95))^{-0.3}] = 4.7924$. Thus, Fréchet quantiles are sensitive to the value of the tail index ξ, and tend to rise with ξ. Conversely, as $\xi \to 0$, the Fréchet quantiles tend to their Gumbel equivalents.

We need to remember that the probabilities in Equations (7.1)–(7.3) refer to the probabilities associated with the extreme loss distribution, not to those associated with the distribution of the 'parent' loss distribution from which the extreme losses are drawn. For example, a 5th percentile in Equation (7.3) is the cut-off point between the lowest 5% of extreme (high) losses and the highest 95% of extreme (high) losses; it is *not* the 5th percentile point of the parent distribution. The 5th percentile of the extreme loss distribution is therefore on the *left*-hand side

[5] We can obtain estimates of EV VaR over longer time periods by using appropriately scaled parameters, bearing in mind that the mean scales proportionately with the holding period h, the standard deviation scales with the square root of h, and (subject to certain conditions) the tail index does not scale at all. In general, we find that the VaR scales with a parameter κ (i.e., so $VaR(h) = VaR(1)(h)^{\kappa}$, where h is the holding period), and empirical evidence reported by Hauksson *et al.* (2001, p. 93) suggests an average value for κ of about 0.45. The square-root scaling rule (i.e., $\kappa = 0.5$) is therefore usually inappropriate for EV distributions.

of the distribution of extreme losses (because it is a *small* extreme loss), but on the *right*-hand tail of the original loss distribution (because it *is* an extreme loss).

To see the connection between the probabilities associated with the distribution of M_n and those associated with the distribution of X, we now let ${M_n}^*$ be some extreme threshold value. It then follows that:

$$\Pr\left[M_n < {M_n}^*\right] = p = \left\{\Pr\left[X < {M_n}^*\right]\right\}^n = [\alpha]^n \tag{7.4}$$

where α is the VaR confidence level associated with the threshold ${M_n}^*$. To obtain the α VaR, we now use Equation (7.4) to substitute $[\alpha]^n$ for p in Equation (7.3), and this gives us:

$$VaR = \mu_n - \frac{\sigma_n}{\xi_n}[1 - (-n\ln(\alpha))^{-\xi_n}] \quad \text{(Fréchet, } \xi > 0) \tag{7.5a}$$

$$VaR = \mu_n - \sigma_n \ln[-n\ln(\alpha)] \quad \text{(Gumbel, } \xi = 0) \tag{7.5b}$$

(Since n is now explicit, we have also subscripted the parameters with n to make explicit that in practice these would refer to the parameters associated with maxima drawn from samples of size n. This helps to avoid errors with the limiting VaRs as n gets large.) Given values for the extreme-loss distribution parameters μ_n, σ_n and (where needed) ξ_n, Equation (7.5) allows us to estimate the relevant VaRs. Of course, the VaR formulas given by Equation (7.5) are meant only for extremely high confidence levels, and we cannot expect them to provide accurate estimates for VaRs at low confidence levels.

Example 7.3 (Gumbel VaR)

For the standardised Gumbel and $n = 100$, the 99.5% VaR is $-\ln[-100 \times \ln(0.995)] = 0.6906$, and the 99.9% VaR is $-\ln[-100 \times \ln(0.999)] = 2.3021$.

Example 7.4 (Fréchet VaR)

For the standardised Fréchet with $\xi = 0.2$ and $n = 100$, the 99.5% VaR is $-(1/0.2)[1 - (-100 \times \ln(0.995))^{-0.2}] = 0.7406$ and the 99.9% VaR is $-(1/0.2)[1 - (-100 \times \ln(0.999))^{-0.2}] = 2.9237$. For $\xi = 0.3$, the 99.5% VaR is $-(1/0.3)[1 - (-100 \times \ln(0.995))^{-0.3}] = 0.7674$ and the 99.9% VaR is $-(1/0.3)[1 - (-100 \times \ln(0.999))^{-0.3}] = 3.3165$.

These results tell us that EV-VaRs (and, by implication, other EV risk measures) are sensitive to the value of the tail index ξ_n, which highlights the importance of getting a good estimate of ξ_n when applying EVT. This applies even if we use a Gumbel, because we should use the Gumbel only if we think ξ_n is insignificantly different from zero.

Example 7.5 (Realistic Fréchet VaR)

Suppose we wish to estimate Fréchet VaRs with more realistic parameters. For US stock markets, some fairly plausible parameters are $\mu = 2\%$, $\sigma = 0.7\%$ and $\xi = 0.3\%$. If we put these into our Fréchet VaR formula Equation (7.5a) and retain the earlier n value, the estimated 99.5% VaR (in %) is $2 - (0.7/0.3)[1 - (-100 \times \ln(0.995))^{-0.3}] = 2.537$, and the estimated

99.9% VaR (in %) is $2 - (0.7/0.3)[1 - (-100 \times \ln(0.999))^{-0.3}] = 4.322$. For the next trading day, these estimates tell us that we can be 99.5% confident of not making a loss in excess of 2.537% of the value of our portfolio, and so on.

It is also interesting to note that had we assumed a Gumbel (i.e., $\xi = 0$) we would have estimated these VaRs (again in %) to be $2 - 0.7 \times \ln[-100 \times \ln(0.995)] = 2.483$ and $2 - 0.7 \times \ln[-100 \times \ln(0.999)] = 3.612$. These are lower than the Fréchet VaRs, which underlines the importance of getting the ξ right.

How do we choose between the Gumbel and the Fréchet? There are various ways we can decide which EV distribution to use:

- If we are confident that we can identify the parent loss distribution, we can choose the EV distribution in whose domain of attraction the parent distribution resides. For example, if we are confident that the original distribution is a *t*, then we would choose the Fréchet distribution because the *t* belongs in the domain of attraction of the Fréchet. In plain English, we choose the EV distribution to which the extremes from the parent distribution will tend.
- We could test the significance of the tail index, and we might choose the Gumbel if the tail index was insignificant and the Fréchet otherwise. However, this leaves us open to the danger that we might incorrectly conclude that ξ is 0, and this could lead us to underestimate our extreme risk measures.
- Given the dangers of model risk and bearing in mind that the estimated risk measure increases with the tail index, a safer option is always to choose the Fréchet.

7.1.2 A Short-cut EV Method

There are also short-cut ways to estimate VaR (or ES) using EV theory. These are based on the idea that if $\xi > 0$, the tail of an extreme loss distribution follows a power-law times a slowly varying function:

$$F(x) = k(x)x^{-1/\xi} \tag{7.6}$$

where $k(x)$ varies slowly with x. For example, if we assume for convenience that $k(x)$ is approximately constant, then Equation (7.6) becomes:

$$F(x) \approx kx^{-1/\xi} \tag{7.7}$$

Now consider two probabilities, a first, 'in-sample' probability $p_{in\text{-}sample}$, and a second, smaller and typically out-of-sample probability $p_{out\text{-}of\text{-}sample}$. Equation (7.7) implies:

$$p_{in\text{-}sample} \approx kx_{in\text{-}sample}^{-1/\xi} \quad \text{and} \quad p_{out\text{-}of\text{-}sample} \approx kx_{out\text{-}of\text{-}sample}^{-1/\xi} \tag{7.8}$$

which in turn implies:

$$\begin{aligned} \frac{p_{in\text{-}sample}}{p_{out\text{-}of\text{-}sample}} &\approx \left(\frac{x_{in\text{-}sample}}{x_{out\text{-}of\text{-}sample}}\right)^{-1/\xi} \\ \Rightarrow \quad x_{out\text{-}of\text{-}sample} &\approx x_{in\text{-}sample}\left(\frac{p_{in\text{-}sample}}{p_{out\text{-}of\text{-}sample}}\right)^{\xi} \end{aligned} \tag{7.9}$$

This allows us to estimate one quantile (denoted here as $x_{out\text{-}of\text{-}sample}$) based on a known in-sample quantile $x_{in\text{-}sample}$, a known out-of-sample probability $p_{out\text{-}of\text{-}sample}$ (which is known

because it comes directly from our VaR confidence level), and an unknown in-sample probability $p_{in\text{-}sample}$. The latter can easily be proxied by its empirical counterpart, t/n, where n is the sample size and t the number of observations higher than $x_{in\text{-}sample}$. Using this proxy then gives us:

$$x_{out\text{-}of\text{-}sample} \approx x_{in\text{-}sample}\left(\frac{np_{out\text{-}of\text{-}sample}}{t}\right)^{-\xi} \tag{7.10}$$

which is easy to estimate using readily available information.

To use this approach, we take an arbitrarily chosen in-sample quantile, $x_{in\text{-}sample}$, and determine its counterpart empirical probability, t/n. We then determine our out-of-sample probability from our chosen confidence level, estimate our tail index using a suitable method, and our out-of-sample quantile estimator immediately follows from Equation (7.10).[6]

7.1.3 Estimation of EV Parameters

To estimate EV risk measures, we need to estimate the relevant EV parameters – μ, σ and, in the case of the Fréchet, the tail index ξ, so we can insert their values into our quantile formulas (i.e., Equation (7.5)). We can obtain estimators using maximum likelihood (ML) methods, regression methods, moment-based or semi-parametric methods.

7.1.3.1 ML estimation methods

ML methods derive the most probable parameter estimators given the data, and are obtained by maximising the likelihood function. To apply an ML approach, we begin by constructing the likelihood or log-likelihood function. In the case of the Gumbel ($\xi = 0$) and with m observations for M_n, the log-likelihood function is:

$$l(\mu_n, \sigma_n) = -m\ln(\sigma_n) - \sum_{i=1}^{m}\exp\left(-\frac{M_n - \mu_n}{\sigma_n}\right) - \sum_{i=1}^{m}\frac{M_n - \mu_n}{\sigma_n} \tag{7.11}$$

Where $\xi \neq 0$ the log-likelihood function is:

$$\begin{aligned} l(\mu_n, \sigma_n, \xi_n) = \; & -m\ln(\sigma_n) - (1 + 1/\xi_n)\sum_{i=1}^{m}\ln\left[1 + \xi_n\left(\frac{M_n - \mu_n}{\sigma_n}\right)\right] \\ & -\sum_{i=1}^{m}\ln\left[1 + \xi_n\left(\frac{M_n - \mu_n}{\sigma_n}\right)\right]^{-\frac{1}{\xi_n}} \end{aligned} \tag{7.12}$$

which would be maximised subject to the constraint that any observation M_n^i satisfies $1 + \xi\left(M_n^i - \mu\right)/\sigma > 0$. The ML approach has some attractive properties (e.g., it is statistically well grounded, parameter estimators are consistent and asymptotically normal if $\xi_n > -1/2$, we can easily test hypotheses about parameters using likelihood ratio statistics, etc.). However, it also lacks closed-form solutions for the parameters, so the ML approach requires the use of an appropriate numerical solution method. This requires suitable software, and there is the

[6] An alternative short-cut is suggested by Diebold *et al.* (2000). They suggest that we take logs of Equation (7.7) and estimate the log-transformed relationship using regression methods. However, their method is still relatively untried, and its reliability is doubtful because there is no easy way to ensure that the regression procedure will produce a 'sensible' estimate of the tail index.

danger that ML estimators might not be robust. In addition, because the underlying theory is asymptotic, there is also the potential for problems arising from smallness of samples.

7.1.3.2 Regression methods

An easier method to apply is a regression method due to Gumbel (1958).[7] To see how the method works, we begin by ordering our sample of M_n^i values from lowest to highest, so $M_n^1 \leq M_n^2 \leq \ldots \leq M_n^m$. Because these are order statistics, it follows that, for large n:

$$E\left[H\left(M_n^i\right)\right] = \frac{i}{1+m} \Rightarrow H\left(M_n^i\right) \approx \frac{i}{1+m} \tag{7.13}$$

where $H\left(M_n^i\right)$ is the cumulative density function of maxima, and we drop all redundant scripts for convenience. (See Equation (7.1) above.) In the case where $\xi \neq 0$, Equations (7.1) and (7.13) together give

$$\frac{i}{1+m} \approx \exp[-\left(1+\xi_n\left(M_n^i - \mu_n\right)/\sigma_n\right)^{-1/\xi}] \tag{7.14}$$

Taking logs twice of both sides yields:

$$\log\left[-\log\left(\frac{i}{1+m}\right)\right] \approx -\frac{1}{\xi_n}\log\left[1+\xi_n\left(\frac{M_n-\mu_n}{\sigma_n}\right)\right] \tag{7.15}$$

and we can obtain least squares estimates of μ_n, σ_n and ξ_n from a regression of $\log[-\log(i/(1+m))]$ against $[1+\xi_n(M_n-\mu_n)/\sigma_n]$. When $\xi = 0$, then the equivalent of Equation (7.14) is:

$$\log\left[-\log\left(\frac{i}{1+m}\right)\right] \approx \left(\frac{M_n-\mu_n}{\sigma_n}\right) \tag{7.16}$$

and the recovery of parameter estimates from a regression is straightforward.

Box 7.1 Moment-based Estimators of EV Parameters

An alternative approach is to estimate EV parameters using empirical moments. Let m_i be the ith empirical moment of our extremes data set. Assuming $\xi \neq 0$, we can adapt Embrechts *et al.* (1997, pp. 321–323) to show that:

$$\begin{aligned} m_1 &= \mu - \frac{\sigma}{\xi}(1-\Gamma(1-\xi)) \\ 2m_2 - m_1 &= \frac{\sigma}{\xi}\Gamma(1-\xi)(2^\xi - 1) \\ 3m_3 - m_1 &= \frac{\sigma}{\xi}\Gamma(1-\xi)(3^\xi - 1) \end{aligned}$$

where the $\Gamma(.)$ is a gamma function. Dividing the last of these into the preceding one gives us an implied estimator $\hat{\xi}$ of ξ. The first two equations can then be rearranged to give us

[7] See Gumbel (1958), pp. 226, 260, 296.

estimators for μ and σ in terms of $\hat{\xi}$ and sample moments m_1 and m_2:

$$\hat{\sigma} = \frac{(2m_2 - m_1)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}} - 1)}$$
$$\hat{\mu} = m_1 + \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma(1-\hat{\xi}))$$

The Gumbel equivalents are obtained by taking the limit as $\xi \to 0$. In this case

$$\hat{\sigma} = \sqrt{\frac{6m_2}{\pi}}$$
$$\hat{\mu} = m_1 + \Gamma(1)\hat{\sigma} = m_1 - 0.57722\hat{\sigma}$$

This moment-based approach is easy to apply, but, it is unreliable because of the poor sampling properties of the second- and higher-order moments.

However, following Hosking *et al.* (1985), we can obtain estimates with superior sampling properties if we replace the m_i in the above expressions with their probability-weighted counterparts w_i, where $w_i = E[X(F(X)^{i-1}]$ for $i = 1, 2, \ldots$. If we wish to, we can also replace the m_i with more general probability-weighted moments $w_{i,r,s}$, where $w_{i,r,s} = E[X^r(F(X)^{i-1}(1 - F(X))^{s-1}]$ for $i, r, s = 1, 2, \ldots$.

7.1.3.3 Semi-parametric estimation methods

We can also estimate parameters using semi-parametric methods. These are typically used to estimate the tail index ξ, and the most popular of these is the Hill estimator. This estimator is directly applied to the ordered parent loss observations. Denoting these from highest to lowest by $X_1, X_2, \ldots, X_n$, the Hill $\xi_{n,k}^{(H)}$ is:

$$\hat{\xi}_{n,k}^{(H)} = \frac{1}{k}\sum_{i=1}^{k} \ln X_i - \ln X_{k+1} \tag{7.17}$$

where k, the tail threshold used to estimate the Hill estimator, has to be chosen in an appropriate way. The Hill estimator is the average of the k most extreme (i.e., tail) observations, minus the $k + 1$th observation, or the one next to the tail. The Hill estimator is known to be consistent and asymptotically normally distributed, but its properties in finite samples are not well understood, and there are concerns in the literature about its small-sample properties and its sensitivity to the choice of threshold k. However, these (and other) reservations notwithstanding, many EVT practitioners regard the Hill estimator as being as good as any other.[8]

The main problem in practice is choosing a cut-off value for k. We know that our tail index estimates can be sensitive to the choice of k, but theory gives us little guidance on what the value of k should be. A suggestion often given to this problem is that we estimate Hill (or Pickands) estimators for a range of k values, and go for k values where the plot of estimators against k-values (hopefully) becomes more or less horizontal: if the plot stabilises and flattens

[8] An alternative is the Pickands estimator (see, e.g., Bassi *et al.* (1998), p. 125 or Longin (1996), p. 389). This estimator does not require a positive tail index (unlike the Hill estimator) and is asymptotically normal and weakly consistent under reasonable conditions, but is otherwise less efficient than the Hill estimator.

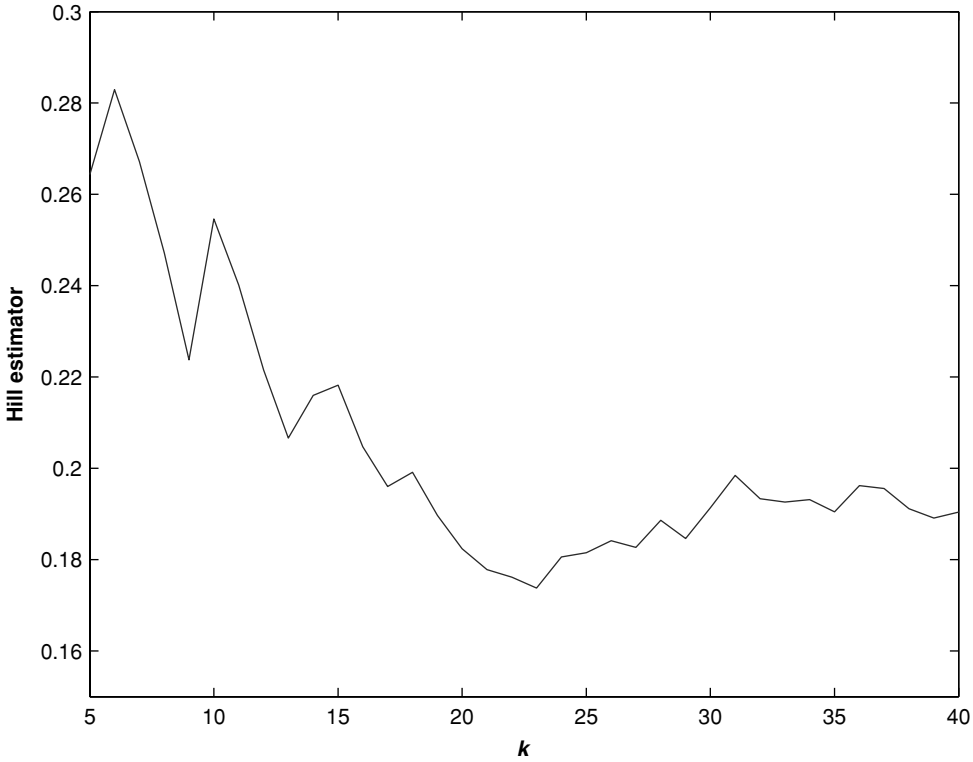

Figure 7.2 Hill plot
Note: Based on 1000 simulated drawings from a standard normal distribution.

out, then the plateau value should give a reasonable estimate of our tail index. This suggestion tries to extract the maximum possible information from all our data, albeit in an informal way.

To show how this might be done, Figure 7.2 shows a 'Hill plot' – a plot of the values of the Hill estimator against k, the tail threshold size used to estimate the Hill estimator, based on 1000 simulated observations from an underlying distribution. As we can see, the Hill estimates are a little unsteady for low values of k, but they become more stable and settle down as k gets larger, and one might suppose that the 'true' value of k lies in the region of 0.18–0.20. Such a value would be plausible for many real situations, so if we met such a situation in practice we could easily persuade ourselves that this was a fair estimate.

However, in coming to such a conclusion, we are implicitly presuming that the values of the Hill estimator do indeed settle down for values of k bigger than 40. Is this assumption justified? The answer, sadly, is that it is often not. We can see why when we extend the same plot for higher values of k: despite the fact that the values of the Hill estimator looked like they were settling down as k approached 40, it turns out that they were doing nothing of the sort. This comes as a shock. In fact, the values of the Hill estimator show no sign of settling down at all. The Hill plot becomes a 'Hill horror plot' and gives us no real guidance on how

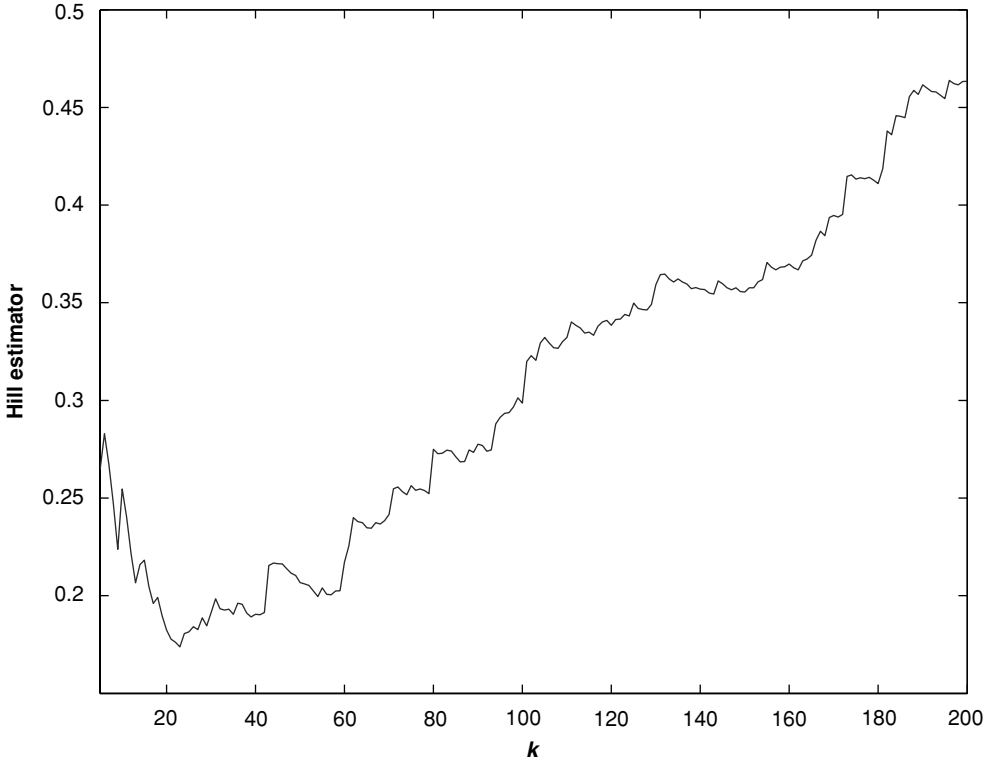

Figure 7.3 Hill horror plot
Note: Based on 1000 simulated drawings from a standard normal distribution.

we might choose k – and this means that it does not help us to determine what the value of the Hill estimator might be.[9] The Hill horror plot is shown in Figure 7.3.

Hill horror plots can be a real problem, and it is sometimes suggested that the best practical response when meeting them is to 'patch' up the estimator and hope for the best. To illustrate this in the present context, I played around a little with the above data and soon discovered that I could obtain a fairly nice Hill plot by making a very small adjustment to the Hill formula.[10] The resulting 'Hill happy plot' is shown in Figure 7.4. In this case, the values of the Hill estimator do settle down as k gets larger, and the plot suggests that we can take the best value of k to be somewhere in the region of 0.15. We have therefore 'solved' the problem of the Hill horror plot. However, this 'solution' comes at a big price: the adjustment is completely ad hoc and has no theoretical basis whatever. More to the point, we don't even know whether the answer it gives us is any good: all we really know is that we have managed to patch up the estimator to stabilise the Hill plot, but whether this actually helps us is quite a different matter.

[9] Purists might point out that we might expect a badly behaved Hill estimator when using data drawn from a normal distribution. This may be true, but it misses the main point of the exercise: Hill horror plots are all too common, and occur with many non-normal distributions as well.

[10] For those who are curious about it, the adjustment used is to add in the extra term $-0.0015k$ to the Hill formula Equation (7.17).

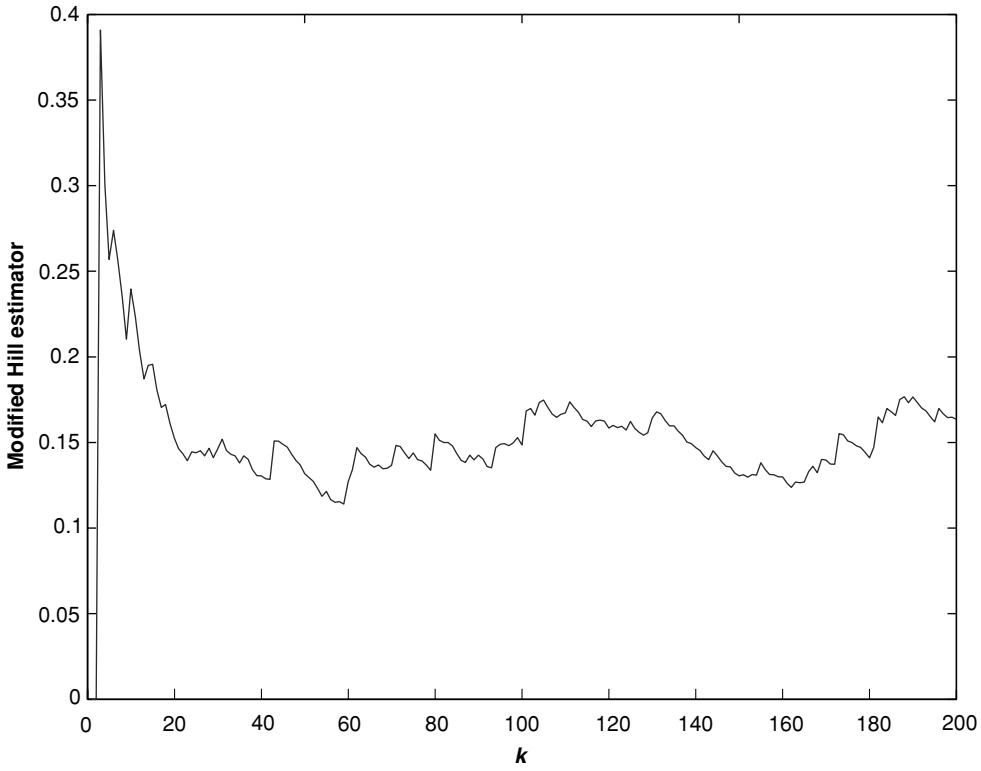

Figure 7.4 'Hill happy plot'
Note: Based on 1000 simulated drawings from a standard normal distribution.

If we have a very large sample size, we can also use an alternative method of gauging the 'right' value of k. Danielsson and de Vries (1997b) have suggested an ingenious (though rather involved) procedure based on the fact that the choice of k implies a trade-off between bias and variance. If we increase k, we get more data and so move to the centre of the distribution. This increases the precision of our estimator (and therefore reduces its variance), but also increases the bias of the tail estimator by placing relatively more weight on observations closer to the centre of our distribution. Alternatively, if we decrease k and move further out along the tail, we decrease bias but have fewer data to work with and get a higher variance. These authors suggest that we choose the value of k to minimise a mean-squared-error (MSE) loss function, which reflects an optimal trade-off, in an MSE sense, between bias and variance. The idea is that we take a second-order approximation to the tail of the distribution function $F(x)$, and exploit the point that the tail size is optimal in an asymptotic mean-squared-error sense where bias and variance disappear at the same rate. This optimal size can be found by a subsample bootstrap procedure. However, this approach requires a large sample size – at least 1500 observations – and is therefore impractical with small sample sizes. In addition, any automatic procedure for selecting k tends to ignore other, softer, but nonetheless often very useful, information, and this leads some writers to be somewhat sceptical of such methods.

Box 7.2 Estimating VaR Under Maximum Domain of Attraction Conditions

We have assumed so far that our maxima data were drawn exactly from the GEV. But what happens if our data are only approximately GEV distributed (i.e., are drawn from the maximum domain of attraction of the GEV)? The answer is that the analysis becomes somewhat more involved. Consider the Fréchet case where $\xi = 1/\alpha > 0$. The far-right tail $\bar{F}(x) = 1 - F(x)$ is now $\bar{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function L. However, the fact that the data are drawn from the maximum domain of attraction of the Fréchet also means that

$$\lim_{n\to\infty} nF(c_n x + d_n) = -\log H_\xi(x)$$

where $H_\xi(x)$ is the standardised (0 location, unit scale) Fréchet, and c_n and d_n are appropriate norming (or scaling) parameters. Invoking Equation (7.1), it follows for large $u = c_n x + d_n$ that

$$\bar{F}(u) \approx \frac{1}{n}\left(1 + \xi\frac{u - d_n}{c_n}\right)^{-\frac{1}{\xi}}$$

This leads to the quantile estimator

$$\hat{x}_p = \hat{d}_n + \frac{\hat{c}_n}{\hat{\xi}}([n(1-p)]^{-\hat{\xi}} - 1)$$

for appropriate parameter estimators $\hat{c}_n$, $\hat{d}_n$ and $\hat{\xi}$, and some high probability (confidence level) p. The problem is then to estimate p-quantiles outside the range of the data where the empirical tail $\bar{F}() = 0$. The standard approach to this problem is a subsequence trick: in effect, we replace n with n/k. This yields the quantile estimator

$$\hat{x}_p = \hat{d}_{n/k} + \frac{\hat{c}_{n/k}}{\hat{\xi}}\left(\left[\frac{n}{k}(1-p)\right]^{-\hat{\xi}} - 1\right)$$

$\hat{c}_{n/k}$ and $\hat{d}_{n/k}$ can be obtained using suitable semi-parametric methods, and $\hat{\xi}$ can be obtained using the usual Hill or other tail index approaches.[11]

7.2 THE PEAKS-OVER-THRESHOLD APPROACH: THE GENERALISED PARETO DISTRIBUTION

7.2.1 Theory

We turn now to the second strand of the EV literature, which deals with the application of EVT to the distribution of excess losses over a (high) threshold. This gives rise to the peaks-over-threshold (POT) or generalised Pareto approach, which (generally) requires fewer parameters than EV approaches based on the generalised extreme value theorem. The POT approach provides *the* natural way to model exceedances over a high threshold, in the same way that GEV theory provides the natural way to model the maxima or minima of a large sample.

[11] For more on estimation under maximum domain of attraction conditions, see Embrechts *et al.* (1997, section 6.4).

If X is a random iid loss with distribution function $F(x)$, and u is a threshold value of X, we can define the distribution of excess losses over our threshold u as:

$$F_u(x) = \Pr\{X - u \leq x | X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{7.18}$$

for $x > 0$. This gives the probability that a loss exceeds the threshold u by at most x, given that it does exceed the threshold. The distribution of X itself can be any of the commonly used distributions: normal, lognormal, t, etc., and will usually be unknown to us. However, as u gets large, the Gnedenko–Pickands–Balkema–deHaan (GPBdH) theorem states that the distribution $F_u(x)$ converges to a generalised Pareto distribution, given by:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if} \quad \xi \neq 0 \\ 1 - \exp(-x/\beta) & \quad\quad\; \xi = 0 \end{cases} \tag{7.19}$$

defined for $x \geq 0$ for $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ for $\xi < 0$. This distribution has only two parameters: a positive scale parameter, β, and a shape or tail index parameter, ξ, that can be positive, zero or negative. This latter parameter is the same as the tail index encountered already with GEV theory. The cases that usually interest us are the first two, and particularly the first (i.e., $\xi > 0$), as this corresponds to data being heavy tailed.

The GPBdH theorem is a very useful result, because it tells us that the distribution of excess losses always has the same form (in the limit, as the threshold gets high), pretty much regardless of the distribution of the losses themselves. Provided the threshold is high enough, we should therefore regard the GP distribution as *the* natural model for excess losses.

To apply the GP distribution, we need to choose a reasonable threshold u, which determines the number of observations, N_u, in excess of the threshold value. Choosing u involves a trade-off: we want a threshold u to be sufficiently high for the GPBdH theorem to apply reasonably closely; but if u is too high, we won't have enough excess-threshold observations on which to make reliable estimates. We also need to estimate the parameters ξ and β. As with the GEV distributions, we can estimate these using maximum likelihood approaches or semi-parametric approaches.

We now rearrange the right-hand side of Equation (7.18) and move from the distribution of exceedances over the threshold to the parent distribution $F(x)$ defined over 'ordinary' losses:

$$F(x) = (1 - F(u))G_{\xi,\beta}(x - u) + F(u) \tag{7.20}$$

where $x > u$. To make use of this equation, we need an estimate of $F(u)$, the proportion of observations that do not exceed the threshold, and the most natural estimator is the observed proportion of below-threshold observations, $(n - N_u)/n$. We then substitute this for $F(u)$, and plug Equation (7.19) into Equation (7.20):

$$F(x) = 1 - \frac{N_u}{n}\left[1 + \xi\left(\frac{x-u}{\beta}\right)\right]^{-1/\xi} \tag{7.21}$$

The VaR is given by the x-value in Equation (7.21), which can be recovered by inverting Equation (7.21) and rearranging to get:

$$VaR = u + \frac{\beta}{\xi}\left\{\left[\frac{n}{N_u}(1-\alpha)\right]^{-\xi} - 1\right\} \tag{7.22}$$

where α, naturally, is the VaR confidence level.

The ES is then equal to the VaR plus the mean–excess loss over VaR. Provided $\xi < 1$, our ES is:

$$ES = \frac{VaR}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \tag{7.23}$$

Example 7.5 (POT risk measures)

Suppose we set our parameters at some empirically plausible values denominated in % terms (i.e., $\beta = 0.8$, $\xi = 0.15$, $u = 2\%$ and $N_u/n = 4\%$; these are based on the empirical values associated with contracts on futures clearinghouses). The 99.5% VaR (in %) is therefore

$$VaR = 2 + \frac{0.8}{0.15}\left\{\left[\frac{1}{0.04}(1-0.995)\right]^{-0.15} - 1\right\} = 3.952$$

The corresponding ES (in %) is

$$ES = \frac{3.952}{1-0.15} + \frac{0.8 - 0.15 \times 2}{1-0.15} = 5.238$$

If we change the confidence level to 99.9%, the VaR and ES are easily shown to be 5.942 and 7.578.

7.2.2 Estimation

To obtain estimates, we need to choose a reasonable threshold u, which then determines the number of excess-threshold observations, N_u. The choice of threshold is the weak spot of POT theory: it is inevitably arbitrary and therefore judgemental. Choosing u also involves a trade-off: we want the threshold u to be sufficiently high for the GPBdH theorem to apply reasonably closely; but if u is too high, we will not have enough excess-threshold observations from which to obtain reliable estimates. This threshold problem is very much akin to the problem of choosing k to estimate the tail index. We can also (if we are lucky!) deal with it in a similar way. In this case, we would plot the mean–excess function, and choose a threshold where the MEF becomes horizontal. We also need to estimate the parameters ξ and β and, as with the earlier GEV approaches, we can estimate these using maximum likelihood or other appropriate methods.[12] Perhaps the most reliable are the ML approaches, which involve the maximisation of the following log-likelihood:

$$l(\xi, \beta) = \begin{cases} -m \ln \beta - (1 + 1/\xi) \sum_{i=1}^{m} \ln(1 + \xi X_i/\beta) & \xi \neq 0 \\ -m \ln \beta - (1/\beta) \sum_{i=1}^{m} X_i & \xi = 0 \end{cases} \quad \text{if} \tag{7.23}$$

[12] We can also estimate these parameters using moment-based methods, as for the GEV parameters (see Box 7.2). For the GPD, the parameter estimators are $\beta = 2m_1m_2/(m_1 - 2m_2)$ and $\xi = 2 - m_1/(m_1 - 2m_2)$ (see, e.g., Embrechts *et al.* (1997), p. 358). However, as with their GEV equivalents, moment-based estimators can be unreliable, and the probability-weighted or ML ones are usually to be preferred.

subject to the conditions on which $G_{\xi,\beta}(x)$ is defined. Provided $\xi > -0.5$, ML estimators are asymptotically normal, and therefore (relatively) well behaved.

7.2.3 GEV vs POT

Both GEV and POT approaches are different manifestations of the same underlying EV theory, although one is geared towards the distribution of extremes as such, whereas the other is geared towards the distribution of exceedances over a high threshold. In theory, there is therefore not too much to choose between them, but in practice there may sometimes be reasons to prefer one over the other:

- One might be more natural in a given context than the other (e.g., we may have limited data that would make one preferable).
- The GEV typically involves an additional parameter relative to the POT, and the most popular GEV approach, the block maxima approach (which we have implicitly assumed so far), can involve some loss of useful data relative to the POT approach, because some blocks might have more than one extreme in them. Both of these are disadvantages of the GEV relative to the POT.
- On the other hand, the POT approach requires us to grapple with the problem of choosing the threshold, and this problem does not arise with the GEV.

However, at the end of the day, either approach is usually reasonable, and one should choose the one that seems to best suit the problem at hand.

7.3 REFINEMENTS TO EV APPROACHES

Having outlined the basics of EVT and its implementation, we now consider some refinements to it. These fall under three headings:

- Conditional EV.
- Dealing with dependent (or non-iid) data.
- Multivariate EVT.

7.3.1 Conditional EV

The EVT procedures described above are all unconditional: they are applied directly (i.e., without any adjustment) to the random variable of interest, X. As with other unconditional applications, unconditional EVT is particularly useful when forecasting VaR or ES over a long horizon period. However, it will sometimes be the case that we wish to apply EVT to X adjusted for (i.e., conditional on) some dynamic structure, and this involves distinguishing between X and the random factors driving it. This conditional or dynamic EVT is most useful when we are dealing with a short horizon period, and where X has a dynamic structure that we can model. A good example is where X might be governed by a GARCH process. In such circumstances we might want to take account of the GARCH process and apply EVT not to the raw return process itself, but to the random innovations that drive it.

One way to take account of this dynamic structure is to estimate the GARCH process and apply EVT to its residuals. This suggests the following two-step procedure:[13]

- We estimate a GARCH-type process (e.g., a simple GARCH, etc.) by some appropriate econometric method and extract its residuals. These should turn out to be iid. The GARCH-type model can then be used to make one-step ahead predictions of next period's location and scale parameters, μ_{t+1} and σ_{t+1}.
- We apply EVT to these residuals, and then derive VaR estimates taking account of both the dynamic (i.e., GARCH) structure and the residual process.

7.3.2 Dealing with Dependent (or Non-iid) Data

We have assumed so far that the stochastic process driving our data is iid, but most financial returns exhibit some form of time dependency (or pattern over time). This time dependency usually takes the form of clustering, where high/low observations are clustered together. Clustering matters for a number of reasons:

- It violates an important premise on which the earlier results depend, and the statistical implications of clustering are not well understood.
- There is evidence that data dependence can produce very poor estimator performance.[14]
- Clustering alters the interpretation of our results. For example, we might say that there is a certain quantile or VaR value that we would expect to be exceeded, on average, only once every so often. But if data are clustered, we do not know how many times to expect this value to be breached in any given period: how frequently it is breached will depend on the tendency of the breaches to be clustered.[15] Clustering therefore has an important effect on the interpretation of our results.

There are two simple methods of dealing with time dependency in our data. Perhaps the most common (and certainly the easiest) is just to apply GEV distributions to block maxima. This is the simplest and most widely used approach. It exploits the point that maxima are usually less clustered than the underlying data from which they are drawn, and become even less clustered as the periods of time from which they are drawn get longer. We can therefore completely eliminate time dependence if we choose long enough block periods. This block maxima approach is very easy to use, but involves some efficiency loss, because we throw away extreme observations that are not block maxima. There is also the drawback that there is no clear guide about how long the block periods should be, which leads to a new bandwidth problem comparable to the earlier problem of how to select k.

A second solution to the problem of clustering is to estimate the tail of the conditional distribution rather than the unconditional one: we would first estimate the conditional volatility model (e.g., via a GARCH procedure), and then estimate the tail index of conditional standardised data. The time dependency in our data is then picked up by the deterministic part of our model, and we can treat the random process as independent.[16]

[13] This procedure is developed in more detail by McNeil and Frey (2000).
[14] See, e.g., Kearns and Pagan (1997).
[15] See McNeil (1998), p. 13.
[16] There is also a third, more advanced but also more difficult, solution. This is to estimate an extremal index – a measure of clustering – and use this index to adjust our quantiles for clustering. For more details on the extremal index and how to use it, see, e.g., Embrechts *et al.* (1997, Chapter 8.1).

7.3.3 Multivariate EVT

We have been dealing so far with univariate EVT, but there also exists multivariate extreme-value theory (MEVT), which can be used to model the tails of multivariate distributions in a theoretically appropriate way. The key issue here is how to model the dependence structure of extreme events. To appreciate this issue, it is again important to recognise how EV theory differs from more familiar central-value theory. As we all know, when dealing with central values, we often rely on the central limit theorem to justify the assumption of a normal (or more broadly, elliptical) distribution. When we have such a distribution, the dependence structure can then be captured by the (linear) correlations between the different variables. Given our distributional assumptions, knowledge of variances and correlations (or, if we like, covariances) suffices to specify the multivariate distribution. This is why correlations are so important in central-value theory.

However, this logic does not carry over to extremes. When we go beyond elliptical distributions, correlation no longer suffices to describe the dependence structure. Instead, the modelling of multivariate extremes requires us to make use of copulas. MEVT tells us that the limiting distribution of multivariate extreme values will be a member of the family of EV copulas, and we can model multivariate EV dependence by assuming one of these EV copulas. In theory, our copulas can also have as many dimensions as we like, reflecting the number of random variables to be considered. However, there is a curse of dimensionality here. For example, if we have two independent variables and classify univariate extreme events as those that occur one time in a 100, then we should expect to see one multivariate extreme event (i.e., both variables taking extreme values) only one time in 100^2, or one time in 10 000 observations; with three independent variables, we should expect to see a multivariate extreme event one time in 100^3, or one time in 1 000 000 observations, and so on. As the dimensionality rises, our multivariate EV events rapidly become much rarer: we have fewer multivariate extreme observations to work with, and more parameters to estimate. There is clearly a limit to how many dimensions we can handle.

One might be tempted to conclude from this example that multivariate extremes are sufficiently rare that we need not worry about them. However, this would be a big mistake. Even in theory, the occurrence of multivariate extreme events depends on their joint distribution, and extreme events cannot be assumed to be independent. Instead, as discussed in the appendix to Chapter 5, the occurrence of such events is governed by the tail dependence of the multivariate distribution. Indeed, it is for exactly this reason that tail dependence is the central focus of MEVT. And, as a matter of empirical fact, it is manifestly obvious that (at least some) extreme events are not independent: a major earthquake can trigger other natural or financial disasters (e.g., tsunamis or market crashes). We all know that disasters are often related. It is therefore important for risk managers to have some awareness of multivariate extreme risks.

7.4 CONCLUSIONS

EVT provides a tailor-made approach to the estimation of extreme probabilities and quantiles. It is intuitive and plausible; and it is relatively easy to apply, at least in its more basic forms. It also gives us considerable practical guidance on what we should estimate and how we should do it; and it has a good track record. It therefore provides *the* ideal, tailor-made, way to estimate extreme risk measures.

EVT is also important in what it tells us *not* to do, and the most important point is not to use distributions justified by central limit theory – most particularly, the normal or Gaussian distribution – for extreme-value estimation. If we wish to estimate extreme risks, we should do so using the distributions suggested by EVT, not arbitrary distributions (such as the normal) that go against what EVT tells us.

But we should not lose sight of the limitations of EV approaches, and certain limitations stand out:

- EV problems are intrinsically difficult, because by definition we always have relatively few extreme-value observations to work with. This means that any EV estimates will necessarily be very uncertain, relative to any estimates we might make of more central quantiles or probabilities. EV estimates will therefore have relatively wide confidence intervals attached to them. This uncertainty is not a fault of EVT as such, but an inevitable consequence of our paucity of data.
- EV estimates are subject to considerable model risk. We have to make various assumptions in order to carry out extreme-value estimations, and our results will often be very sensitive to the precise assumptions we make. At the same time, the veracity or otherwise of these assumptions can be difficult to verify in practice. Hence, our estimates are often critically dependent on assumptions that are effectively unverifiable. EVT also requires us to make ancillary decisions about threshold values and the like, and there are no easy ways to make those decisions: the application of EV methods involves a lot of subjective 'judgement'. Because of this uncertainty, it is especially important with extremes to estimate confidence intervals for our estimated risk measures and to subject the latter to stress testing.
- Because we have so little data and the theory we have is (mostly) asymptotic, EV estimates can be very sensitive to small sample effects, biases, non-linearities, and other unpleasant problems.

In the final analysis, we need to make the best use of theory while acknowledging that the paucity of our data inevitably limits the reliability of our results. To quote McNeil,

> We are working in the tail . . . and we have only a limited amount of data which can help us. The uncertainty in our analyses is often high, as reflected by large confidence intervals . . . . However, if we wish to quantify rare events we are better off using the theoretically supported methods of EVT than other ad hoc approaches. EVT gives the best estimates of extreme events and represents the most honest approach to measuring the uncertainty inherent in the problem.[17]

Thus EVT has a very useful, albeit limited, role to play in risk measurement. As Diebold *et al*. nicely put it:

> EVT is here to stay, but we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations – as well as its strengths. When the smoke clears, the contribution of EVT remains basic and useful: It helps us to draw smooth curves through the extreme tails of empirical survival functions in a way that is guided by powerful theory. . . . [But] we shouldn't ask more of the theory than it can deliver.[18]

[17] McNeil (1998, p. 18).
[18] Diebold *et al*. (2000), p. 34.

8

Monte Carlo Simulation Methods

This chapter and the next deal with the use of Monte Carlo simulation methods to estimate measures of financial risk. These methods have a long history in science and engineering, and were first developed in the 1940s to help deal with some of the calculations involved in nuclear physics. They then became widely used for other problems, such as those involved with numerical integration. In the finance area, they have been used since the late 1970s to price derivatives and estimate their Greek hedge ratios, and they have been adapted more recently to estimate VaRs and other financial risk measures. They are extremely flexible and powerful, and can be used to tackle all manner of otherwise difficult calculation problems.

The idea behind Monte Carlo methods is to simulate repeatedly from the random processes governing the prices or returns of the financial instruments we are interested in. For example, if we were interested in estimating a VaR, each simulation would give us a possible value for our portfolio at the end of our holding period. If we take enough of these simulations, the simulated distribution of portfolio values will converge to the portfolio's unknown 'true' distribution, and we can use the simulated distribution of end-period portfolio values to infer the VaR.

This simulation process involves a number of specific steps. The first is to select a model for the stochastic variable(s) of interest. Having chosen our model, we estimate its parameters – volatilities, correlations, and so on – on the basis of 'judgement' or whatever historical or market data are available. We then construct fictitious or simulated paths for the stochastic variables. Each set of 'random' numbers then produces a set of hypothetical terminal price(s) for the instrument(s) in our portfolio. We then repeat these simulations enough times to be confident that the simulated distribution of portfolio values is sufficiently close to the 'true' (but unknown) distribution of actual portfolio values to be a reliable proxy for it. Once that is done, we can infer the VaR from this proxy distribution.

MCS methods can be used to address problems of almost any degree of complexity, and can easily address factors – such as path dependency, fat tails, non-linearity and optionality – that most other approaches have difficulty with. Simulation approaches are also particularly useful when dealing with multidimensional problems (i.e., where outcomes depend on more than one risk variable) and, as a rule, we can say that they become relatively more attractive as the complexity and/or dimensionality of a problem increases.

However, there is no point using such powerful methods in cases where simpler approaches are adequate: there is no point using a sledgehammer to crack open a walnut. So if we are trying to price a Black–Scholes vanilla call option, there would be no point using simulation methods because we can solve this problem very easily using the Black–Scholes pricing equation; similarly, if we are trying to estimate a normal VaR, we would use an appropriate formula that is known to give us the correct answer. We would therefore use simulation methods only in more difficult situations where such simple solutions are unavailable.

Given that we wish both to explain Monte Carlo methods and to discuss the many ways in which they can be applied to risk estimation problems, it is convenient to divide our discussion into two parts: an initial explanation of the methods themselves, and a more detailed discussion of the ways in which they can be used to estimate market risk measures. Hence this chapter gives the initial explanation, and the next chapter looks at the risk measurement applications.[1]

8.1 USES OF MONTE CARLO SIMULATION

To motivate our discussion, we begin by illustrating some of the main financial applications of Monte Carlo methods. One such application is to price a derivative. To do so, we simulate sample paths of the underlying, say stock, price S in a risk-neutral world (i.e., typically, we assume a GBM process and run sample paths taking the expected return to be the risk-free return r instead of μ), and calculate the payoff from the derivative at the end of each path (e.g., so the payoff from a standard Black–Scholes call with strike price X would be $\max(S_T - X,0)$) where S_T is the terminal stock price. We do this a large number (M) of times, calculate the sample mean payoff to our derivative, and discount this at the risk-free rate to obtain our derivative price. More details of the procedure involved are given in the next section.

To give a practical illustration, suppose we apply this method to a standard (i.e., vanilla) Black–Scholes call option with $S_0 = X = 1$, $\mu = r = 0$, $\sigma = 0.25$ and a maturity of 1 year, but with M taking values up to 5000. The results of this exercise are presented in Figure 8.1, and show that the simulated call price is initially unstable, but eventually settles down and converges towards the 'true' Black–Scholes call price of 0.0995. However, the figure also makes it clear that we need a lot of simulation trials (i.e., a large M value) to get accurate results.

A second use of MCS is to estimate the Greek parameters of option positions. The idea is to estimate the value of our derivatives position for two (or more, as relevant) slightly different values of the underlying value. The results of these exercises then give us estimates of the Greek parameters. For example, the delta, δ, of a standard European call is approximately equal to the ratio of the change in option price to the corresponding (small) change in the underlying stock price:

$$\delta \approx \frac{\Delta c}{\Delta S} = \frac{c(S+h) - c(S-h)}{2h} \tag{8.1}$$

where the option price, c, say, is written as a function of the underlying variable, S, and the S-values are perturbed slightly each way so that their difference, ΔS, is equal to $2h$. When estimating these parameters, each of the two sets of underlying prices (i.e., $S + h$ and $S - h$) is subject to random 'sampling' error, but we can reduce their combined effect and the number of calculations needed by using the same set of simulated S-values to determine both sets of underlying prices: in short, we run one set of simulations for S, perturb the S-values each way (i.e., up by h and down by h), determine two sets of option values, and thence obtain an

[1] Traditional Monte Carlo methods are based on a drawing from a random (or strictly, pseudo-random) number generator. However, more recently, newer Monte Carlo methods have been suggested based on quasi-random (or low discrepancy) numbers. Neither approach uses real random numbers: the former uses numbers that appear to be random, and the latter uses numbers that spread evenly and don't even look random. The quasi methods are very promising, but in the interests of brevity (and bearing in mind that quasi methods have yet to have much impact on the financial risk measurement literature) we do not discuss them here. For more on them and their financial applications, a good starting point is Jäckel (2002).

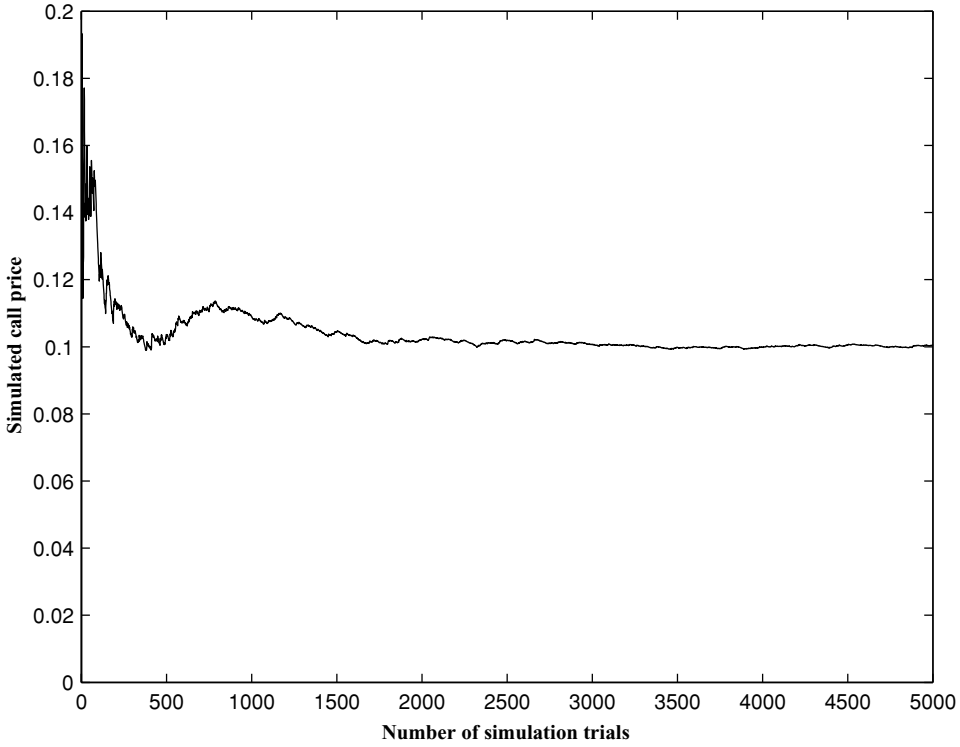

Figure 8.1 Monte Carlo simulation of a vanilla call price
Note: Based on assumed parameter values, $S = X = 1$, $r = \mu = 0$, $\sigma = 0.25$, and maturity = 1 year.

estimate of the delta.[2] We can obtain estimates of the other Greek parameters in a similar way, using discrete approximations of their defining formulas.[3]

A third use of MCS is, of course, to estimate risk measures. For example, if we wish to estimate the VaR of a vanilla call position, say, we run M simulations of the terminal stock value. However, in doing so we would use the 'real' stock-price process rather than the risk-neutralised one used to price derivatives and estimate their Greeks (i.e., we use the process with μ as the drift term rather than r). The value of T now corresponds to the end of our VaR holding period, and we revalue our option for each simulated terminal stock price (e.g., using the Black–Scholes pricing equation or, if the option expires at T, the option payoff function) and subtract from this value the current price of our option. This gives us M simulated P/L values for a portfolio consisting of one option, and we obtain the position P/L by multiplying these values by the number of options in our position. The result is a set of M simulated P/L values, and we can take the VaR as the relevant order statistic or read the VaR off from the distribution of simulated P/L values.

[2] We have to make sure that we obtain both our up and down paths from the same set of underlying simulations. If we run two separate sets of simulated underlying price paths, and estimate the delta by plugging these into Equation (8.1), the variance of our delta estimator will be of order $1/h^2$, so the variance will get very large as h gets small. Such estimates are clearly very unsatisfactory. On the other hand, if we use one set of simulated underlying price paths, the variance of our delta estimator will be of order 1, and will therefore get small as h gets small. See Boyle *et al.* (1997), pp. 1304–1305.

[3] See, e.g., Boyle *et al.* (1997), pp. 1302–1309; or Clewlow and Strickland (1998), p. 105.

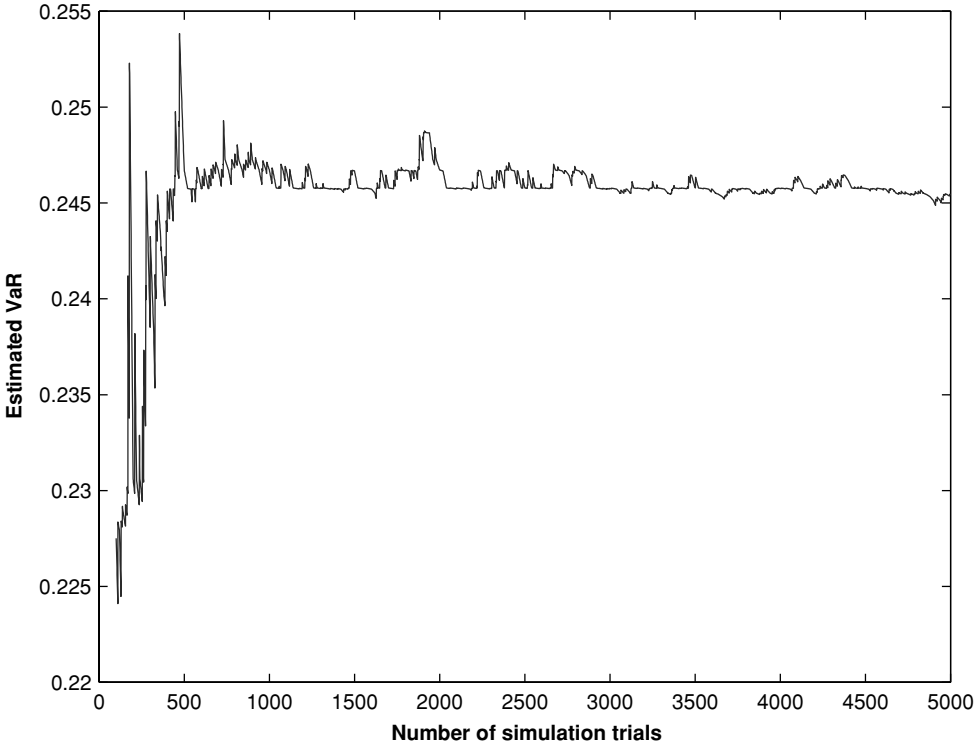

Figure 8.2 Monte Carlo simulation of a vanilla call VaR
Note: Based on assumed parameter values $S = 1, X = 0.5, r = \mu = 0, \sigma = 0.25$, maturity = 1 year, $hp =$ 5 days, confidence level = 0.95 and an investment of \$1.

To illustrate MC simulation of VaR, suppose we invest \$1 in a vanilla Black–Scholes call option with the same parameters as in Figure 8.1. We now assume a confidence level of 95% and a holding period of 5 days, and simulate the VaR of this position with M-values of up to 5000. The results of this exercise are presented in Figure 8.2, and show that the simulated VaR is initially unstable, but slowly settles down and (very) gradually converges to its 'true' value of around 0.245. However, the figure also makes it clear that we need a large number of trials to get accurate results. It suggests, too, that to achieve any given level of accuracy, we generally need a larger number of trials when estimating risk measures than when pricing options. This makes intuitive sense, because with option pricing we are concerned about the mean of the (payoff) distribution, whereas with risk measures we are concerned about the tail of the (P/L) distribution, and we know that for any given sample size, estimates of means are more accurate than estimates of tail quantities.

This same approach can also be extended to estimate other measures of financial risk. Essentially, we would use Monte Carlo to produce a simulated distribution P/L or loss distribution, and we then apply one of the 'weighted average quantile' methods discussed in Chapter 3 to obtain an estimate of the desired risk measure.[4]

[4] Monte Carlo simulation can also be used for other purposes: for example, it can be used to estimate confidence intervals for risk measures (see Box 6.4) and to estimate model risk (see Chapter 16, section 16.3).

Box 8.1 Generating Random Numbers

Monte Carlo simulations depend on drawings from a 'random' number generator. However, strictly speaking, these 'random' numbers are not random at all. They are 'pseudo'-random numbers generated from an algorithm using a deterministic rule (i.e., a rule that does not have any random elements), and such a rule *cannot* produce actual random numbers. These rules take some initial value, a 'seed' number, and then generate a series of numbers that *appear* random and ought, if the number generator is well designed, to pass most tests for randomness. There are many 'pseudo'-random number generators available (linear congruential generators, multiple-recursive generators, Mersenne twisters, etc.), and care needs to be taken when choosing one to use: even standard computer packages have been known to have problematic generators, such as the infamous RANDU generator in early IBM packages. Such problems manifest themselves in 'hidden' structure, such as repeated patterns in high dimensions. It is therefore good practice for users of Monte Carlo methods to build up a small library of alternative generators, and periodically check results across different generators.

We should also keep in mind that a random number generator will always generate the same sequence of numbers from the same initial 'seed' number. Eventually, the seed number will recur and the sequence of 'random' numbers will repeat itself all over again. All random number generators therefore cycle, but generators with higher cycle periods are generally better than those with smaller periods. We also have to make sure that the cycle is long enough for the problem at hand, and one useful rule of thumb is that the cycle should be well over the square of the number of pseudo-random numbers needed in any application: so, for example, if our application requires $2^{20} = 1\,048\,576$ pseudo-random numbers, then we require a generator with a cycle well in excess of $2^{40} = 1.0995$ billion billion(!). We should therefore check whether this condition holds when working with calculation-intensive applications of Monte Carlo simulation.[5]

8.2 MONTE CARLO SIMULATION WITH A SINGLE RISK FACTOR

We now look at the mechanics of Monte Carlo in more detail, and start with the simulation of a single risk factor. Suppose, for example, that we wish to simulate a stock price. Our first task is to choose a model to describe the behaviour of the stock price over time. Assume that the stock price S follows a geometric Brownian motion process:

$$\frac{dS}{S} = \mu dt + \sigma dx \tag{8.2}$$

where μ is its expected (per unit-time) rate of return, and σ is the volatility of the stock price. dx is known as a Wiener process, and can be written as $dx = \phi(dt)^{1/2}$, where ϕ is a drawing from a standard normal distribution. If we substitute out dx, we get:

$$\frac{dS}{S} = \mu dt + \sigma\phi(dt)^{1/2} \tag{8.3}$$

[5] For more on pseudo-random number generation, see, e.g., Jäckel (2002, Chapter 7).

This is the standard model used in quantitative finance, at least for stock prices,[6] and it holds that the (instantaneous) rate of increase of the stock price dS/S evolves according to its mean drift term μt and realisations from the random term ϕ. In practice, we would often work with this model in its discrete-form equivalent. If Δt is some small time increment, we can write:

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \phi \sqrt{\Delta t} \tag{8.4}$$

ΔS is the change in the stock price over the time interval Δt, and $\Delta S/S$ is its (discretised) rate of change. Equation (8.4) tells us that the rate of change of the stock price is normally distributed with mean $\mu \Delta t$ and standard deviation $\sigma \sqrt{\Delta t}$.

Now suppose that we wish to simulate the stock price over some period of length T. We would usually divide T into a large number N of small time increments Δt (i.e., we set $\Delta t = T/N$). The simplest way to simulate S is the Euler method: we take a starting value of S, $S(0)$, and draw a random value of ϕ to update S using Equation (8.4); this gives $S(\Delta t)$ from $S(0)$; we then derive $S(2\Delta t)$ from $S(\Delta t)$ in the same way; and carry on until we have a terminal value for S, $S(T)$. To spell out the process in more detail, we first rewrite Equation (8.4) as:

$$\Delta S = S\mu \Delta t + S\sigma \phi \sqrt{\Delta t} \quad \Rightarrow \quad S(t + \Delta t) = S(t)(1 + \mu \Delta t + \sigma \phi \sqrt{\Delta t}) \tag{8.5}$$

Starting from $S(0)$, we take a drawing from ϕ – say $\phi(\Delta t)$, where the term in brackets again refers to the time the drawing is taken – and so obtain a value for $S(\Delta t)$ using:

$$S(\Delta t) = S(0)(1 + \mu \Delta t + \sigma \phi(\Delta t) \sqrt{\Delta t}) \tag{8.6}$$

We now take another drawing from ϕ, $\phi(2\Delta t)$, and obtain a value for $S(2\Delta t)$ using:

$$\begin{aligned} S(2\Delta t) &= S(\Delta t)(1 + \mu \Delta t + \sigma \phi(2\Delta t) \sqrt{\Delta t}) \\ &= S(0)[1 + \mu \Delta t + \sigma \phi(\Delta t) \sqrt{\Delta t}][1 + \mu \Delta t + \sigma \phi(2\Delta t) \sqrt{\Delta t}] \end{aligned} \tag{8.7}$$

The stock price after two time increments depends on the initial stock price $S(0)$ and the realisations of ϕ after time increments of Δt and $2\Delta t$. Carrying on in this way, we eventually find that:

$$\begin{aligned} S(T) &= S(T - \Delta t)(1 + \mu \Delta t + \sigma \phi(T) \sqrt{\Delta t}) \\ &= S(0) \prod_{i=1}^{N} (1 + \mu \Delta t + \sigma \phi(i \Delta t) \sqrt{\Delta t}) \end{aligned} \tag{8.8}$$

The simulated terminal stock price thus depends on the initial stock price, the parameters μ and σ, and each of N realisations of ϕ. To simulate the behaviour of the stock price, we now use a random number generator to produce a series of simulated values of $\phi(\Delta t)$, $\phi(2\Delta t)$, ..., $\phi(T)$, and substitute these into Equations (8.6), (8.7) and so on, to produce a series of simulated stock prices $S(\Delta t)$, $S(2\Delta t)$, ..., $S(T)$.

This Euler method provides a good illustration of the mechanics of Monte Carlo simulation. In general, it produces estimates with two sources of error. The first are discretisation errors:

[6] The GBM model is widely used for equity prices because it is simple and it accommodates the main stylised features of equity prices, namely, that stock prices are non-negative, random, and tend to drift upwards (which we can accommodate by letting μ be positive). However, we should keep in mind that a GBM process is questionable even for stock prices, because it does not accommodate any excess kurtosis in stock returns. It is even less suitable for other random variables. Most obvious among these are spot interest rates, mainly because the GBM process fails to account for the well-established tendency of interest/spot rates to revert towards their mean. The GBM process is also inappropriate where the underlying variable is subject to jumps, stochastic volatility, asymmetric barriers or policy targets (e.g., as exchange rates are in some exchange rate regimes), and similar features.

given that the Brownian motion process is a continuous one, taking a discrete approximation to it will produce errors proportional to the size of the discretisation (i.e., of order Δt).[7] This error therefore falls as Δt gets smaller, and goes to zero as Δt approaches zero. The second source of error comes from the number of simulation trials: each trial produces a single simulated path for our random variable, culminating in a single terminal value for that variable. If there are M independent simulation trials, and the terminal value has mean υ and standard deviation ω, the standard error of our estimate of υ is $\omega/\sqrt{M}$. The accuracy of our estimates therefore increases with the square root of the number of trials: if we wish to double the accuracy of our estimates, we must quadruple the number of trials, and so on.[8]

However, it is often possible to cut down on the calculations involved and/or speed up the accuracy of our results. For instance, in the case of geometric Brownian motion, we can apply Itô's lemma to rewrite the stock price process as:

$$d \ln S = (\mu - \sigma^2/2)dt + \sigma dx \tag{8.9}$$

We then take a discretisation of Equation (8.9), set $t = 0$ and solve for $S(t + \Delta t)$:

$$\begin{aligned} \ln S(\Delta t) - \ln S(0) &= (\mu - \sigma^2/2)\Delta t + \sigma\phi(\Delta t)\sqrt{\Delta t} \\ \Rightarrow S(\Delta t) &= S(0)\exp[(\mu - \sigma^2/2)\Delta t + \sigma\phi(\Delta t)\sqrt{\Delta t}] \end{aligned} \tag{8.10}$$

Equation (8.10) can be used repeatedly to simulate a path at times $\Delta t, 2\Delta t$, and so on. Equation (8.10) is more useful than Equations (8.6)–(8.8) because it holds exactly, whereas the earlier equations are only true in the limit as Δt approaches zero. Consequently, if we are only interested in the terminal stock price (i.e., $S(T)$), we can jump from $S(0)$ to $S(T)$ in one giant step (i.e., we set $N = 1$ or $\Delta t = T$) using:

$$S(T) = S(0)\exp[(\mu - \sigma^2/2)T + \sigma\phi(T)\sqrt{T}] \tag{8.11}$$

Provided we are only interested in the terminal stock value, this approach is both more accurate and less time consuming than the Euler method.

8.3 MONTE CARLO SIMULATION WITH MULTIPLE RISK FACTORS

MCS can easily handle problems with more than one random risk factor. If we have two risky stock prices, our discretised geometric Brownian motion process is:

$$\begin{aligned} \Delta S_1/S_1 &= \mu_1\Delta t + \sigma_1\phi_1\sqrt{\Delta t} \\ \Delta S_2/S_2 &= \mu_2\Delta t + \sigma_2\phi_2\sqrt{\Delta t} \\ \Rightarrow \quad S_1(t + \Delta t) &= S_1(t) + S_1(t)\mu_1\Delta t + S_1(t)\sigma_1\phi_1\sqrt{\Delta t} \\ S_2(t + \Delta t) &= S_2(t) + S_2(t)\mu_2\Delta t + S_2(t)\sigma_2\phi_2\sqrt{\Delta t} \\ \text{or} \quad \begin{bmatrix} S_1(t + \Delta t) \\ S_2(t + \Delta t) \end{bmatrix} &= \begin{bmatrix} S_1(t)(1 + \mu_1\Delta t) \\ S_2(t)(1 + \mu_2\Delta t) \end{bmatrix} + \begin{bmatrix} \sigma_1 S_1(t)\phi_1\sqrt{\Delta t} \\ \sigma_2 S_2(t)\phi_2\sqrt{\Delta t} \end{bmatrix} \end{aligned} \tag{8.12}$$

[7] However, there also exist more refined methods that reduce the size of the error to something in the order of $(\Delta t)^2$ or even less. See, e.g., Broadie and Glasserman (1998), p. 182.

[8] There is also a third possible source of error. In many applications, we do not have closed-form solutions for the derivative value (e.g., as with some American options). In such cases, we may have to resort to new simulations (or other methods, such as binomial or trinomial methods) to determine derivatives values, and our estimates are then subject to error from whatever simulation or other method is used to determine these values.

where we use the obvious notation. However, in this case we allow the random terms, ϕ_1 and ϕ_2, to be correlated, which means that their expectation, $E[\phi_1\phi_2]$, is equal to ρ, the correlation between S_1 and S_2.

We now want to generate these correlated random variables, ϕ_1 and ϕ_2, and a common approach is by means of a Choleski decomposition.[9] Suppose we write the ϕ_i terms as a 2×1 matrix $\boldsymbol{\varphi}$:

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \tag{8.13}$$

If $\boldsymbol{\varepsilon}$ is a 2×1 matrix of uncorrelated standard normal variables, we can then write $\boldsymbol{\phi}$ as:

$$\boldsymbol{\phi} = \mathrm{A}\boldsymbol{\varepsilon} \tag{8.14}$$

where $\mathbf{A}$ is an appropriate 2×2 matrix. If we now post-multiply each side of Equation (8.14) by its transpose, we get:

$$\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}} = \mathrm{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\mathrm{A}^{\mathrm{T}} \tag{8.15}$$

We then take expectations of each side, noting that the expectation of the left-hand side is also equal to the correlation matrix $\mathbf{C}$:

$$E[\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}] = \mathbf{C} = \mathrm{A}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]\mathrm{A}^{\mathrm{T}} = \mathrm{AIA}^{\mathrm{T}} = \mathrm{AA}^{\mathrm{T}} \tag{8.16}$$

i.e.,

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{T}}$$

which tells us that $\mathbf{A}$, the matrix of a_{ij} terms, is the 'square-root matrix' of the correlation matrix $\mathbf{C}$. Provided $\mathbf{C}$ is positive definite, the easiest solution is a Choleski decomposition:

$$\mathbf{A} = \begin{bmatrix} 1, & 0 \\ \rho, & \sqrt{1-\rho^2} \end{bmatrix} \tag{8.17}$$

(The reader can easily verify this result by post-multiplying Equation (8.17) by its transpose to give the correlation matrix $\mathbf{C}$.) Hence, once we have the correlation matrix $\mathbf{C}$, we take its Choleski decomposition, given by Equation (8.17), and then use Equation (8.14) to generate our correlated random variables $\boldsymbol{\phi}$ from a set of uncorrelated variables $\boldsymbol{\varepsilon}$.

This two-dimensional analysis generalises easily to n random variables. Whatever the number of assets involved, the $\mathbf{A}$ matrix is still the $n \times n$ matrix square root of the $n \times n$ correlation matrix $\mathbf{C}$, as defined in Equation (8.16).

All we then need is a means of taking the Choleski decomposition when n exceeds 2, and the best solution is to program it. If we have n correlated random variables ϕ_i, and if c_{ij} denotes the individual terms of the $\mathbf{C}$ matrix, then the individual terms a_{ij} of the $\mathbf{A}$ matrix are:

$$a_{11} = \sqrt{c_{11}} \tag{8.18a}$$

$$a_{j1} = \frac{c_{j1}}{a_{11}} \quad \text{for } j = 2, \ldots, n \tag{8.18b}$$

[9] The Choleski decomposition procedure is efficient when the matrix Σ is positive definite, and more details on the approach are given in many places (e.g., Kreyszig (1999, pp. 896–897) or Hull (2003), p. 413). The main alternatives to Choleski decomposition are the eigenvalue decomposition and singular value decomposition approaches, but these are more computationally intensive. However, they have the advantage over the Choleski decomposition approach in that they are reliable when the $\boldsymbol{\Sigma}$ matrix is positive semi-definite, which the Choleski decomposition procedure is not.

$$a_{jj} = \sqrt{c_{jj} - \sum_{s=1}^{j-1} a^2_{js}} \quad \text{for } j = 2, \ldots, n \tag{8.18c}$$

$$a_{pj} = \frac{1}{a_{jj}} \sqrt{c_{pj} - \sum_{s=1}^{j-1} a_{js} a_{ps}} \quad \text{for } p = j+1, \ldots, n; j \geq 2 \tag{8.18d}$$

The Choleski decomposition therefore has a nice recursive structure that makes it straightforward to program.

The correlated random variables ϕ are then obtained from Equation (8.14), and it turns out that the individual correlated random variables ϕ_i are given by

$$\phi_i = \sum_{k=1}^{i} a_{ik} \varepsilon_k \tag{8. 19}$$

which also has a nice recursive structure: ϕ_1 is set equal to ε_1; we then solve the relevant a-equations to determine ϕ_2 from ε_1 and ε_2; after that, we calculate ϕ_3 from ε_1, ε_2 and ε_3; and so on.

Box 8.2 Full, Grid and other MCS Procedures

There are a number of ways of applying MCS with multiple risk factors. The most elaborate and (usually) most demanding is full MC, which involves the exact valuation of (i.e., the computing of price paths for) every instrument in our portfolio.

However, there are approximation procedures that can cut down on the calculations required. One of these is grid MC, which is essentially MCS applied to a mapped position: we map our portfolio onto a grid of factor values and then carry out a full valuation of the mapped portfolio. However, the number of calculations grows geometrically with the number of primary factors, so if the number of factors is large, we would need to use some other method or simplify the grid procedure further. One possible way to do this – a modified grid MC approach – is suggested by Pritsker (1997). An alternative is to apply full valuation to an approximation of the actual portfolio, the typical case being a delta-gamma approximation.

Pritsker's simulation results suggest that full MC generates the most accurate VaR estimates, but is also the most computer time-intensive procedure. He also found that the other two approaches were comparable in accuracy, but the delta–gamma MC approach was eight times faster.

8.4 VARIANCE-REDUCTION METHODS

For most MCS procedures, the accuracy of our results will vary with the square root of the number of trials, M. MCS can therefore be very computer intensive, particularly when we need a high level of accuracy. Naturally, the number of calculations required also depends on the number of random drawings we take in each trial. If we have to carry out M trials to get results of acceptable accuracy, and we have n random variables, then we would therefore have to take drawings of at least nM random variables. Moreover, where we need to take account of path dependency, each trial path for each of our n random variables would require N drawings of its own. This means we would need at least nMN random drawings, and this can be a very large

number. For example, if we have, say, $n = 50$, $M = 10\,000$, and $N = 20$ – and these figures are by no means untypical – then we would need $50 \times 10\,000 \times 20$ or 10 million random drawings. We can easily find ourselves having to run many millions of calculations.

However, it is possible to obtain great improvements in accuracy, for any given number of calculations, using one or more of a number of variance-reduction methods.

8.4.1 Antithetic Variables

One of these is the use of antithetic variables, which are often used for derivatives pricing. We begin in the usual way by generating a random number or random path ϕ, and use this number/path to produce a derivatives value, $f(+\phi)$ say. We now replace ϕ with its negative, $-\phi$, and obtain the corresponding derivatives value $f(-\phi)$. We then take our derivatives value as the average of $f(+\phi)$ and $f(-\phi)$ (i.e., $\bar{f} = [f(+\phi) + f(-\phi)]/2$), so producing one derivatives value $\bar{f}$ from one value/path ϕ. We repeat this process M times, and take our derivatives price as the average of our M values of $\bar{f}$.

Both the $f(+\phi)$'s and the $\bar{f}$'s give us estimates of the derivatives price, but the standard error of the latter is generally much lower. The reason for this can be appreciated by seeing $f(+\phi)$ as coming from an unhedged portfolio and $\bar{f}$ as coming from a hedged one. Both portfolios have the same expected payoff – and hence produce estimates of the value of our derivatives position – but our hedged portfolio has a lower standard deviation.[10] Hence, the use of antithetics enables us to estimate derivatives prices (and of course Greeks) with much greater accuracy. Antithetic routines are also easy to incorporate into MCS programs.[11]

8.4.2 Control Variates

Another approach often used to price derivatives is to use a control variate. Suppose we wish to price a derivative A, and there exists some similar derivative B that also has a closed-form solution. (Obviously, A clearly doesn't have an analytic solution, because otherwise we wouldn't need to use MCS in the first place.) The derivative A might be a call option with no analytic solution (e.g., a European call with stochastic volatility, as in Hull and White (1987)[12]), and B might be a vanilla call or a delta hedge. The idea behind the control variate method is that we revise our MCS estimate of the value of A by adjusting for the error in the MCS estimate of B. This works because the MCS errors for both derivatives are similar, and the MCS error for B is known. If f_A^{MCS} and f_B^{MCS} are the MCS estimates of the prices of A and B, and f_B is the true (i.e., analytic) value of B, then our control-variate estimate of f_A is:

$$f_A = f_A^{MCS} - f_B^{MCS} + f_B \tag{8. 20}$$

As with antithetics, we can also interpret the control-variate technique in hedging terms: B, being similar to A, constitutes a hedge, and the control-variate technique adds a zero-value hedge to A. The value of the 'hedged' portfolio is therefore the same as the value of the unhedged portfolio (i.e., A), but the standard deviation of the hedged portfolio is often much less.

The effect of control variates on the error in the MCS of the value of A can also be illustrated in terms of standard statistical theory. If σ_A^2 and σ_B^2 are the variances of the MCS estimates

[10] This explanation is, however, rather loose, and the effectiveness of antithetic sampling can vary considerably on the particular application at hand. For more on the effect of antithetic sampling on variance reduction, see Boyle *et al*. (1997, p. 1273) and Broadie and Glasserman (1998, pp. 184–186).

[11] For some examples, see, e.g., Clewlow and Strickland (1998), pp. 89–91.

[12] Hull and White (1987).

of the values of A and B, then the variance of the control-variate estimate of the value of A, $\sigma^2_{control}$, is:

$$\sigma^2_{control} = \sigma^2_A + \sigma^2_B - 2\rho\sigma_A\sigma_B \tag{8. 21}$$

where ρ is the correlation between A and B. If we now make the convenient simplifying assumption that $\sigma^2_A \approx \sigma^2_B$, then:

$$\sigma^2_{control} \approx 2\sigma^2_A(1 - \rho) \tag{8. 22}$$

This is very revealing, as it tells us that the effectiveness of the control variate method largely depends on the correlation coefficient ρ: the higher the correlation, the lower the variance of the control-variate error. Indeed, a very good control (i.e., one with $\rho \approx 1$) can bring the error variance down to negligible ones – and a bad one (i.e., one with $\rho < 1/2$) can actually increase the variance of our error. [13]

The control-variate technique can also be extended to incorporate gamma and other hedges; however, the effectiveness of Greek-based control variates – delta hedges and so on – can depend on the frequency with which the hedge is assumed to be rebalanced, so we need to take account of this frequency in our simulations.[14]

The effectiveness of this technique depends to a large extent on using control variates that are good for the problem at hand: in particular, we want easily computed controls that are highly correlated with the object we are interested in. The most effective controls are those that take advantage of special features, and a good example is where we might use an Asian option on a geometric mean, an option with a known closed-form solution, as a control for an Asian option on an arithmetic mean. Broadie and Glasserman present results showing that if we wish to price an arithmetic Asian option, using the underlying variable as a control leads to a variance reduction of 66% and using a vanilla call leads to a reduction of 72%; however, using a geometric Asian leads to a reduction of 99%.[15] To some extent, these falls in the variance also reflect the correlations of the controls with the option to be hedged. The effectiveness of the control thus largely depends on the extent to which the control 'matches' the derivative we are interested in. When using control variates to estimate the VaR, we can also use a delta–gamma approximation or a related VaR (e.g., a VaR for a similar instrument for which we have good analytical or algorithmic solutions) as a control variate.[16]

8.4.3 Importance Sampling

Another variance-reduction technique is importance sampling. Loosely speaking, importance sampling adjusts the probabilities to reflect the importance of the paths or events concerned for the purpose at hand. For example, if we wish to simulate the value of an out-of-the-money option, we know that most paths will lead to a zero payoff, and these paths are a waste of

[13] However, the method outlined in the text is not efficient and can be improved. If we replace Equation (8.20) with a more general expression, $f_A = f_A^{MCS} + \beta(f_B - f_B^{MCS})$, where Equation (8.20) is the special case of this expression with $\beta = 1$, then we can easily show that the variance-minimising value of β is $\beta^* = cov(f_A^{MCS}, f_B^{MCS})/\sigma^2_{f_B^{MCS}}$. β^* may or may not be close to 1 in any given instance, and where it is not close to 1, this refinement allows us to achieve notably greater variance reduction than we could achieve using the method in the text. Its estimator $\hat{\beta}^*$ also has the attraction of being a regression coefficient, which means that we can easily estimate it by standard regression (e.g., by regressing $f_A - f_A^{MCS}$ against $f_B - f_B^{MCS}$).

[14] Where we have multiple controls, we can easily estimate the optimal (i.e., variance-minimising) coefficients β_1^*, β_2^*, etc. using an appropriate generalisation of the expression used in the last footnote. This is not only easy to do, but is also very helpful because we might not otherwise have much idea what values these coefficients might take (Boyle *et al*. (1997), p. 1276).

[15] See Broadie and Glasserman (1998), p. 194.

[16] See Cárdenas *et al.* (1999) or Glasserman *et al.* (1999a).

computational effort because they contribute little or nothing to help determine the option value. To get around this problem, we can sample only from paths where the option ends up in the money: if F is the distribution function for the underlying, and p is the probability of the option ending up in the money, we might work with $G = F/p$, which is the distribution function for the underlying conditional on the option ending up in the money. The estimate of the option value is then the average discounted payoff multiplied by p. The trick with importance sampling is to change the probability measure to produce a more efficient estimator.

To estimate VaR using IS, Glasserman *et al.* suggest we first estimate a preliminary high loss value (or VaR), say L.[17] We could base L on a quadratic or delta–gamma approximation. Having obtained L, we would then use an appropriate change of probability measure, and a good candidate would be an exponential change in measure, as this is known to be close to optimal if VaR is close to L. This change in measure ensures that a loss in excess of L is no longer rare, and so leads to a much more accurate estimator of our VaR.

As we have just seen, IS is often used to make rare events less rare, and the efficiency gains from using IS will tend to rise as the 'original' rare event becomes rarer – that is, the benefits from IS increase as our tails become more extreme. These features of IS therefore make it very useful for pricing out-of-the-money options and for estimating VaR and ES, particularly at high confidence levels:

- In the options context, simulation results for down-and-in calls reported by Boyle *et al.* indicate that IS can lead to very substantial variance reductions: they get variance reduction by factors ranging from a minimum of 7 to a maximum of over 1100, and these efficiency gains are particularly impressive when the barriers are well below the current option price (i.e., when the rare event of the option paying off becomes even rarer).[18]
- We can also get very impressive efficiency gains when using IS to estimate VaR. Thus, Glasserman *et al.* report results showing that simple IS applied to a set of stock option portfolios can reduce VaR variances by factors ranging from 6 to 54, and the higher the confidence level, the greater the variance reduction.[19]

8.4.4 Stratified Sampling

Another powerful variance-reduction technique is stratified sampling. Suppose we wish to generate, say, 100 values, $Z_1, Z_2, \ldots, Z_{100}$, to approximate some chosen distribution. If we generate these values using a standard random number generator, then a random sample of independent draws will tend to leave gaps in the approximated distribution and underrepresent the tails. Stratified sampling enables us to avoid these problems. To apply it, we first set:

$$V_j = \frac{j - 1 + U_j}{100}, \; j = 1, \ldots, 100 \tag{8. 23}$$

where $U_1, U_2, \ldots, U_n$ are independently and identically distributed uniform numbers lying in the range [0,1). This transformation ensures that each V_j is distributed between the $(j - 1)$ and jth percentiles (e.g., so V_1 lies in the range [0,0.01), V_2 lies in the range [0.01,0.02), etc.). Since

[17] See Glasserman *et al.* (1999b).
[18] Boyle *et al.* (1997), p. 1287.
[19] Glasserman *et al.* (1999b), p. 355.

these values can also be interpreted as cdf-values for our chosen distribution, we can obtain a stratified sample from our chosen distribution by applying the appropriate inverse distribution function. For example, if our chosen distribution is a standard normal, we apply the standard normal inverse transform:

$$Z_j = \Phi^{-1}(V_j) \tag{8. 24}$$

Our values $Z_1, Z_2, \ldots, Z_n$ are then a stratified sample from a standard normal distribution (i.e., we have a Z-value corresponding to each percentile of the normal distribution, so eliminating gaps and the underrepresentation of tail values).

Stratified sampling is guaranteed to reduce variance, because the variance from a random sample can be decomposed into within-strata variance and across-strata variance, and stratified sampling eliminates the second term.

However, one drawback of stratified sampling is that it destroys the independence of our simulated Z-values, and this makes it more difficult to estimate confidence intervals: because our sample values are no longer independent, we can no longer obtain reliable estimates of confidence intervals by the usual method of estimating the sample standard deviation and then invoking the central limit theorem, which presupposes independence. Instead, we have little practical alternative but to resort to batching: if we have a computation budget of, say, M replications, we might run n independent stratified samples of size M/n, and then estimate the variance of our simulated output variable from the n sample outputs; since *these* are independent, we can then apply the central limit theorem to them and thence estimate confidence intervals in the usual way. This batching procedure enables us to recover our confidence intervals, but at the price of some loss in variance-reduction compared with 'regular' stratified sampling.

Stratified sampling is particularly suited to the estimation of VaR or tail quantiles, because we can target our strata to focus on the tail we are interested in. If we are estimating VaR, there is no point using equal-percentile strata because that would involve generating a lot of central-mass values that would tell us relatively little about our VaR. It would therefore make more sense to choose our strata so that we generate a lot of values around and in the tail and relatively few central ones – or, put differently, we oversample from around the tail, and undersample from the rest of the distribution. Since these simulations would tell us much more about VaR than equal-percentile stratified sampling, we could expect this approach to yield considerably more accurate VaR estimates.

The only question then is how to allocate samples to our strata, and the answer to this question is not always obvious. However, one way to resolve this problem is to run a preliminary simulation exercise and use the results of this exercise to select our strata. We could also use IS to determine sample allocation, or use one or more of the heuristic rules (or rules of thumb) suggested by Glasserman *et al.* (1999b, pp. 353–354, 356), and their results suggest that these have the potential to capture a significant fraction of the potential variance reduction achievable using an optimal allocation of samples to strata. A final possibility is to use some 'learning rule' (e.g., such as a neural network approach) that starts with a given set of strata allocations, and periodically updates these as the simulations continue and interim VaR results come in, so gradually 'fine-tuning' the strata allocations so they provide a better and better focus on points around the VaR.

The theory of stratified sampling also applies to higher dimensions. If we wish to generate a stratified sample from a d-dimensional hypercube with n strata, we can partition the unit

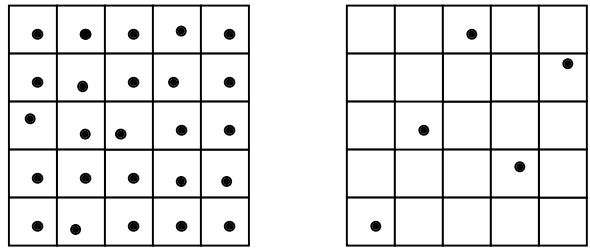

Figure 8.3 Stratified sampling vs a Latin hypercube
Note: The box on the left represents the data points generated using full stratified sampling, with $d = 2$ and $n = 5$, and that on the right represents the points generated with a Latin hypercube.

hypercube into n^d equal-volume cubes with length $1/n$ along each dimension, these being the d-dimensional equivalents of our earlier partition of the unit interval [0,1] into n subdivisions. We then sample from within each hypercube to obtain a stratified sample from the hypercube $[0,1]^d$, and transform them using the appropriate inverse function to obtain the stratified sample corresponding to our chosen distribution function.

Unfortunately, multidimensional stratified sampling rapidly becomes impractical as d gets large, because the number of data points we need to generate, n^d, grows rapidly with n. One solution to this problem is to keep down n, the number of strata, but this is often not particularly desirable because it can mean that we lose many of the benefits of stratification in the first place.

A more promising solution[20] to this problem is to use a Latin hypercube approach, the idea behind which is to select n stratified sample points and then project permutations of these points onto the n^d points in a d-dimensional hypercube. The way this is done is illustrated in Figure 8.3 for $d = 2$. This figure shows two boxes each representing a two-dimensional hypercube, and each of which is divided into $n = 5$ rows and the same number of columns. Full stratification is illustrated in the box on the left-hand side, and requires $n^2 = 25$ random drawings, one for each little box. However, with a Latin hypercube, we choose n random points, subject to the constraint that each row and each column is represented exactly once. This is represented in the second box, where there are only $n = 5$ data points, and each row and each column has exactly one data point.

In both cases, the points within the boxes are selected according to a two-dimensional uniform distribution, and the difference between the two relates to the way in which the hypercubes are selected. To see how the LH approach does this, let U_j^i be identically and independently distributed over the interval [0,1], where $j = 1, \ldots, d$ and $i = 1, \ldots, n$. Now let $\pi_1, \pi_2, \ldots, \pi_d$ be independent random permutations of $\{1, \ldots, n\}$ sampled uniformly from the complete universe of $n!$ possible such permutations. We now set:

$$V_j^i = \frac{\pi_j(i) - 1 + U_j^i}{n} \tag{8. 25}$$

[20] It is, however, not the only alternative solution. We could also use a Brownian bridge method, which generates points in a smaller dimensional space and uses these to generate the most important points along a path. An example might be where we strip down the dimensionality of our problem using principal components analysis, and then apply stratified sampling to a space whose dimensionality is determined by the most important principal components. For more on the Brownian bridge approach, see Broadie and Glasserman (1998), pp. 187–188.

The n points $V^1 = (V_1^1, \ldots, V_d^1), \ldots, V^n = (V_1^n, \ldots, V_d^n)$ make up a Latin hypercube sample of size n in dimension d. Since the calculations involved are fairly easily programmed, the LH approach gives us a very practical means of applying stratified sampling to higher dimensional problems.

8.4.5 Moment Matching

A final variance-reduction method is moment matching (see Barraquand (1995)). The idea behind this is to adjust the simulated data so that one or more of their sample moments matches (i.e., is equal to) the relevant moments from the assumed theoretical distribution. This adjusted sample then has the theoretically 'correct' sample moments, free of sample error, and this is very useful because it can ensure that the underlying is correctly priced. We can regard moment matching as attempting to exploit what we already know, in this case our (assumed) information about the 'true' moments of a distribution. Moment matching can also be regarded as a generalisation of the antithetic variate approach, since the antithetic approach ensures that the sample mean of our 'random' numbers is zero; however, unlike the antithetic approach, the moment-matching approach usually extends this treatment to the second and sometimes higher moments of our sample.

We can apply moment matching in various ways. For example, if our random sample values Z_i are drawn from a normal distribution with mean μ and standard deviation σ, but have sample mean m and sample standard deviation s, we can adjust our simulated values using:

$$\tilde{Z}_i = (Z_i - m)\frac{\sigma}{s} + \mu \tag{8. 26}$$

to ensure that the adjusted values $\tilde{Z}_i$ have (the first two) sample moments equal to their population moments. We would then feed the $\tilde{Z}_i$-values into our underlying stochastic process (e.g., Equation (8.11)) instead of the original Z_i-values. Alternatively, we could match (say, the first two) moments of the sample distribution of the underlying, using something like:

$$\tilde{S}_i(T) = (S_i(T) - m_{S_T})\frac{\sigma_{S_T}}{s_{S_T}} + \mu_{S_T} \tag{8. 27}$$

where m_{S_T} and s_{S_T} are the sample mean and standard deviation, and μ_{S_T} and σ_{S_T} are the assumed 'true' mean and standard deviation. There is, therefore, some ambiguity with moment matching in terms of which variable should be chosen for matching and which features of its distribution should be matched.

Moment matching has both plus and minus points. On the plus side, it ensures that we make effective use of our assumptions about the moments of the variable(s) involved, and so rids us of at least some sample error. Moreover, as Broadie and Glasserman (1998, p. 192) point out, it produces exact prices for any instrument depending linearly and exclusively on the quantities with matched means. This implies, for example, that if zero-coupon bond prices have been matched, then we will get exact prices for coupon bonds and swaps. More generally, we can safely say that the more linear the payoff, the greater the likely error reduction.

On the negative side, all moment-matching methods produce biased estimates with non-linear payoffs, although the bias goes to zero as M gets large. This bias is likely to be small in most financial problems, although there are extreme exceptions, and we can sometimes, though not always, correct for it using standard bias adjustments. The other drawback of these methods is one that they share with stratified sampling: because the adjustments are

applied across all sample paths, they introduce dependence among the adjusted sample values when these would otherwise be independent. This dependence complicates the problem of constructing confidence intervals and, as with stratified sampling, the only practicable way to get confidence intervals may be to apply moment matching to independent batches of runs and then estimate the standard error from the batch means – and this inevitably reduces the efficiency of the moment-matching approach.[21]

Box 8.3 Markov Chain Monte Carlo Methods

Markov chain Monte Carlo methods are very useful methods that can be applied to a great range of problems. They are also useful for risk measurement problems, such as the estimation of multivariate VaR in high dimensions, the estimation of subjective (or Bayesian) VaR, and the estimation of model risk. A Markov chain is a process in which the pdf of future values of a random (possibly vector) variate depends on the last known realised value of the random variate, and not on earlier values as well. An MCMC method is a method that uses a Markov chain to simulate the distribution of a parameter (vector) $\boldsymbol{\theta}$ from some data set $\mathbf{X}$: we simulate a Markov process on the parameter space, and under suitable (weak) regularity conditions, the distribution of simulated parameter values converges to the 'true' (multivariate) distribution $P(\boldsymbol{\theta}|\,\mathbf{X})$. An MCMC method thus enables us to simulate from a multivariate distribution without specifying the distribution itself.

The easiest MCMC method is the Gibbs sampler. To see how this works, suppose that we have two parameters θ_1 and θ_2, and we wish to sample from their MV distribution without knowing what this distribution actually is. All we know are conditional distributions $f_1(\theta_1|\theta_2, \mathbf{X})$ and $f_2(\theta_2|\theta_1, \mathbf{X})$. We now set some arbitrary starting values $\theta_{1,0}$ and $\theta_{2,0}$; we draw a random sample $\theta_{1,1}$ from $f_1(\theta_1|\theta_{2,0}, \mathbf{X})$ and a random sample $\theta_{2,1}$ from $f_2(\theta_2|\theta_{1,0}, \mathbf{X})$; and then we repeat the process again and again, each time drawing a sample $\theta_{1,i}$ from $f_1(\theta_1|\theta_{2,i-1}, \mathbf{X})$ and a sample $\theta_{2,i}$ from $f_2(\theta_2|\theta_{1,i-1}, \mathbf{X})$. This gives us a sample of m, say, parameter pairs, $(\theta_{1,0}, \theta_{2,0}), \ldots, (\theta_{1,m}, \theta_{2,m})$. Provided m is sufficiently large, the distribution of $(\theta_{1,m}, \theta_{2,m})$ will converge under suitable (weak) regularity conditions to the 'true' multivariate distribution $P(\boldsymbol{\theta}|\,\mathbf{X})$. In effect, if m is sufficiently large, the simulated values of $(\theta_{1,m}, \theta_{2,m})$ 'forget' the assumed starting values. So we take a large sample of size m, throw the first n away because of their dependence on the assumed starting values, and end up with a sample of size $m - n$, which we can consider as drawings from the 'true' distribution function $P(\boldsymbol{\theta}|\mathbf{X})$. This approach enables us to decompose a high-dimensional problem into a series of lower-dimensional ones (e.g., we might decompose a d-dimensional problem into d one-dimensional problems). The Gibbs sampler is easy to apply, but in practice care has to be taken to ensure that n is sufficiently high.[22]

[21] One other approach worth mentioning is conditional Monte Carlo simulation (CMCS). This is based on the idea that the conditional variance of a random variable X given some other random variable Y will usually be less, and never more, than the unconditional variance of X: in effect, replacing the unconditional variance by the conditional one reduces the variance because we are doing part of the integration analytically so leaving less for MCS to do. This approach was used by Hull and White (1987) to price options in the presence of stochastic volatility, and Boyle *et al.* (1998, pp. 1288–1290) apply it to price other options. The results reported in the latter study suggest that these methods can produce substantial reductions in variance; however, they also take more calculation time because of the need to do more work analytically, so the optimal degree of conditioning reflects a trade-off between the time spent on simulation calculations and time spent on analytical calculations.

[22] For more on MCMC methods, see, e.g., Tsay (2002, Chapter 10). An application to VaR is given in Wang and Wan (2001).

8.5 ADVANTAGES AND DISADVANTAGES OF MONTE CARLO SIMULATION

8.5.1 Advantages

MCS methods have many advantages:

- They are easy to use once the routines/programs have been set up, and there is plenty of good software available.
- They can easily accommodate more elaborate/sophisticated stochastic processes than GBM ones, unlike many closed-form or analytic approaches.
- They have no problems dealing with multiple risk factors, correlations and fat tails.
- They have no problems dealing with the complexities of path dependency, unlike many analytic alternatives.
- They can be applied to portfolios with heterogeneous and/or complex instruments (e.g., credit derivatives, mortgage-backed securities, ART, etc.).
- They are easy to modify.
- They are capable of considerable refinement to increase accuracy and/or reduce calculation time.
- They give us indications of the accuracy of their results, and estimating confidence intervals is very easy.
- We can increase the accuracy of our results simply by running more trials.
- They are more powerful than finite-difference methods when dealing with moderate or high-dimension problems.

8.5.2 Disadvantages

The main disadvantages of MCS methods are:

- They can be slow because of the number of calculations involved, particularly when we have lots of risk factors.
- MCS approaches are less efficient than lattice and finite-difference methods when we have very low dimensional problems.
- Standard MCS approaches often have difficulty handling early exercise features. The problem here is that because MCS methods work forward in time, they have no easy way of establishing whether instruments should be exercised early. In this respect, methods that work forward in time can be inferior to methods that work back in time, such as binomial trees or finite-difference methods, which (precisely because they work from the terminal date backwards) have no problem checking for possible early exercise. However, methods have been developed that allow Monte Carlo methods to handle early exercise, so these difficulties are not insuperable.[23]

8.6 CONCLUSIONS

MCS approaches a number of major attractions: they are easy to use once set up; they are very flexible and powerful; they can easily handle sophisticated stochastic processes, multiple risk factors and their correlations, and the complexities of path dependency; and so on. On the

[23] For some examples, see Tilley (1993), Andersen (2000), or Longstaff and Schwartz (2001).

other hand, they can also be intensive, both in terms of computing time and in terms of the intellectual/human resources needed to run them; and as noted elsewhere, they have difficulties (though not insuperable ones) handling early exercise features.

Despite these drawbacks, there are good reasons to expect simulation methods to become much more widely used in the future. Simulation methods depend largely on raw computing power, and IT costs have been falling at a rate of 25–30% or more a year for decades, and improvements in computing power have been even more rapid. We can expect these trends to continue into the foreseeable future, and they, in turn, will continue to make simulation methods increasingly fast and more user-friendly. Simulation methods will therefore become more and more attractive – and increasing widely used – as times goes by.

9 Applications of Stochastic Risk Measurement Methods

This chapter follows on from the last and discusses applications of simulation methods to different types of risk estimation problem. Simulation methods are ideally suited to a great range of risk measurement problems, and will often provide the best way of dealing with the ones we are likely to encounter in practice: they are very flexible and powerful, and are particularly good at dealing with complicating factors – such as 'badly behaved' stochastic processes, path dependency, non-linearity, optionality and multidimensionality – that often defeat other approaches. The downsides of simulation methods are that they are less easy to use than some alternatives (unless one uses tailor-made software), require a lot of calculations, and can have difficulty with early exercise features. However, these problems are less severe than they used to be, and will diminish further over time: calculation software is becoming more user-friendly, improvements in computer power mean that large numbers of calculations are less and less of a problem, and (as mentioned at the end of the last chapter) stochastic methods are available to handle early exercise problems.

This chapter looks at some of the more important risk measure estimation problems to which simulation methods are suited. These are the estimation of:

- Risk measures in the face of dynamic portfolio management strategies (i.e., when we relax the assumption that the portfolio is given over the holding period).
- Risk measures of fixed-income positions and of positions in interest-sensitive derivatives.
- Risk measures of portfolios with credit-related risks.
- Risk measures of insurance portfolios.
- Risk measures of pension portfolios.

In each of these cases, the best approach is to construct a tailor-made solution to each specific problem and, where appropriate, make good use of the variance-reduction methods available for that problem.[1]

However, before discussing these, it is a good idea to step back for a moment to look at some general methodological issues first. These relate to the selection of suitable stochastic processes (section 9.1) and methods to reduce the computational burden of multidimensional problems (section 9.2). Sections 9.3–9.8 then discuss the various topics mentioned in the bullet points above, and section 9.9 concludes.

9.1 SELECTING STOCHASTIC PROCESSES

One of the first steps in stochastic simulation is to choose a suitable process. To do that, it is a good idea to establish the stylised facts of the particular empirical process we are trying

[1] The list of topics mentioned here is by no means comprehensive. Some further topics are options risks (covered in the next chapter), energy risks, which are only dealt with tangentially here and real estate or property risks, which are not addressed at all in this book.

to model. Once we have established what these are, we look for a stochastic process that mirrors these stylised facts. In doing so we should keep in mind that a simulated process will be discrete, whereas our assumed 'true' process will be continuous, so we should also take care to avoid unnecessary discretisation errors.

To illustrate these points, suppose that we wish to model a stock price S. The main stylised facts of a stock price process are that stock prices are random, non-negative and drift upwards, and a convenient process that has these characteristics is geometric Brownian motion:

$$\frac{dS}{S} = \mu dt + \sigma dx \tag{9.1}$$

where dx is a Wiener process and μ and σ have their standard interpretations. This process is continuous, but we have to simulate in discrete time. It is therefore tempting to simulate our stock price from a discrete approximation to Equation (9.1), such as:

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \Delta x = \mu \Delta t + \sigma \phi \sqrt{\Delta t} \tag{9.2}$$

However, this procedure produces discretisation errors proportional to the size of the discretisation (i.e., of order Δt). These errors can be avoided if we apply Itô's lemma to Equation (9.1) (as in the previous chapter) and solve for $S(t)$:

$$S(t) = S(0) \exp[(\mu - \sigma^2/2)t + \sigma \phi(t) \sqrt{t}] \tag{9.3}$$

Equation (9.3) avoids discretisation error because it holds exactly, whereas Equation (9.2) is only true in the limit as Δt approaches zero. Naturally, any estimates we get from using Equation (9.3) will still be subject to random sampling error, but such error is unavoidable and can be reduced using some of the variance-reduction methods discussed in the previous chapter.

A simulated path for this process is shown in Figure 9.1. This path shows the characteristics we are seeking to model: the stock price is non-negative, random and has a slight upward drift.

Nonetheless, there are many situations where this process would be too simplistic. For example, it fails to accommodate heavy tails. However, we can build tail heaviness into our process in many different ways: we can apply normality conditionally, we can replace the random process $\phi(t)$ with a fat-tailed process (e.g., a t or a Lévy), we can use a mixture process, etc.

There are also many situations where we want a stochastic process that exhibits mean-reversion. These are particularly important when modelling interest-rate processes, and there are many different ones to choose from: Vasicek, Cox–Ingersoll–Ross (which is discussed further below in section 9.4), Ho–Lee, and so on. We might also want more sophisticated processes. A good example is in the energy field, where we are often dealing with processes that exhibit both strong mean-reversion and very pronounced 'spikes' arising from factors such as unexpected outages and transmission constraints. If we are to model energy prices credibly, we need a price process that incorporates both mean-reversion and a jump process to account for the occasional spikes.

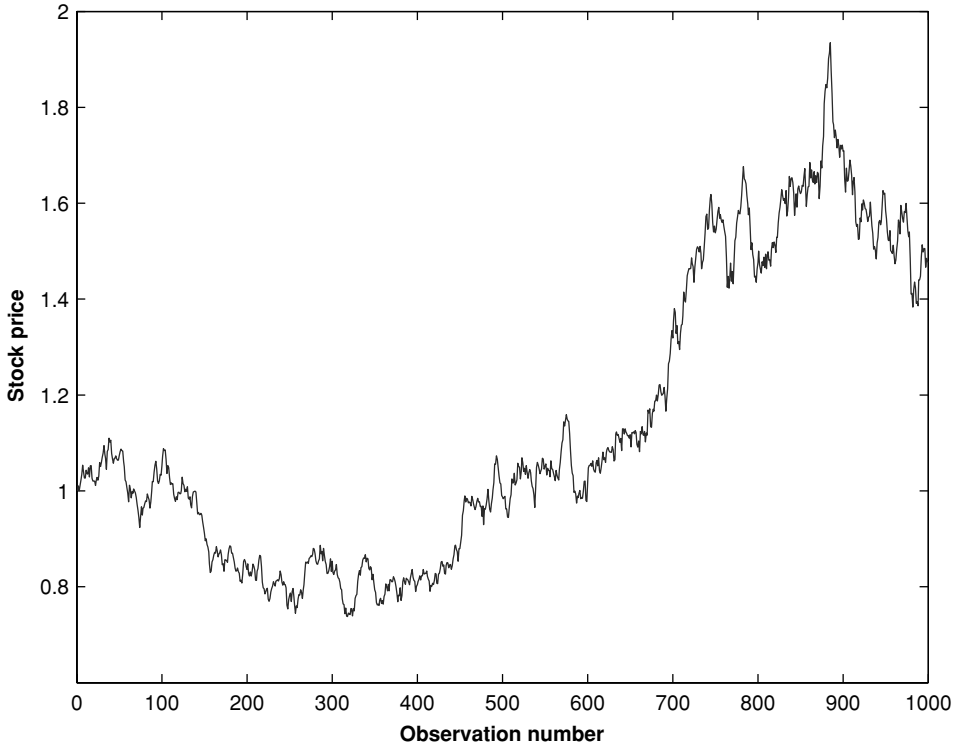

Figure 9.1 A simulated stock price path
Note: Based on assumed annualised values of $\mu = 0.10$ and $\sigma = 0.25$, with observations of daily frequency.

A convenient continuous-time process that incorporates these properties is the arithmetic Ornstein–Uhlenbeck process with jumps:[2]

$$dx = \eta_E(\bar{x} - x)dt + \sigma_E dz_E + dq_E \tag{9.4}$$

where x is the natural log of the spot energy price E, $\bar{x}$ is the (log) long-run equilibrium price to which the (log) spot price x tends to revert, $\eta_E > 0$ indicates the mean-reversion rate, σ_E is the volatility (rate) of the diffusion process, $dz_E = \varepsilon_E(dt)^{0.5}$ is a Wiener diffusion process with a standard normal random variable ε_E, and dq_E is a Poisson process taking the value 0 with probability $(1 - \lambda_E)dt$, and a random value ϕ_E with probability $\lambda_E dt$. λ_E, in turn, is the frequency rate with which jumps occur, and the random jump factor, ϕ_E, is assumed to be normally distributed with mean κ_E and volatility δ_E.[3] The first term on the right-hand side of Equation (9.4) provides for mean-reversion: it shows that, other things being equal, prices

[2] Apart from the properties of mean-reversion, diffusion and jumps, this process also has two other attractive properties – it ensures that the price is always positive, and it permits an exact discretisation. Without these latter properties, we will often get simulated prices that are negative, and the possibility of such 'degenerate' outcomes can make our simulation routines unreliable.

[3] If we wished to, we can incorporate seasonal effects by adding dummy variables to Equation (9.4) or by making volatility or jump parameters dependent on the season. If our model was applied to intra-day (e.g., hourly) periods, we could also do the same for intra-day effects (e.g., to take account of the times of the day when spikes are most likely to occur, etc.). For more on these effects, see, e.g., Blanco and Soronow (2001), pp. 83–84.

have a tendency to rise if $x < \bar{x}$ and to fall if $x > \bar{x}$. The second term is the standard diffusion component, which provides for prices to disperse over time, and the third term gives us our occasional price jumps.[4]

Following Dias (2001), we then take an exact discretisation of Equation (9.4) for simulation purposes. Using obvious notation, our simulation equation is:

$$x_t = x_{t-1}\exp(-\eta_E) + \bar{x}[1 - \exp(-\eta_E)] + \sigma_E\sqrt{[1 - \exp(-2\eta_E)]/(2\eta_E)}N(0, 1) + jumps_x \tag{9.5}$$

We now recover the spot price in the standard way using:

$$E_t = \exp[x_t - 0.5\sigma(x_t)], \quad \text{where } \sigma(x_t) = [1 - \exp(-2\eta_E t)]\frac{\sigma_E^2}{2\eta_E} \tag{9.6}$$

Suitably parameterised, we then use Equations (9.5) and (9.6) to simulate the energy price.[5]

The simulated price is illustrated in Figure 9.2, based on an empirically plausible set of parameter values.[6] This simulated process shows the occasional spikes and strong mean-reversion observed in market energy prices, and the simulated prices – a mean of around \$30 per kWh, with short sharp spikes hitting \$200 or more – are empirically plausible. It is also very different from the GBM process shown in Figure 9.1.

9.2 DEALING WITH MULTIVARIATE STOCHASTIC PROCESSES

When dealing with multivariate stochastic processes, we face the additional problem of having to specify the nature of the dependence between the random processes and we might model this dependence using either correlation- or copula-based approaches.

There are also major issues of computational efficiency to consider, especially in large dimension problems. For example, we might consider Markov chain Monte Carlo methods (discussed in Box 8.3), which help to model large dimension problems using simulations from lower dimension ones.

In this regard, it is often helpful to simulate the principal components of our random variables, rather than the 'original' variables themselves. This is especially the case where we have a large number of possibly highly correlated risk factors, such as a series of points along a spot-rate (or zero-coupon rate) curve. Simulating the principal components can then produce large gains in efficiency, because we have fewer random variables to handle.

9.2.1 Principal Components Simulation

To simulate PCA applied to, say, a fixed-income portfolio, we begin by choosing a vector of n key zero-coupon rates, $[r_1, r_2, \ldots, r_n]$, each with its own different maturity. We then

[4] For more on mean-reversion processes with jumps, and their applications to energy risk problems, see, e.g., Clewlow and Strickland (2000) or Dias (2001).

[5] Strictly speaking, $\sigma(x_t)$ should also incorporate an additional term for the effect of jumps on volatility. However, provided jumps are not too frequent, the error in ignoring the effects of jumps on $\sigma(x_t)$ will be small. Consequently, practitioners can (and, I believe, sometimes do) ignore this term without too much apparent difficulty.

[6] Parameter estimation is discussed elsewhere (see Box 6.1) and is by no means a trivial matter. However, for present purposes, we can take the parameter estimates as given.

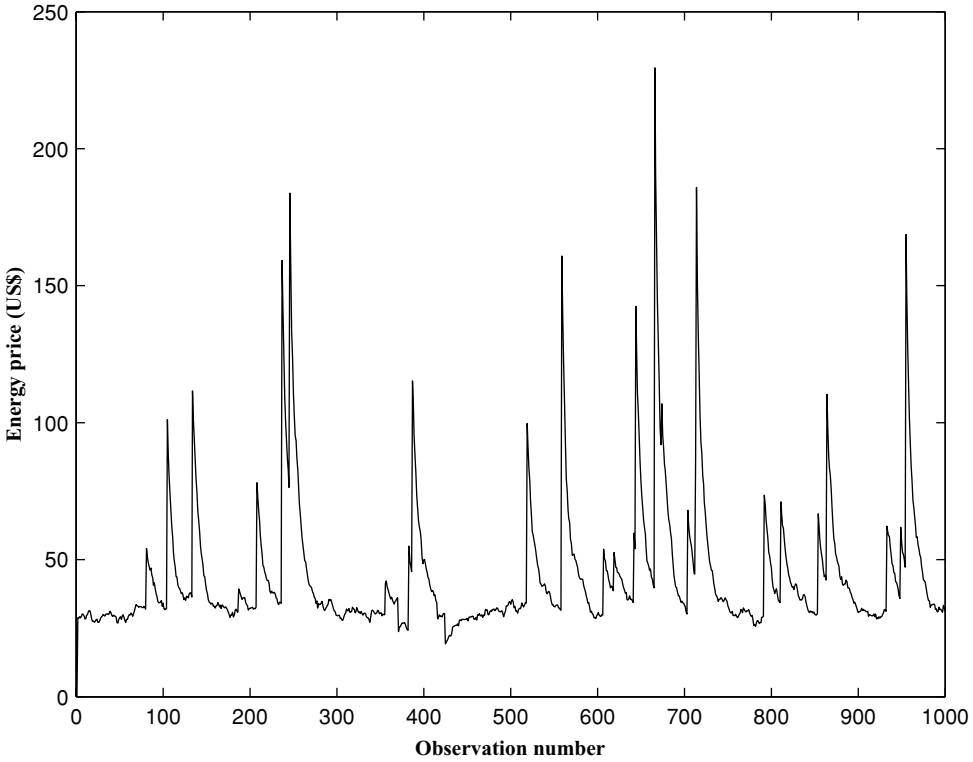

Figure 9.2 A simulated energy price process
Note: Simulated with parameter values using Equations (9.6) and (9.7) with $E_0 = \bar{E} = 30$, $\eta_E = 30$, $\sigma_E = \delta_E = \kappa_E = 0.5$, and $\lambda_E = 10$.

postulate a stochastic process for each rate, such as the simple geometric Brownian motion process suggested by Jamshidian and Zhu (1997):

$$dr_i/r_i = \mu_i(t)dt + \sigma_i dz_i(t) \tag{9.7}$$

where $z_i(t) \sim N(0, \sqrt{t})$ and the $z_i(t)$ would in general be correlated with each other.[7] We will often find that the first three principal components of a spot-rate curve will explain 95% or more of the movement in it, and that these have ready spot-rate interpretations: the first principal component can be interpreted as its shift factor, the second as its slope or twist factor, and the third as its curvature or butterfly factor. These findings suggest that we might often wish to focus on the first three principal components.

We now determine the principal components. If we let $\mathbf{C} = [\rho_{ij}]_{i,j=1,\ldots,n}$ be the $n \times n$ correlation matrix of the n key rates, and let the jth eigenvector of $\mathbf{C}$ be $\boldsymbol{\beta}_j = [\beta_{1j}, \ldots, \beta_{nj}]^T$, then, by the definition of an eigenvector:

$$\mathbf{C}\boldsymbol{\beta}_j = \lambda_j \boldsymbol{\beta}_j, \; j = 1, \ldots, n \tag{9.8}$$

[7] This process is chosen for illustrative purposes and does not allow for mean-reversion in interest rates; however, we can easily adapt it to incorporate mean-reversion by adding a supplementary equation such as the Ornstein–Uhlenbeck process $dz_i = -kz_i(t)dt + dB_i$, where $B_i(t)$ is a geometric Brownian motion.

where λ_j is the jth eigenvalue. We now normalise the $\boldsymbol{\beta}_j$ so that $|\boldsymbol{\beta}_j|^2 = \sum_{i=1}^{n} \beta_{ij}^2 = \lambda_j$, and relabel the λ_i terms so that λ_1 is no less than λ_2, λ_2 is no less than λ_3, and so on. The jth principal component is then β_j.

We can now write dz_i in terms of the principal components:

$$dz_i = \sum_{j=1}^{n} \beta_{ij} dw_j \approx \beta_{i1} dw_1 + \beta_{i2} dw_2 + \beta_{i3} dw_3, \quad i = 1, \ldots, n \tag{9.9}$$

where $E[w_k w_j] = 0$ for $k \neq j$ and $dw_j = \sqrt{dt}$, and assuming that the fourth and further principal components are not particularly significant. Using this approximation, Equation (9.7) becomes:

$$dr_i/r \approx \mu_i(t)dt + \sigma_i \beta_{i1} dw_1 + \sigma_i \beta_{i2} dw_2 + \sigma_i \beta_{i3} dw_3 \tag{9.10}$$

The dimensionality of the problem is reduced from n to 3 and, because the principal components are independent, we now have a correlation matrix with three terms rather than the $n(n+1)/2$ terms we had before.

We would now simulate the principal components and derive our risk estimates: we would run trials for dw_1, dw_2 and dw_3, each trial would (via Equation (9.10)) produce a set of dr_i/r_i values, and we would use these to estimate our risk measures from the distribution of simulated dr_i/r_i values in the usual way.

9.2.2 Scenario Simulation

This PCA procedure can still involve a large number of calculations, but we can cut down even further on the number of calculations using the 'scenario simulation' approach suggested by Jamshidian and Zhu. Each of our three principal components, dw_1, dw_2 and dw_3, is allowed to take one of a limited number of states, and we can determine which state occurs by taking a random drawing from a multinomial distribution. Thus, if a particular principal component is allowed to take one of $m+1$ states, the probability of state i, $P(i)$, is:

$$P(i) = 2^{-m} \frac{m!}{i!(m-i)!}, \quad i = 0, \ldots, m \tag{9.11}$$

So, for example, if we have five states, their corresponding probabilities are:

$$\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16} \tag{9.12}$$

The middle state is more likely than either of the adjacent states, and these in turn are more likely than the extreme states. Hence, the middle state might be 'no change' in the principal component concerned, the adjacent states might be 'moderate up' and 'moderate down' changes, and the other two might be 'extreme up' or 'extreme down' changes. Since the first principal component is more important than the second, and the second more important than the third, it would make sense to allow the first principal component to have more states than the second, and the second more than the third. Jamshidian and Zhu suggest that we might allow seven states for the first component, five for the second, and three for the third. We can then define a spot-rate curve scenario to be a set of states for each of our principal components, and the total number of

scenarios is equal to the product of the number of different states (so, e.g., if the first component has seven different states, the second five and the third three, the total number of scenarios is $7 \times 5 \times 3 = 105$). Moreover, since the states are independent, their probability is equal to the product of the multinomial probabilities in Equation (9.11). In other words, if the first principal component has $m_1 + 1$ possible states, the second has $m_2 + 1$ possible states, and the third has $m_3 + 1$ possible states, then the scenario (i, j, k), in which the first principal component takes state i, the second takes state j, and the third takes state k, occurs with probability:

$$\begin{aligned} &P(component1 = i) \,.\, P(component2 = j) \,.\, P(component3 = k) \\ &= \left[2^{-m_1} \frac{m_1!}{i!(m_1 - i)!}\right] \times \left[2^{-m_2} \frac{m_2!}{j!(m_2 - j)!}\right] \times \left[2^{-m_3} \frac{m_3!}{k!(m_3 - k)!}\right] \end{aligned} \tag{9.13}$$

So, basically, we randomly select scenarios using Equation (9.13), and each scenario gives us a particular spot-rate curve, and a particular P/L on our portfolio. If we take a reasonable number of trials, we should be able to get a good estimate of risk measures from the sample of trial P/L values. Observe, too, that we only ever have to make $(m_1 + 1)(m_2 + 1)(m_3 + 1)$ evaluations (i.e., one for each scenario), regardless of the number of trials. As each trial only involves the selection of a scenario using Equation (9.13), but not its evaluation, it follows that the trials are not computer intensive, and this suggests that even a large number of trials should be calculated quickly. We would therefore expect scenario simulation to perform very well compared to principal components MCS and the results reported by Jamshidian and Zhu seem to confirm this expectation.[8]

The scenario simulation approach also allows us to pay attention to extreme moves. For example, in Jamshidian and Zhu's own example with 105 separate scenarios, some of these occur with a probability of only 0.024%. These low-probability scenarios occur because of the compounding of low-probability states. If we want more extreme scenarios, we can obtain them by increasing the number of states, and so get a better handle on extreme outcomes. Scenario simulation can also be used to conduct 'what if' experiments in which we make certain changes to state outcomes (or state probabilities) and see how they affect the VaR. This can make simulation very useful for stress testing purposes.

Scenario simulation can also be extended to more complex portfolios such as multicurrency portfolios, and portfolios involving credit exposures (e.g., swap portfolios). However, as the number of risk factors increases, we need to keep the number of scenarios down. The answer is to make use of independence assumptions where possible. In the multicurrency context, we might assume that exchange rates and interest rates are uncorrelated, or we might assume that the ith principal components in different countries are correlated, but that the ith principal component in one country is independent of the jth principal component in another, for $i \neq j$. In the credit risk context, we might assume that market and credit risks are independent, or that defaults are independent of each other. Some of these assumptions are more plausible than others, but there is a trade-off between realism and speed, and assumptions such as these help to prevent the number of scenarios from exploding and so enable us to carry large numbers of trials on any given computation budget.

[8] However, the results of Abken (2000) suggest that the performance of scenario simulation can sometimes be erratic and results can be significantly biased compared to those of standard Monte Carlo simulation and principal components approaches, particularly for non-linear portfolios. Convergence can also be slow for some types of portfolio, and Abken recommends that users of scenario simulation should periodically check its results against those of standard simulation or principal components methods.

Box 9.1 Numbers of Scenarios under Monte Carlo Simulation, Principal Components Analysis and Scenario Simulation

One way to appreciate the computational differences between MCS, PCA and scenario simulation is to examine the number of scenarios involved in each. As a benchmark, if we have n risk factors that can take (for the sake of argument) each of d possible outcomes by the end of some period, then the total number of scenarios is d^n. Under classical MCS, n and d would both be relatively large, and the number of scenarios would be almost incalculable. To give an example, n might be 10 rates on a spot-rate curve, in which case the number of scenarios would be d^{10}. Under PCA, we keep n down, so our 10 spot rates might be reduced to, say, three principal components: we then have d^3 scenarios, which will typically be a tiny if not almost infinitesimal fraction of the number of scenarios under classical MCS. However, d would still typically be a large number under PCA, so the number of PCA scenarios is still very large. Under scenario simulation, we keep both n and d down, and so bring the number of scenarios right down: for example, we might have $n = 3$ and $d = 4$, in which case the number of scenarios comes down to $d^n = 4^3 = 64$. Obviously, having a smaller number of scenarios is coarser, other things being equal, but the gains in calculation time are enormous. These gains mean that we can afford to run much larger numbers of trials on any given calculation budget, and this enables us to get more accurate estimates faster in real time.

9.3 DYNAMIC RISKS

We turn now to consider some specific applications of stochastic simulation to the estimation of risk measures. One such application is to the estimation of a risk measure in the presence of a dynamic portfolio strategy. This is a difficult problem to handle using analytical methods, but is a straightforward problem for simulation methods. The method is simple: we specify the dynamic portfolio management strategy, run a simulation exercise taking this strategy into account, obtain a simulated P/L sample, and derive the VaR in the usual way from the simulated sample.

To illustrate the general approach, we consider two alternative dynamic portfolio strategies: a stop-loss strategy, and a filter rule strategy.

A stop-loss strategy is one in which we exit the market and move into something safer as soon as our loss reaches a specified level. In this particular case, we assume that the safer asset is actually risk-free, and that the initial portfolio is invested entirely in a single risky asset whose (arithmetic) return is distributed as a standard lognormal. A stop-loss strategy is contingent on a new parameter, the stop-loss limit, and the impact of this limit is illustrated in Figure 9.3.

The impact of the stop-loss strategy on the VaR depends on the size of the loss limit: if the loss limit is low, the probability of running into the limit will be relatively high, and the maximum likely loss at this confidence level will be the loss limit itself; the higher the limit, the less likely it is to be breached; and if the limit is sufficiently high, the VaR will fall below the loss limit and be close to the static-portfolio VaR we would have obtained with no stop-loss strategy at all.[9] The stop-loss strategy can therefore have a big impact on the VaR, depending on the value of the loss limit.

[9] Naturally, this treatment of loss limits is rather simplistic and makes additional assumptions that we might sometimes be wary of making in practice: most particularly, the assumption that we can 'cut and run' without incurring any further costs, and this (rather optimistically) presupposes that the market for our asset is fully liquid. We will come back to liquidity issues in Chapter 14.

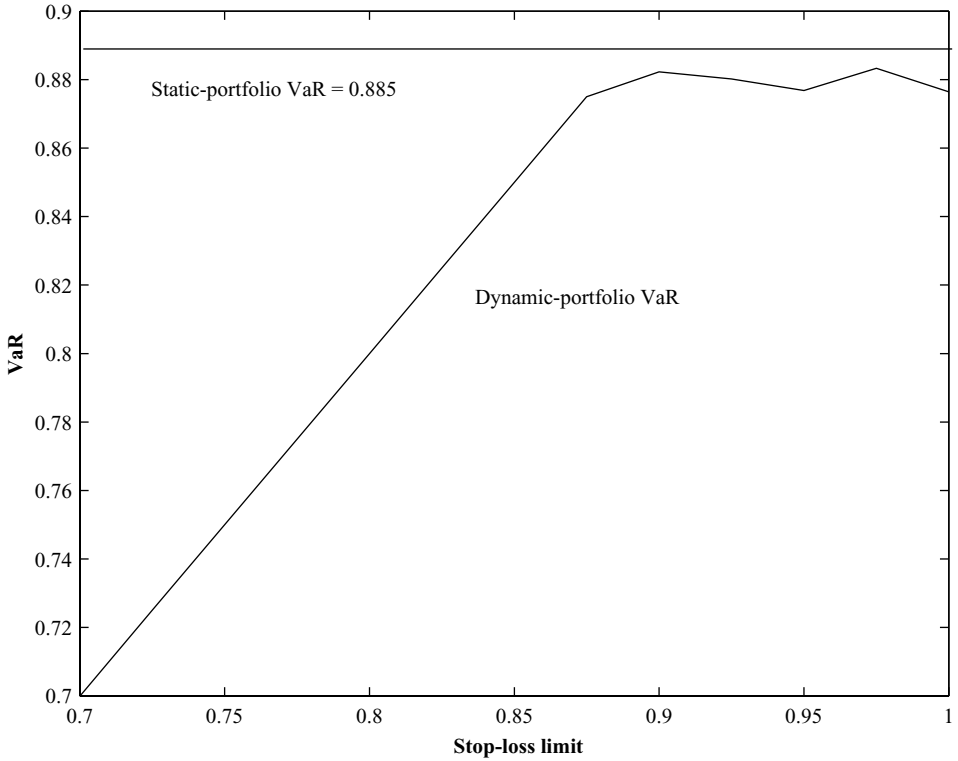

Figure 9.3 VaR with a stop-loss portfolio management strategy
Note: Based on an assumed investment of \$1, annualised arithmetic returns of mean 0 and standard deviation 1, lognormally distributed, 5000 MCS trials, 100 incremental time steps in each trial, and VaR predicated on a confidence level of 0.95 and holding period of 1 day.

A filter rule strategy can be applied where we have two assets, one riskier than the other. The strategy postulates that we increase the proportion of the portfolio held in the more risky asset as the value of our portfolio rises, and we decrease that proportion as the value of our portfolio falls. In this particular case, we again assume that the alternative asset is risk-free, and that the portfolio is initially divided equally between the risky and risk-free assets. We also assume that when the portfolio value rises (or falls), a proportion θ of the increase (or decrease) in the portfolio value is invested in (or disinvested from) holdings of the risky asset. The parameter θ, sometimes known as a participation rate, indicates the degree of responsiveness of the portfolio composition to the change in the value of the portfolio: if θ is low, then the portfolio responds relatively sluggishly to changes in the value of the portfolio, and if θ is high, the portfolio responds more aggressively to such changes. The special case where $\theta = 0$ (i.e., no change) corresponds to the usual static portfolio strategy often assumed in VaR exercises.

The impact of a filter rule strategy on VaR is illustrated in Figure 9.4, which plots an illustrative VaR against varying values for the participation rate. A participation rate of zero (i.e., a static portfolio) gives a VaR of almost 0.45; however, as the participation rate rises, the VaR falls and a participation rate of 1 produces a VaR of little over 0.05. The participation rate therefore has a very considerable, negative, effect on the VaR. The explanation is that a bigger

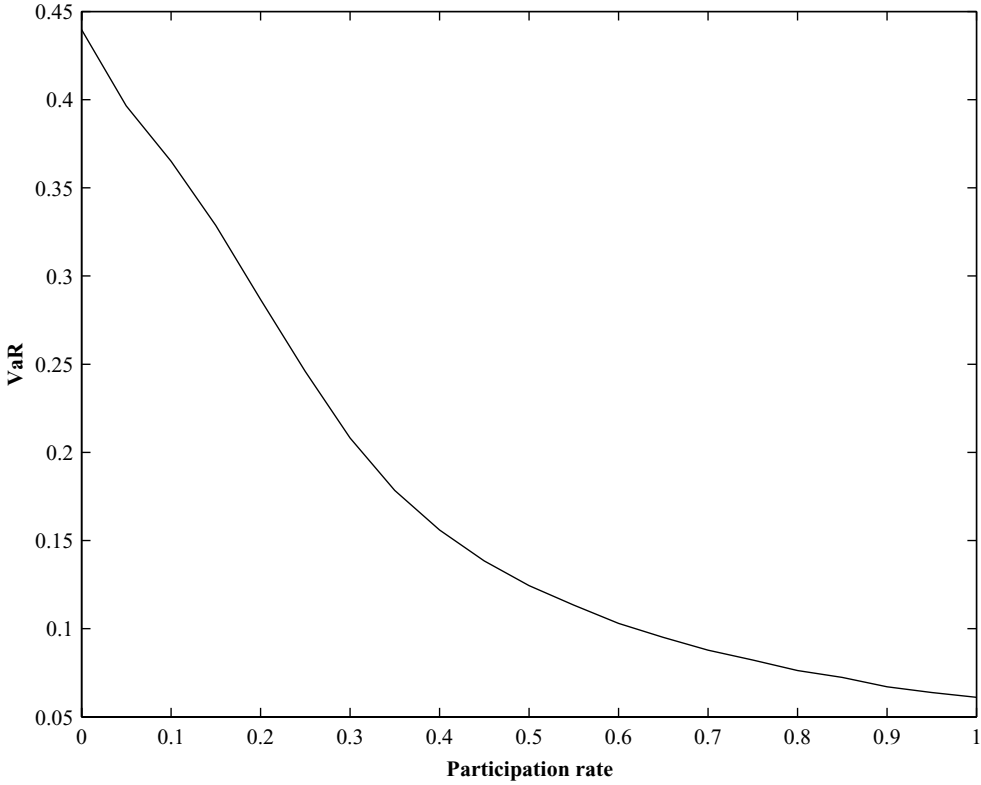

Figure 9.4 VaR with a filter-rule portfolio management strategy
Note: Based on an assumed investment of $1, initially divided equally between two assets, a lognormally distributed asset with annualised arithmetic returns of mean 0 and standard deviation 1, and a zero-mean risk-free asset; 5000 MCS trials; and VaR predicated on a confidence level of 0.95 and holding period of 1 day.

participation rate implies a larger cutback on loss-making positions, and hence a lower loss and lower VaR.

Thus, a dynamic portfolio management strategy can have a very large impact on the VaR, and the impact depends critically on the type of dynamic strategy and the values of the parameters concerned.

9.4 FIXED-INCOME RISKS

Our next topic is the estimation of risk measures of positions in interest-sensitive instruments, usually known as fixed-income instruments: these are bonds, floating rate notes, structured notes, interest-rate derivatives such as interest-rate swaps and futures, and swaptions. Fixed-income problems can be difficult to handle using analytical or algorithmic methods because they usually involve a range of spot rates across the term structure, not just one or two spot interest rates; and fixed-income problems can be particularly difficult to handle where they involve interest-rate options, because of the extra complications of optionality (e.g., the need

to take account of volatilities, etc.). But, fortunately, fixed-income problems are often very amenable to simulation methods.[10]

9.4.1 Distinctive Features of Fixed-income Problems

When estimating fixed-income risk measures, we generally have to take account of two distinctive features of fixed-income positions: the types of stochastic processes governing interest rates, and the term structure.

Interest-rate processes differ from stock price processes in that interest rates are usually taken to be mean-reverting – if interest rates are high, they have a tendency to fall; and if interest rates are low, they have a tendency to rise. In this respect interest rates differ from stock prices which show no mean-reversion: for instance, under the archetypal assumption of a random walk or martingale, the best forecast of a future stock price is today's stock price, regardless of how high (or low) that price might be. One of the most popular interest-rate processes is the Cox–Ingersoll–Ross process:

$$dr = k(\mu - r)dt + \sigma\sqrt{r}dz \tag{9.14}$$

where μ is the long-run mean interest rate, or the level to which interest rates tend to revert, σ is the annualised volatility of interest rates, k indicates the speed with which interest rates tend to revert to their mean values, and dz is a standard normal random variable. This process is popular because it captures the three major stylised facts of empirical interest-rate processes[11] – namely, that interest rates are stochastic, positive and mean-reverting – and because we can simulate from a CIR process exactly and so avoid discretisation errors.

The other distinctive feature of fixed-income positions is more important and harder to deal with. Most fixed-income positions involve payments that are due on multiple future dates, so the valuation of fixed-income instruments requires us to take account of a number of different points along the spot-rate term structure. For example, if we have a coupon-paying bond that pays a coupon every six months, we would want information about the spot rates at six-monthly intervals along the term structure. There are exceptions – most notably those relating to zero coupon bonds, which can be priced with only one spot rate – but the main point is inescapable: in general, we need information about the spot-rate term structure, and not just about one individual spot rate, to be able to price interest-sensitive instruments.

9.4.2 Estimating Fixed-income Risk Measures

To estimate a risk measure for a fixed-income position, we need to simulate the distribution of possible values of a fixed-income portfolio at the end of the holding period. If we are dealing with the simpler fixed-income instruments – such as bonds, floating-rate notes or swaps – then a terminal term structure provides us with enough information to price the instruments and, hence, to value the portfolio at the end of the holding period. The term structure information

[10] There is a very diverse specialist literature on fixed-income VaR, and I haven't space in this book even to begin to cover it properly. However, a useful starting point is D'Vari and Sosa (2000), Niffikeer *et al.* (2000) and Vlaar (2000). For those interested in the VaR of mortgage-backed securities, a good reference is Jakobsen (1996).

[11] However, none of the simpler models of interest-rate processes can fully capture the dynamics of interest rates. We tend to find that after a model has been fitted to the data, the goodness of the fit tends to deteriorate over time, and this makes the use of such models for pricing purposes somewhat problematic unless they are frequently recalibrated. For more on these issues and how to deal with them, see, e.g., James and Webber (2000).

is sufficient because we can value these instruments using standard pricing methods based on present-value calculations of the remaining payments due.

However, if we are dealing with interest-rate options – such as interest-rate calls, puts, floors, ceilings, etc. – then information *only* about the term structure of spot rates will *not* be enough to value our instruments. To value positions involving options, we need information about terminal volatilities as well, and where the options are interest-rate options, we generally need information about the terminal volatility term structure (i.e., we want the terminal volatilities associated with each of a number of spot rates). Moreover, in these circumstances, we can no longer price our instruments using simple present-value methods, but need to resort to an appropriate option pricing model.

There is a very large literature on such models, and the models themselves vary a great deal in their sophistication and flexibility. Fortunately, given the availability these days of good software, it is no longer difficult to implement some of the better models. The main problem then remaining is how to obtain the terminal volatilities, and we can forecast these volatilities either by assuming that the terminal volatilities are equal to the currently prevailing volatilities (which is permissible if we can assume that volatilities are random walks or martingales, which are best predicted by their current values) or by using GARCH or similar approaches.

Since accuracy and calculation time are often significant issues in fixed-income problems, it makes sense to look for ways of improving accuracy and/or reducing computation time. We should therefore explore ways to use variance-reduction methods. We might also simulate principal components of the spot-rate curve instead of the spot rates themselves. For example, we might use scenario simulation. If we follow Jamshidian and Zhu (see section 9.2.2 above) and have three principal components, with seven possible states for the first principal component, five for the second and three for the third, then we would have 105 distinct principal component scenarios. This would also give us 105 spot-rate (or term-structure) scenarios, and the only simulation required would be to pick a random scenario from the multinomial scenario-selection process. It should then be possible to obtain results quickly in real time.

Box 9.2 Estimating the VaR and ES of Coupon Bonds

Suppose we wish to estimate the VaR and ES at the 95% confidence level of a \$1 position in a coupon-paying bond with 10 years to maturity, with a coupon rate of 5%, over a holding period of 1 year. These coupon payments should be accounted for in our P/L, and one simple way of doing so is to assume that coupon payments are reinvested at the going spot rate until the end of the holding period. We assume for convenience that the term structure is flat at 5%, so all spot rates are 5%. We also assume that the spot interest-rate process is a CIR process like that given in Equation (9.14), with $\mu = 0.05$, $k = 0.01$, and $\sigma = 0.05$, and we discretise this process into $N = 10$ steps, so $dt = 1/10$ measured in years, and assume a number of trials, M, equal to 1000. Given these parameters, our estimated VaR and ES turn out to be about 0.011 and 0.014 respectively. The coding strategy is set out in the MMR function 'bondvares'.

9.5 CREDIT-RELATED RISKS

Simulation approaches are also well suited for estimating the risk measures of positions with credit-related risks. Credit-related or default risks arise in a huge variety of contexts, ranging from portfolios of simple default-risky bonds, at one extreme, to portfolios of sophisticated,

credit-sensitive, interest-rate derivatives, including credit derivatives, at the other.[12] Credit-related risks involve at least three possible complicating factors:

- The first is the need to model the default process. Since a default event is a binary variable (i.e., it takes a value 1 if default occurs and 0 if it does not), this means that we are dealing with a risk factor, or set of risk factors, that is very non-normal. The P/L associated with default losses is therefore also non-normal, except in unusual cases where we have a lot of 'small' independent default risks that allow us to appeal to the central limit theorem.
- The second complicating factor is the need to take account of how the risk of default affects the prices of the instruments concerned. Market risk is not independent of credit risk, and the price difference (or, if we prefer, the yield–spread difference) between default-risky and default-free bonds will depend on the default risk in a complex way. We might also want to take account of the way in which the default probability might change over time (i.e., we should account for its transition process). In addition we should take account of the recovery rate and how that might change over time as well. Furthermore, since default probabilities and recovery rates depend on other factors (e.g., the stage of the business cycle, sector-specific changes, etc.), we might also want to model default probabilities and recovery rates as dependent random variables, and take account of the factors on which they depend.
- With many credit-related positions, there are also various institutional or contractual factors to consider, and these can have an important bearing on the actual losses we might suffer in the event of default. These factors include netting agreements, collateral requirements, credit triggers, recouponing features, credit guarantee arrangements, and mutual termination options. These 'credit enhancement' features help institutions manage their credit risks, but also make risks more complicated, and in the process complicate the estimation of market risks and liquidity risks as well.

As we might expect, these complicating features make credit-related problems amenable to simulation methods in ways that other methods are often not.

As with many other simulation problems with complex problem-specific features, the strategy is to build a model that is tailor-made to the problem. To illustrate the basic issues, consider the simple problem of estimating the VaR of a coupon bond with default risk. To make this concrete, suppose we have a bond with one year to run, with coupon payments due in six and 12 months. The coupon rate is c, and bonds of this credit rating are assumed for ease of illustration to have a flat spot-rate curve with a universal spot rate r. If the issuer defaults, the holder is assumed to recover a fraction δ of the current market value of the bond, and the probability of default is p. If the issuer defaults in six months' time, the holder obtains the recovery value and is assumed to invest it at the risk-free rate until the end of the holding period, which is assumed to coincide with the maturity period of the bond. We now write a program to estimate the VaR of our bond portfolio, and the details can be found in the function 'defaultriskybondvar'. This function estimates the VaR once we specify the values of the parameters concerned: for example, if $r = c = 0.05$, $r_f = 0$, $\sigma = 0.25$, $p = 0.05$, $\delta = 0.5$, $hp = 360$, $\alpha = 0.95$, and an amount invested of \$1, then a 1000-trial MCS exercise gives us a VaR estimate of 0.452.

[12] For obvious reasons, I can't do justice here to the very large literature on credit-risk measurement: the Merton model, the KMV model, credit migration models (e.g., CreditMetrics), latent variable and copula models, etc. However, there are good treatments of credit-risk measurement around, and for those who wish to pursue some of these issues, I recommend Crouhy *et al.* (2001, Chapters 7–11).

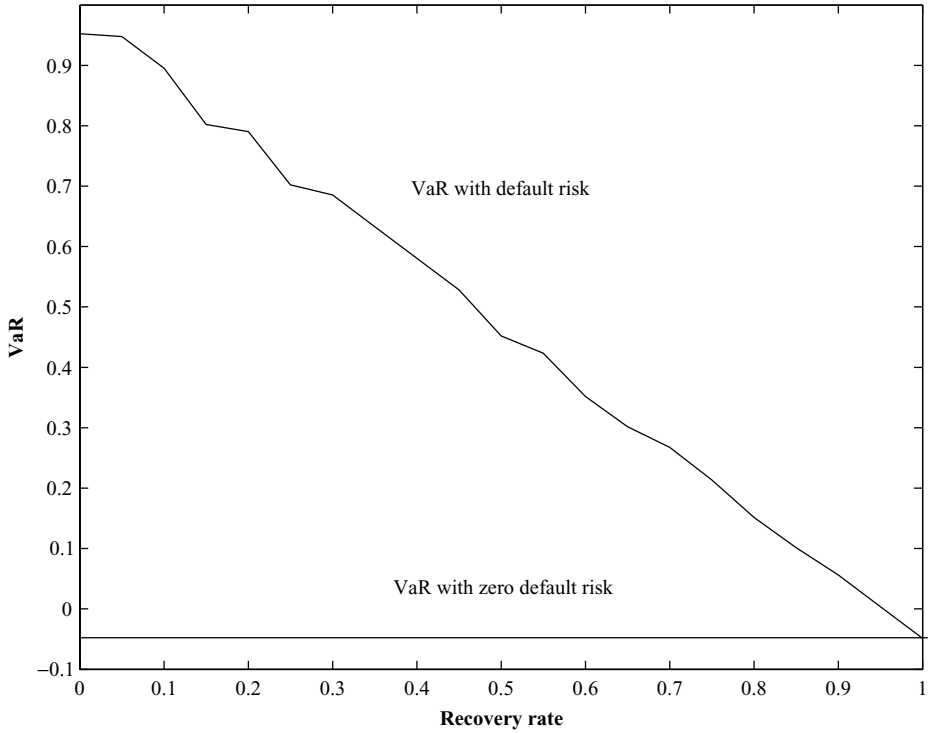

Figure 9.5 VaR and the recovery rate for a default-risky bond
Note: Obtained using the 'defaultriskybondvar' function with $r = c = 0.05$, $r_f = 0$, $\sigma = 0.25$, $p = 0.05$, $\delta = 0.5$, $hp = 360$, $\alpha = 0.95$, an amount invested of \$1, and 1000 MCS trials.

This type of methodology is also useful for shedding light on how our risks vary with underlying parameters. For example, a key parameter for a default-risky bond is the recovery rate: the higher the recovery rate, the less we lose in the event of default and the lower the VaR. We can examine the impact of the recovery rate by plotting the VaR against a range of recovery rates, as in Figure 9.5. This plot shows that the recovery rate has a big impact on the VaR, and confirms that the VaR falls as the recovery rate improves. Observe, too, that as the recovery rate approaches 1, the default-risky VaR approaches the VaR we would get with a default-free bond (which is as we should expect).[13]

9.6 INSURANCE RISKS

There are also many applications of simulation risk estimation methods in insurance. These problems fall under three broad headings:

- Underwriting risk measurement problems: these are risk measurement problems associated with the 'core' business of insurance companies, which is buying or selling insurance. These would include the risks of losses on insurance portfolios – and we will give a couple

[13] We can also perform similar exercises for the other parameters – the risk of default p, the term to maturity, the holding period, and so on. Obviously, with more complicated instruments, there will also be more parameters or other distinguishing features to be considered.

of examples presently – but also include the calculation of reserve or capital requirements (e.g., reserves against losses that are reported and adjusted, reserves against losses that are reported but not adjusted, etc.).

- Investment risk measurement problems: these cover the risks associated with the investment of insurance companies' financial assets. In theory, these risks are qualitatively similar to the investment risks faced by other financial institutions, although insurance companies often operate to longer horizon periods and traditionally have a preponderance of fixed-income investments (which can also be subject to some limited default risk).
- Enterprise-wide insurance risk measurement problems: these cover risks such as those associated with asset-liability management and a firm's overall risk exposure, and are closely related to risks associated with the securitisation of insurance company risks, such as those associated with alternative risk transfer (ART) and integrated risk management products.

9.6.1 General Insurance Risks

To illustrate the potential of simulation methods to estimate the risks of general insurance portfolios (or property and casualty portfolios, as they are known in the US), suppose we wish to estimate the VaR of a general insurance portfolio pool consisting of a large number of identical contracts (e.g., a pool of motor insurance contracts). To be specific, let us assume:

- There are n contracts, each paying a fixed premium to the insurance company, and these contracts are priced so the insurance company expects to make a target profit of θ on each contract in the portfolio.
- The insurable loss events are independent of each other, and each occurs with probability p.
- When any of these events occur, the associated loss L is determined by a drawing from a lognormal distribution, and $\log L$ has mean μ and standard deviation σ.
- Each contract has a deductible D, and we assume that there are no 'no claims' bonuses or similar inducements to encourage policyholders not to recover losses from the insurance company. Hence, when a loss occurs, the contract holder will bear the whole loss if $L \leq D$ and will bear a loss of D if $L > D$; and the insurance company will bear a loss of 0 if $L \leq D$ and a loss of $L - D$ if $L > D$.

Obviously, many real-world insurance contracts will be much more complicated and insurance portfolios would also be more heterogeneous, but this simple example illustrates the basic ideas.

We can solve this problem by writing a suitable program to estimate the VaR, the details of which can be found in the MMR Toolbox function 'insurancevar'. The gist of the program is that in each trial, we take a random drawing from a binomial distribution for each contract to determine if it incurs a loss; for each loss event, we also take a random drawing from a lognormal to determine the size of the loss. We then adjust the losses for the deductible and pricing policies, and determine the VaR and ES from the sample of adjusted losses. A parameterised example is given in Figure 9.6, which shows the histogram of simulated loss values and the associated VaR and ES estimates – in this case, the loss histogram has a long right-hand tail (which is characteristic of the lognormal) and the VaR and ES are respectively equal to 17.786 and 24.601.

We can also adapt this approach for related purposes: we can change parameters and examine the resulting impact on VaR, and so on.

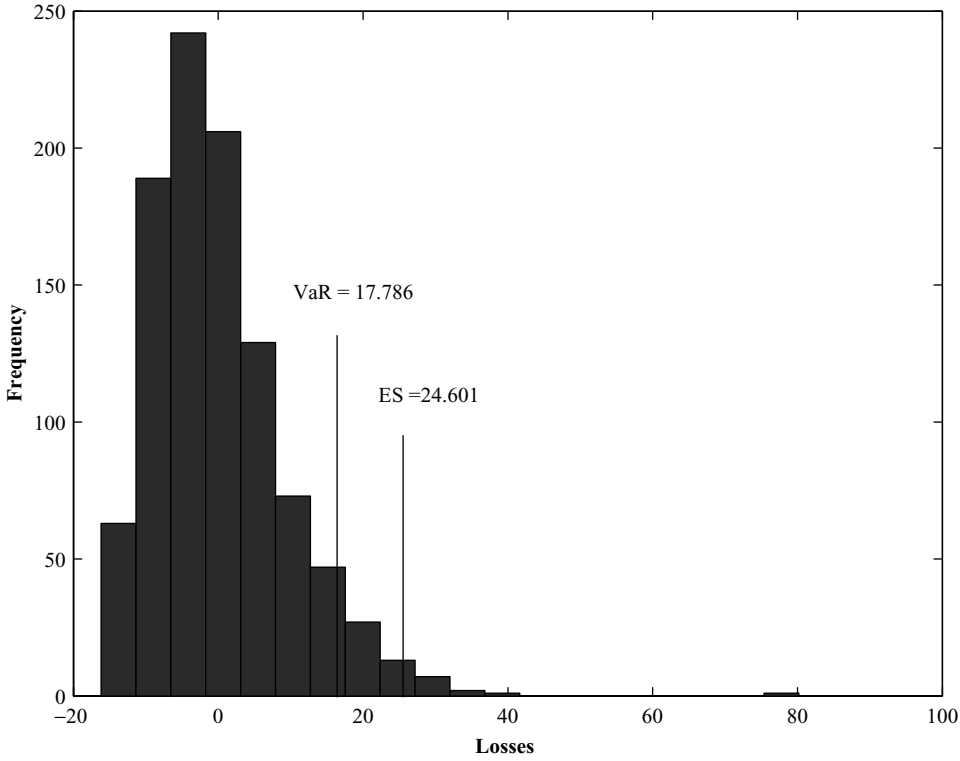

Figure 9.6 General insurance risks
Note: Obtained using the 'insurancevares' function with $\mu = 0$, $\sigma = 1$, $p = 0.2$, $n = 100$, $\theta = 0.1$, $D = 1$, 1000 trials, and $\alpha = 0.95$.

9.6.2 Life Insurance Risks

There are also many applications of stochastic risk estimation methods in life insurance. These are the risks associated with annuity and life insurance portfolios, and the underlying risks are aggregate mortality risks.

To give an example, suppose an insurer has sold a large number (n, equal to 10 000) of annuities to a group of ostensibly similar 65-year-old individuals, each of which involves a commitment to pay \$1 a year to each annuitant for the rest of their lives. We now let S_t be the number of annuitants who survive to t, where $t = 0, 1, 2, \ldots$. Hence, $S_0 = n = 10\,000$, and S_t gradually falls over time as annuitants die off. S_t is also equal to the payments to be made in period t. The individual mortality process is binomial and each individual survives to t with probability p_t, but for large n, S_t will be approximately normal with mean np_t and variance $np_t(1 - p_t)$. We now need to model the p_t process, and one possibility is to assume that p_t evolves as follows:

$$p_t = (p_t^*)^{1-\varepsilon_t} \tag{9.15}$$

This tells us that initial anticipations of survival probabilities, p_t^*, are subject to random shocks ε_t, and we want these to fall in the range $[-1,+1]$. If $\varepsilon_t > 0$, mortality unexpectedly improves,

and $\varepsilon_t < 0$ mortality unexpectedly diminishes. We also assume that ε_t obeys a transformed beta distribution. The beta is a very flexible distribution; it has two parameters, a and b, the values of which we can set to achieve the particular type or shape of distribution we desire. However, a beta-distributed random variable is constrained to the range [0,1], and this range is inappropriate for our purposes because we want to allow for the possibility of unexpected mortality deterioration ($\varepsilon_t < 0$). To get around this constraint, we take a beta-distributed random variable, y_t say, and take ε_t to be:

$$\varepsilon_t = 2y_t - 1 \tag{9.16}$$

This transformation expands the range of the random shock to [−1,+1] as desired, and the general shape of the ε_t distribution reflects that of the underlying beta distribution.

Assuming a set of values for the parameters ($a = b = 10$) in this case, we can now simulate the insurer's risk exposure. This exposure is illustrated in Figure 9.7. This figure plots the means, 5-percentiles and 95-percentiles of the distributions of the insurer's prospective

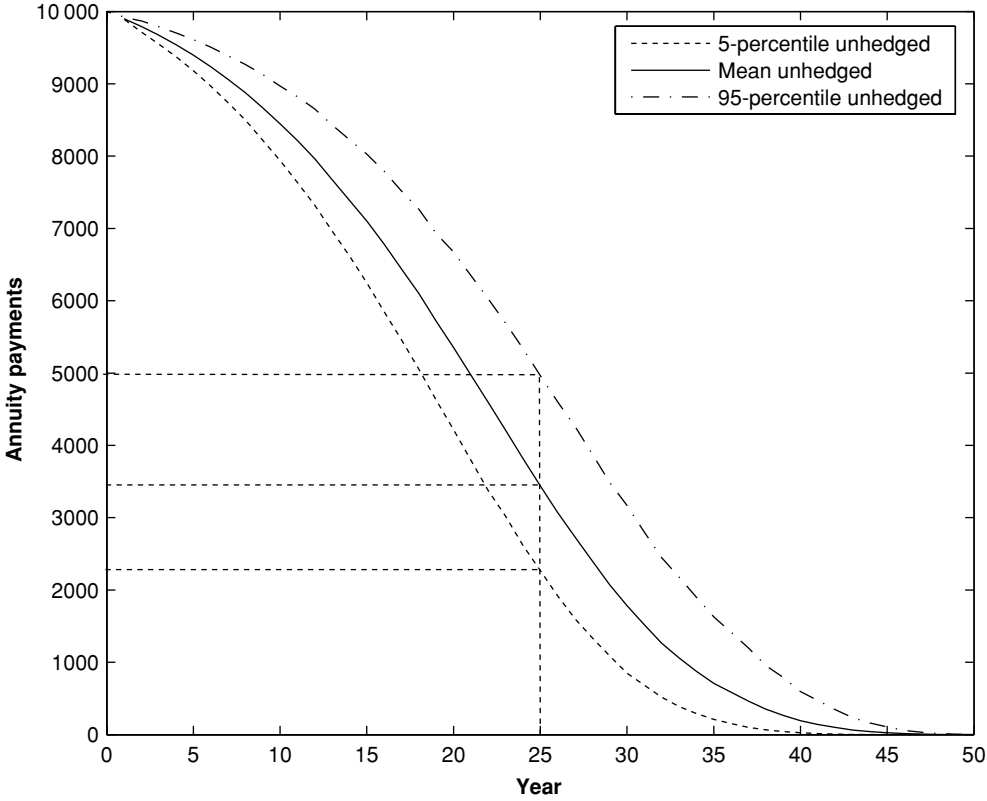

Figure 9.7 Random payments on an annuity portfolio
Notes: Based on projected cumulative survival probabilities taken from the 1996 US Annuity 2000 Basic Mortality Table, base age = 65 10 000 Monte Carlo simulation trials using Equations (9.15)–(9.16) and illustrative assumptions $a = b = 10$. The assumptions about a and b imply that the estimated mean and standard deviation of ε_t are ≈0 and 0.218.

payments for each of the next 50 years. The means are the payments that can be expected given the assumptions made, and the differences between the percentiles give us the payments' 90% confidence intervals. For example, if we take $t = 25$, we see that expected payments are \$3442, and the 90% confidence interval is the range [\$2268,\$4966]. This wide range indicates the extent of the uncertainty faced by the insurer. We also see that the earlier and very distant confidence intervals tend to be narrow, but the intervals in the middle are quite wide. This tells us that most of the uncertainty attaches to payments in the middle period, and is what we would expect bearing in mind that the earlier survival rates can be predicted reasonably confidently, whereas the most distant ones, although subject to the greatest uncertainty, will nevertheless be very low.

9.7 MEASURING PENSIONS RISKS

Some final applications of simulation approaches are to the measurement of pensions risks. These are complicated problems, and the general method is to build a model that allows the pension fund to grow in line with pension fund contributions and the (risky) returns made on past pension fund investment; if the model is sophisticated, it would also allow for the effects of the pension fund portfolio management strategy, which might also be dynamic, and for the possibility of interrupted contributions (e.g., due to the holder of the pension scheme being unemployed). When the holder retires, one of two things might happen, depending on whether the pension scheme is a defined-benefit (DB) scheme or a defined-contribution (DC) scheme:

- With a DB scheme, the holder gets a predefined retirement income, usually specified in terms of a certain proportion of their final salary or the average of their last few years' salary, plus possible add-ons such as inflation adjustment. In these schemes, the pension plan provider bears a considerable amount of risk, because the holder's pension benefits are defined and yet their (and, where applicable, their employers') contributions might not cover the provider's pension liability. In such cases, the pension risk we are interested in is related to the probability and magnitude of the holder's accumulated pension fund falling short of the amount needed to meet the costs of the holder's defined pension benefits. Pension risk in this context is borne by the pension provider, and involves the risk of the pension provider's assets falling short of their liabilities.
- With a DC scheme, the holder gets a pension based on the value of its accumulated fund and the terms on which the fund is converted to an annual retirement income.[14] Usually, the accumulated fund would be converted into an annual retirement income by purchasing an annuity, in which case the retirement income depends not just on the value of the accumulated fund, but also on the going interest rate and the holder's expected mortality. With these schemes, the holder bears the pension risk, and the risk we are interested in is the risk of a low pension relative to some benchmark (e.g., such as final salary). Pensions risk in this context is the risk of the holder having a low pension.

We now consider each of these two types of scheme in turn.

[14] The reader who is interested in the mechanics and simulation of DC pension schemes might wish to explore the PensionMetrics work of Blake *et al.* (2001).

9.7.1 Estimating Risks of Defined-benefit Pension Plans

To estimate the risks of a defined-benefit pension scheme, we should first clarify the precise features of the scheme concerned, and it is helpful to do so by focusing on the pension provider's assets and liabilities.

On the asset side, we need to make assumptions about the start and end dates of the scheme, the amounts contributed to the pension fund, the unemployment (or other contribution interruption) risk, and the way in which the pension fund is invested. To illustrate, we might assume that the pension plan holder:[15]

- Starts contributing at age 25 and aims to retire at age 65.
- Has a salary of $25 000 at age 25, contributes 20% of gross income to the fund, and expects salary to grow at 2% a year in real terms until retirement.
- Faces a constant unemployment risk of 5% a year, and contributes nothing to the scheme when unemployed.
- Does not benefit from any employer contributions to their pension fund.

We also assume that for convenience the fund is invested entirely in one risky asset (e.g., equities), and we assume that the return to this asset is normally distributed with annualised mean and standard deviation both equal to 0.1.

On the liability side, we can think of the pension provider as being obliged to purchase an annuity when the holder retires to give him/her a specified retirement income, and the provider's liability is the cost of this annuity plus the cost of any add-ons such as guarantees against inflation. The provider's liability therefore depends on the final salary, the formula used to determine retirement income, the annuity rate, and the holder's mortality prospects conditional on reaching retirement. It follows, then, that in general the provider's liability is subject to (at least) five (!) different sources of risk:

- Risk arising from uncertainty about the holder's salary at or near retirement.
- Risk arising from uncertainty about the timing of retirement.
- Risk arising from uncertainty about employment, which creates uncertainty about the number of years the holder will contribute into the scheme.
- Risk arising from uncertainty about the interest rate at retirement. This is important, because the interest rate determines the cost of the annuity necessary to provide a given retirement income: the lower the interest rate, the greater the cost of the annuity, other things being equal.
- Risk arising from uncertainty about the holder's mortality prospects on retirement, which is important because of its effect on the cost of the annuity.

To make the analysis as transparent as possible, we now make the following illustrative assumptions:

- The pension is equal to the final salary times the proportion of years in which the holder has contributed to the fund. This implies that if the holder has worked and contributed throughout his/her working life, then he/she will get a pension equal to their final salary.

[15] These assumptions are unrealistically precise, but we need to make these or similar assumptions to estimate the risks involved, and these rather simple assumptions help to illustrate the approach. In practice, we might want to modify them in many ways: we might want to use a variety of alternative assumptions about start/end dates or contribution rates, real income growth rates, and so on. We might also wish to make real income growth stochastic, allow a wage profile that peaks before retirement age and possibly depends on the holder's profession, allow unemployment risk to vary with age or profession, and allow for more realistic investment strategies (e.g., a mix of bonds and equities, etc.)

- The annuity rate on retirement is taken as 4%.[16]
- There are no add-ons, so the provider's liability is only the cost of the annuity.

Having established the structure of our model, we now program it, and one way to do so is provided by the function 'dbpensionvar'. Leaving aside details, the programming strategy is to model the asset and liability sides separately, work out the terminal values of each under each simulation trial, and take the pension provider's P/L to be the difference between the two. Once we have a set of simulated sample P/L values, we can then estimate VaR or ES in the usual way. With the asset side, we build up the value of the pension fund over time, bearing in mind that the fund is equal to current contributions plus the last period's fund value plus the return earned on last period's fund value; and pension fund contributions depend on whether the holder is working that period. We therefore need to take random drawings to determine the rate of return earned on previous pension fund investments and to determine whether the holder is working in the current period. On the liability side, we determine the final salary value and the number of contribution years, use these to determine the pension entitlement, and then apply the annuity rate to obtain the value of the pension annuity.

After all of this, we now choose our confidence level and run the program to estimate our VaR. Given the assumptions made, our 95% VaR turns out to be $259 350. This is shown in Figure 9.8, which also shows the provider's simulated L/P. Perhaps the most striking feature of this figure is the very wide dispersion in the L/P series: the provider can get outcomes varying from a huge (i.e., $400k) profit to an even bigger loss (i.e., over $1.5m). The business of providing DB pensions is clearly a very risky one. The other point that stands out is that the VaR itself is quite high and this in part is a reflection of the high risks involved. However, the high VaR is also partly illusory, because these prospective outcomes are 40 years off in the future, and should therefore be discounted to obtain their net present values. If we then discount the future VaR figure at, say, a 5% annual discount rate, we obtain a 'true' VaR of about $36 310 – which is about 14% of the amount shown in the figure. When dealing with outcomes so far off into the future, it is important to discount future values and work with their present-value equivalents.

9.7.2 Estimating Risks of Defined-contribution Pension Plans

Defined-contribution pension schemes share much the same asset side as DB schemes; however, they differ in not having any distinct liability structure: so instead of matching assets against specified liabilities, they (usually) convert the assets available into an annuity to provide the pension.[17] This implies that the pension is determined by the size of the fund built up, and by the terms on which that fund can be annuitised.

To model a DC scheme, we could have much the same asset structure as the earlier DB scheme: hence, we can assume that our pension plan holder starts contributing at age 25, aims to retire at 65, has a starting salary of $25 000, and so on. Once he/she reaches retirement age, we have to make some assumption about how the accumulated pension fund is used. One possibility is to assume that the accumulated fund is converted into an annuity at the going rate, in which case the pension income obtained will also depend on the prevailing annuity rate and

[16] We therefore assume away uncertainty associated with both the annuity rate and mortality prospects. In practical applications we would certainly want to make the annuity rate stochastic, but would also want to ensure that the rate used was positive and consistent with other contemporaneous rates. The correct way to treat the annuity rate is discussed in Blake *et al.* (2001).

[17] The practice of annuitising funds on retirement is, however, not necessarily the best way to convert the fund into a pension. These issues are explored in more detail in Blake *et al.* (2003) and Milevsky (1998).

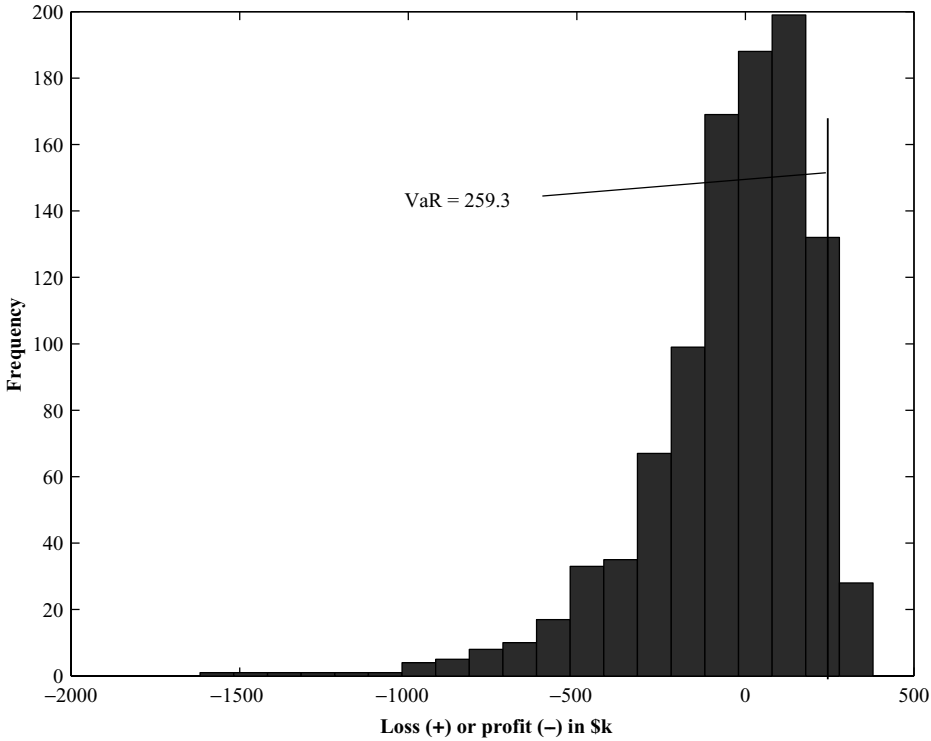

Figure 9.8 VaR of a defined-benefit pension scheme
Note: Obtained using the 'dbpensionvar' function with initial income = \$25k, income growth rate of 0.02, conditional life expectancy = 80, contribution rate = 0.15, $\mu = 0.1$, $\sigma = 0.1$, $p = 0.05$, annuity rate = 0.04, 1000 trials, and $\alpha = 0.95$.

the holder's mortality prospects.[18] To complete the model, we therefore also need assumptions about the annuity rate, and we may as well make the same assumptions about these as we did before. However, with DC schemes, the notion of pensions risk refers to the value of the pension itself, not to a possible pension fund shortfall, and it is convenient to express this risk in terms of the pension divided by the final salary.

We now program the model, and a program is provided by the MMR function 'dcpensionvar'. The programming strategy is to model the terminal value of the pension fund under each simulation trial, annuitise each trial fund value at the going rate, and then divide these by the final salary to obtain a set of normalised pension values. The distribution of these pension values then gives us an indication of our pension risks.

If we carry out these calculations, we find that that the pension ratio has a sample mean of 0.983 and a sample standard deviation of 0.406 – which should indicate that DC schemes can be very risky, even without looking at any VaR analysis. The pension VaR – the likely worst

[18] As viewed from the time when the scheme is first set up, this means that the pension is subject to a number of sources of risk. Leaving aside retirement-age and contribution-rate risks, there are risks arising from the returns earned on pension fund investments, risks arising from the possibility of unemployment and interrupted contributions, and risks arising from uncertainty about the annuity rate and mortality prospects that will prevail at the time of retirement. These risks are borne entirely by the plan holder, and not by the pension plan provider, who essentially bears no risk.

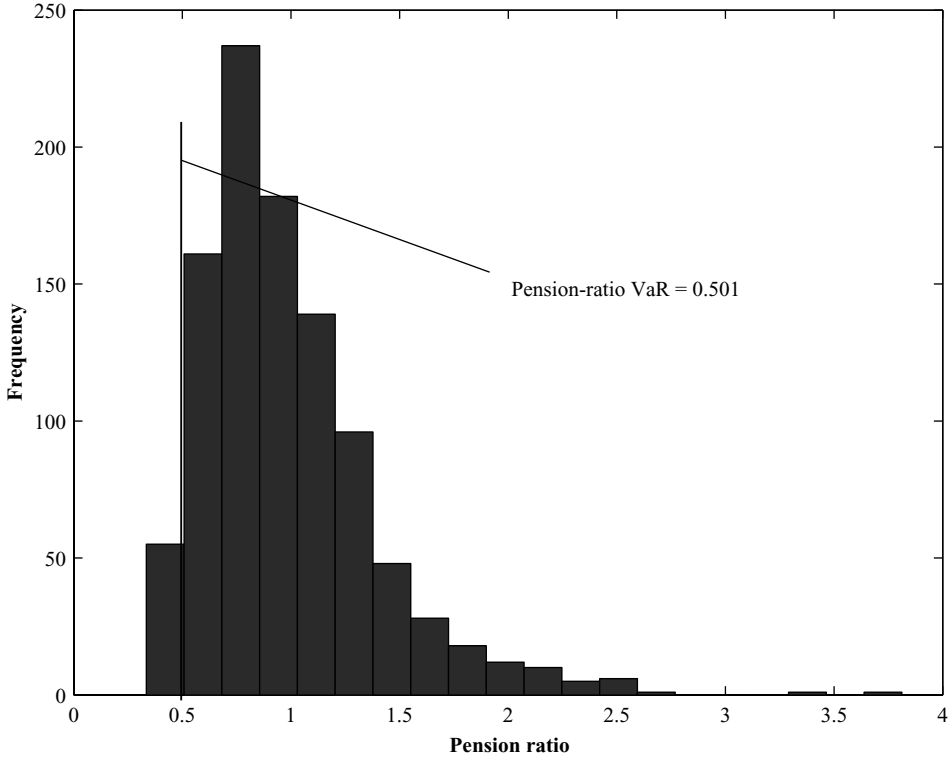

Figure 9.9 VaR of defined-contribution pension schemes
Note: Obtained using the 'dcpensionvar' function with initial income = \$25k, income growth rate of 0.02, conditional life expectancy = 80, contribution rate = 0.15, $\mu = 0.1$, $\sigma = 0.1$, $p = 0.05$, annuity rate = 0.04, 1000 trials, and $\alpha = 0.95$.

pension outcome at the relevant (and in this case 95%) confidence level – is 0.501, and this indicates that there is a 95% chance of a pension ratio higher than 0.501, and a 5% chance of a lower pension ratio.

9.8 CONCLUSIONS

This chapter has looked at a number of illustrative applications of simulation methods to risk measurement. Our discussion indicates that simulation methods can be extremely effective for a great variety of practical problems that are too complicated or too messy for analytical or algorithmic approaches, and simulation methods are particularly good at handling complications like 'badly behaved' stochastic processes, path dependency, non-linearity, optionality, and multiple dimensions. We should also keep in mind that the applications considered here merely scratch the surface of useful applications of simulation methods. In applying simulation methods, the key is to think in terms of how to program particular solutions to particular problems. However, they can be time consuming, especially when we are dealing with large numbers of dimensions. In such cases, we should consider variance-reduction techniques and principal components simulation methods such as scenario simulation.

10

Estimating Options Risk Measures

This chapter discusses the risk measurement of options positions. This is a very important subject for at least two reasons:

- There is a huge variety of different types of options (starting at one end with vanilla calls, puts, etc. and ranging to all manner of strange exotics at the other end). Many other types of financial instrument also have embedded options of one sort or another (e.g., bonds might be callable, convertible, callable and convertible, etc., futures and forwards might have embedded delivery options of some sort, etc.), and there are countless types of real options.
- Options are also notoriously risky, and we often want to get an idea of what we stand to lose on an options position. This is particularly the case where we have written options ourselves (i.e., have short positions) or where we have an options position that is short gamma or short vega. Being able to assess the riskiness of an options position is often critical for successful risk management.

Unfortunately, estimating the risks of options positions is also one of the most difficult areas of market risk measurement, even though many different estimation methods have been suggested. The approaches suggested fall under three broad headings:

- Analytical and algorithmic approaches.
- Simulation approaches.
- Delta–gamma and related approaches.

The analytical and algorithmic methods are the most straightforward, and should be used wherever possible. Unfortunately, they are very limited in their applicability. Simulation methods are more powerful, and can be applied to a much larger range of options positions. However, the application of simulation methods to estimate options risk measures is an underdeveloped area, and its full potential has yet to be realised. The delta–gamma and related methods are the best known methods and most widely used, but also have major limitations: in particular, they can be difficult to implement and are potentially very unreliable.

10.1 ANALYTICAL AND ALGORITHMIC SOLUTIONS FOR OPTIONS VaR

With simple options positions, it is sometimes possible to derive analytical (i.e., algebraic closed-form) or algorithmic (i.e., computer-search) solutions for options VaR.[1] Where they exist, such solutions are very useful, as they are accurate and easy to compute.

The process of solving options VaR analytically is illustrated in Figure 10.1. This figure is broken into three subfigures. The first is on the lower left-hand side, and shows how the tail

[1] For more on these, see Fong and Lin (1999).

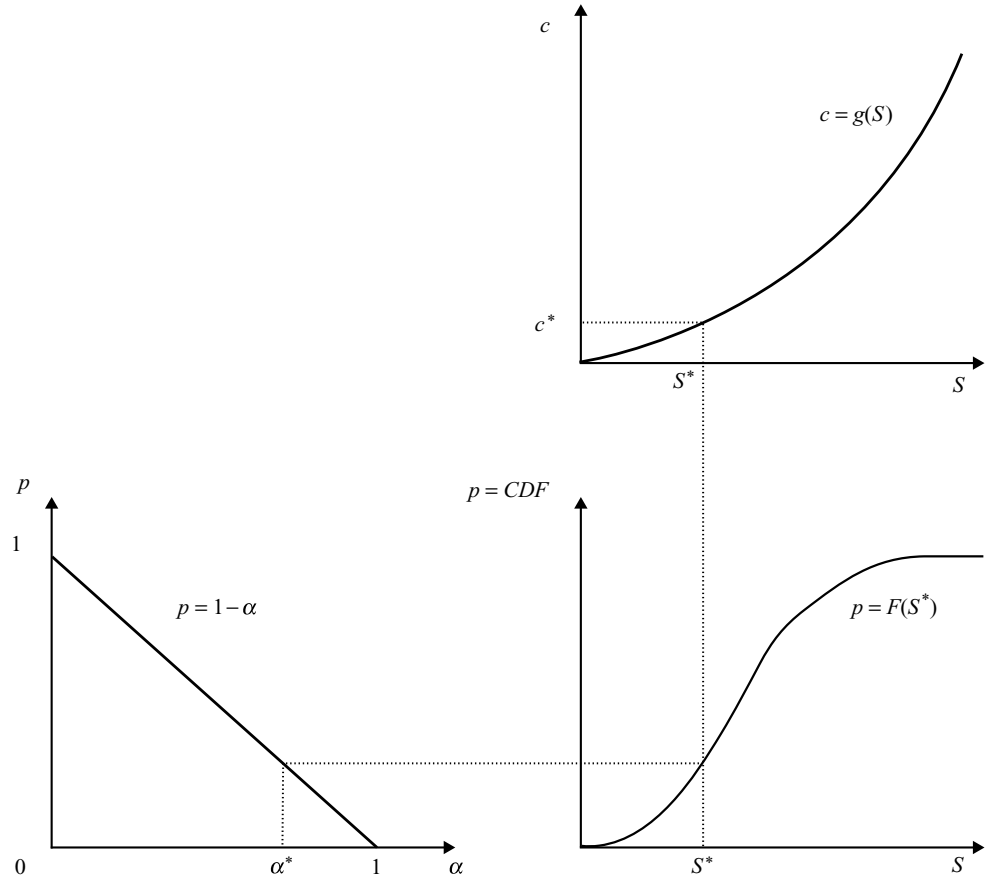

Figure 10.1 Diagrammatic illustration of analytic call option VaR

probability p is related to the confidence level α. This relationship is very basic, as $p = 1 - \alpha$. (We are implicitly assuming here that our option position is a long one in a call option, but we will address short positions presently.) The second subfigure is to the right of the first, and shows how the value of the underlying S relates to p, where p can now also be interpreted as the value of the underlying's cumulative density function. The third subfigure, on the top, shows how the value of the option, say c, relates to S.

To interpret this diagram, we start with some chosen confidence level α^*, and use the first subfigure to determine p^*, the corresponding critical value of p. We then use the second subfigure to read off S^*, the critical value of S corresponding to p^*, and after that we use the third subfigure to determine the corresponding critical option value, c^*. Our VaR at the confidence level α^* is then $c_0 - c^*$, where c_0 is the initial option value.

These subfigures correspond to the following three equations:

$$p = 1 - \alpha \tag{10.1}$$

$$p = F(S) \tag{10.2}$$

$$c = g(S) \tag{10.3}$$

where $p = F(S)$ is the distribution function for S, and $c = g(S)$ is the analytic solution for the option value. We now solve these equations to determine c^* in terms of a chosen confidence level α^*:

$$c^* = g(F^{-1}(1 - \alpha^*)) \tag{10.4}$$

The critical value c^* is also the lowest likely value of c we can get at the chosen confidence level. Hence, $c_0 - c^*$ is the highest likely loss at this confidence level (i.e., the VaR). Consequently, these equations, read in the context of Figure 10.1, give us an analytical solution for our options VaR.

But what are we assuming when we derive VaR in this way? The answer is that we are making two key assumptions:[2]

- We are assuming that we have a solution for the option value (i.e., we are assuming that $c = g(S)$ actually exists).
- More subtly, we are assuming that the function $g()$ is a monotonic function of S: we are assuming that $g()$ always rises or always falls as S rises.[3] This assumption enables us to obtain the VaR from c^*, for any chosen confidence level α^*.[4]

The importance of this latter assumption is not obvious at first sight, but becomes clearer if we examine what might happen if we don't make this assumption.

So let us assume for the moment that $g()$is not a monotonic function of S, but is instead a function that first rises and then falls with S, as illustrated in Figure 10.2.

Now choose two confidence levels, a high confidence level α^{hi} and a low confidence level α^{lo}, so that they produce the critical S-values $S^*(\alpha^{hi})$ and $S^*(\alpha^{lo})$ shown in the figure. Observe that the high confidence level produces an option value that is higher than the option value produced by the low confidence level (i.e., $c^*(\alpha^{hi}) > c^*(\alpha^{lo})$) – which is made possible by the fact that the curve is not monotonic. This implies that $VaR(\alpha^{hi}) = c - c^*(\alpha^{hi}) < c - c^*(\alpha^{lo}) = VaR(\alpha^{lo})$, so the VaR associated with the higher confidence level is less than the VaR associated with the lower confidence level – and this is impossible given the definition of VaR. We can now see why monotonicity is so useful: monotonicity ensures that the term $c_0 - c^*$ is a VaR.

If we have a short option position, the first equation, Equation (10.1), is replaced by:

$$p = \alpha \tag{10.5}$$

[2] We are also making a third assumption – namely, that the distribution function $p = F(S)$ is invertible. This is essential, but also relatively innocuous.

[3] The diagram in Figure 10.1 clearly presupposes that $g()$ always rises as S rises, but the important point is that $g()$ either always rises or always falls as S rises. If $g()$ always falls, then the high loss is made when S rises, rather than when it falls, and we should change Equation (10.1) to $p = \alpha$.

[4] Strictly speaking, this assumption is also overly restrictive. There are cases where $g(S)$ might be monotonically non-decreasing, or monotonically non-increasing, implying that the curve can sometimes be horizontal, and where we can still get reasonable analytical solutions for VaR. More importantly, where the $g(S)$ is not too badly behaved, we can sometimes solve for or approximate the VaR by suitable numerical algorithms. For example, Fong and Lin (1999) show that we can solve for the VaR of a straddle in this way, because a straddle has a convex $g(S)$ function with a single minimum. They also give other examples of such algorithms in cases where $g(S)$ might have multiple minima or be a function of more than one risk factor. However, these algorithms depend on local approximations around identified minima in the $g(S)$ function, and these minima are not always easily identified (e.g., as with path-dependent options). So, while the Fong–Lin algorithmic approach is sometimes useful in cases where $g(S)$ is not monotonic, there are many cases where the algorithms are too messy, or insufficiently accurate, or not applicable. An alternative algorithmic approach is suggested by Duffie and Pan (1999), and uses numerical integration methods to calculate VaR from characteristic functions. This approach is quite flexible and can accommodate jump diffusion, fat tails, skewness and credit-related risk, besides basic optionality. However, this approach is not easy to use and is not as flexible as simulation methods.

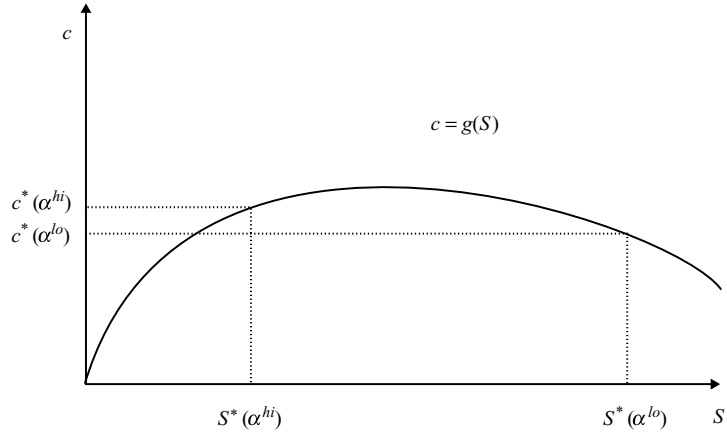

Figure 10.2 *g(S)* as a non-monotonic function of *S*

and the first subfigure is replaced by its mirror image, with a positive sloping curve. In this case, Equation (10.4) becomes:

$$c^* = g(F^{-1}(\alpha^*)) \tag{10.6}$$

and our VaR becomes $c^* - c_0$ instead of $c_0 - c^*$.

Example 10.1 (Analytical solution for long European call VaR)

Suppose we have a Black–Scholes call option with parameters as follows: current stock price $S_0 = 1$, strike $X = 1$, risk-free rate $r = 0.05$, volatility rate $\sigma = 0.25$, and period-to-expiry (in days) $t = 30$. We are also told that the trend growth term μ in the lognormal stock price process is 0.10. We wish to use an analytical solution to estimate the 95% VaR of this option over a horizon of 5 days.

We first note that the VaR is equal to $c_0 - c^*$, where c_0 is the current option price, and c^* is the critical end-holding-period option price. We use Black–Scholes to estimate c_0 as 0.0306.

To determine c^*, we begin by using Equation (10.1) to tell us that the tail probability $p = 1 - \alpha = 1 - 0.95 = 0.05$. We then need to determine the critical terminal stock price, S^*, associated with this probability, i.e., we seek $S^* = F^{-1}(p) = F^{-1}(0.05)$ using the inverse of Equation (10.2). To do so, we note that if S_t is lognormal, then $\ln S_t$ is distributed as normal with mean $\ln S_0 + (\mu - \sigma^2/2)t$ and variance $\sigma^2 t$, where t is here measured in years (see Chapter 6, Equation (6.17)). Now recall that if X is distributed as normal with mean μ_X and variance σ_X^2, then X^*, the critical value of X at the probability 0.05, is given by $\mu_X + \sigma_X z_{0.05} = \mu_X - \sigma_X 1.645$. We now use $\ln(S_t^*) = X^*$ and substitute out μ_X and σ_X^2 to get $\ln(S_t^*) = X^* = \mu_X - \sigma_X 1.645 = \ln S_0 + (\mu - \sigma^2/2)t - 1.645\sigma\sqrt{t}$. Hence, $S_t^* = \text{Exp}[\log S_0 + (\mu - \sigma^2/2)t - 1.645\sigma\sqrt{t}]$ for any t. Setting $t = 5/365 = 0.0137$, then given our parameters, $S_t^* = \text{Exp}[(0.10 - 0.25^2/2) \times 0.0137 - 0.25 \times \sqrt{0.0137} \times 1.645] = 0.954$. We now plug this value into the Black–Scholes formula to obtain the critical option price at the end of the holding period. This gives us $c^* = 0.0096$, and the VaR is $c_0 - c^* = 0.0306 - 0.0096 = 0.021$.

Example 10.2 (Analytical solution for long European put VaR)

Suppose we wish to estimate the VaR of a Black–Scholes put option given the same parameters as in Example 10.1.

We first note that the VaR is equal to $p_0 - p^*$, where p_0 is the current put price, and p^* is the critical end-holding-period put price. We use Black–Scholes to estimate p_0 as 0.0265. To determine p^*, we again note that Equation (10.1) tells us that the tail probability is $1 - \alpha = 0.05$. We then seek $S^* = F^{-1}(0.05)$ using the same approach as before, but in this case $X^* = \mu_X + \sigma_X 1.645$. (Because the option is a put, the critical stock price is now on the right-hand tail of the stock price distribution.) Using $X^* = \ln(S_t^*)$, we get $S_t^* = \text{Exp}[\log S_0 + (\mu - \sigma^2/2)t + 1.645\sigma\sqrt{t}]$, so $S_t^* = \text{Exp}[(0.10 - 0.25^2/2) \times 0.0137 + 0.25 \times \sqrt{0.0137} \times 1.645] = 1.0503$. We now insert this value into the Black–Scholes formula to obtain the critical option price at the end of the holding period. This gives us $p^* = 0.0080$, and the VaR is $p_0 - p^* = 0.0265 - 0.0080 = 0.019$.

Example 10.3 (Analytical solution for short European call VaR)

Suppose we wish to estimate the VaR of a short position in a Black–Scholes call option given the same parameters as in Example 10.1.

In this case, the VaR is equal to $c^* - c_0$, where c^* is the critical end-holding-period option price. However, c^* is now associated with the right-hand tail of the option price distribution, which is associated with the right-hand tail of the stock price distribution.

As in Example 10.1, we use Black–Scholes to estimate c_0 as 0.0306. To determine c^*, repeat the analysis of Example 10.1, but in this case end up with $X^* = (\mu - \sigma^2/2)t + 1.645\sigma\sqrt{t}$. Hence, as in Example 10.2, $S_t^* = \text{Exp}[(0.10 - 0.25^2/2) \times 0.0137 + 0.25 \times \sqrt{0.0137} \times 1.645] = 1.0503$. We then plug this value into Black–Scholes to get the critical option price $c^* = 0.0617$. The VaR is therefore $c^* - c_0 = 0.0617 - 0.0306 = 0.031$.

The reader will note that the VaR of the short position is about 50% larger than the VaR of the corresponding long position. It can of course be much more: the potential loss on a long position is bounded by the value of the option itself, but the potential loss on a short position is unbounded.

10.2 SIMULATION APPROACHES

A more powerful and flexible alternative is to estimate options risk measures using simulation. Simulation methods can be applied to a much greater range of options risk problems than closed-form or algorithmic methods, and they can be adapted to accommodate the early exercise features of American options, which the earlier methods cannot.

Simulation routines are also fairly straightforward to program, at least for European options: we simulate a sufficient number of paths for our underlying random variable; we ascertain the value of our option at the end of each simulation path; we construct the critical

terminal option price from the histogram of simulated end-holding-period option prices; and thence infer an estimate of whatever risk measure we are interested in. The key is to have a means of ascertaining the value of our option at the end of the holding period: depending on the context, this might involve the use of a lattice or simulation option-pricing engine.[5]

In dealing with the American-style options, we also have to accommodate the possibility that an option might be exercised before the end of the holding period. Where this happens, the end-holding-period option value is given by the option's exercised value carried forward to the end of the holding period, *not* by the contemporaneous price of an unexercised option. More specifically, suppose a hypothetical path for the underlying over the holding period leads our option position to have some particular value at the end of the holding period.[6] If the option was not exercised over the holding period, the end-holding-period value of the option position will equal the contemporaneous American option price. But if the option is exercised during the holding period, then the end-holding-period option-position value will equal the exercised value of our option position carried forward to the end of the holding period, and the latter will depend on a holding strategy that specifies what is done with the proceeds of an option exercised during the holding period. Ignoring the time value of money for convenience, and assuming the option to be a put if the option is exercised at time t and we make the reasonable assumption that the exercised position is converted to cash, then the value of a put position at the end of the holding period is $(X - S_t)$, where S_t is the stock price at t. So the end-holding-period value of our option position is the contemporaneous American put price if the option was not exercised over the holding period, and $(X - S_t)$ otherwise.

We now simulate a large number of paths for the underlying over the holding period. We then subdivide each path into a reasonable number of steps spanning the subperiods 0 to Δt, Δt to $2\Delta t$, and so on, up to the subperiod ending at h. For each path, we check for early exercise during each subperiod in turn (i.e., we check for early exercise at 0, then Δt, then $2\Delta t$, etc.) and so determine when, or if, the option is exercised over that path. This information gives us the end-holding-period value of our option position for each simulated path. Provided we simulate a sufficiently large number of paths, we can regard the distribution of end-holding-period option-position values as a surrogate for the true, but unknown, pdf of such values. We then obtain the critical end-holding-period option-position value OPV^* from this pdf, and the VaR is:

$$VaR = P - OPV^* \tag{10.7}$$

where P is the initial price of the American put.

This type of approach can be used to estimate the risks of a great range of different options, with or without early exercise features, including barrier options, Asian options, and many others. It can also accommodate alternative underlying stochastic processes (e.g., jump diffusion, etc.), multiple underlyings (e.g., term structures) and portfolios of heterogeneous options, and can be easily applied to estimate other risk measures besides the VaR. Whatever the option or the underlying process, the basic approach is always the same: to simulate the

[5] If we had a closed-form solution for the option price, there would be no point using a simulation method: the VaR could be inferred directly from the closed-form solution, as explained in section 10.1.

[6] In doing so, we also need to check if or when the option is exercised before the end of the holding period. To do so, we need to break up our holding period into discrete segments and check for early exercise at each node in the tree. This requires, either, that we have some pricing engine to give us the value of an unexercised option at each node, or that we can identify the early exercise boundary, and use that to determine interim option values. These refinements are fine in theory, but can add massively to the computational burden.

critical terminal option price or (where early exercise is possible) the critical terminal value of the original option position. The VaR is then inferred from the difference between this value and the original option price.

These methods can be supplemented with variance-reduction techniques to improve speed and accuracy, and there are many variance-reduction methods to choose from:

- *Antithetics*: Antithetics can deliver modest variance reductions. They require little extra code, and are therefore always worth checking out.
- *Control variates*: We should use control variates where we reasonably can, but whether we can find good controls depends largely on the specifics of the problem at hand. For example, if we are dealing with reverse knockouts, good controls are hard to come by, because the option is not well correlated with obvious candidate controls such as vanilla options or the underlying. On the other hand, an arithmetic Asian option has many good controls, including geometric Asian options, some of the standard approximations to arithmetic Asians (e.g., Turnbull and Wakeman, etc.), vanilla options, and the underlying itself.
- *Importance sampling*: IS methods are very well suited to VaR and ES estimation, particularly at high confidence levels, and can produce very large variance reductions. However, their usefulness depends on being able to find a good preliminary VaR estimate, such as a good quadratic approximation.
- *Stratified sampling*: Stratified sampling methods are well suited to VaR and ES estimation, particularly if we target the allocation of our strata around the VaR (or the tail, if we are estimating ES). On the other hand, basic stratified sampling can run into difficulties in high dimensional problems, and in such cases we might want to make use of Latin hypercube or Brownian bridge approaches. Stratified sampling also complicates the estimation of confidence intervals, and forces us to obtain such estimates using batch methods.
- *Moment matching*: These methods are easy to implement and have the potential to reduce variance by respectable amounts; however, their results can be biased and we have to obtain estimates of confidence intervals by batch methods.

Many of these methods can often be used together to produce very substantial cumulative reductions in variance. The use of such methods is especially recommended where we have to take account of possible early exercise, as this can add very considerably to the computational burden.

In practice, the major constraints with Monte Carlo methods relate to early exercise and the number of risk factors. It is also interesting to note that the number of different types of option in our options portfolio is not, as such, a problem. For example, suppose we start with a portfolio consisting of a single option, and then add a new option to it. If the new option is European and is written on the same underlying, the new option is accommodated merely by changing the payoff to the portfolio as a whole. Monte Carlo methods therefore have no difficulties with diverse European options written on the same underlyings. However, in practice, they eventually run into difficulties if we keep adding new American-style options, or if we keep adding new options that increase the number of risk factors on which the portfolio value depends: in particular, there are issues of calculation time, especially if we have to resort to calculation-intensive methods to deal with early exercise features. The severity of these problems is to a large extent context specific (e.g., it depends on how many different types of American option we have, etc.) and much more work is needed to establish how best to deal with them.

Box 10.1 Lattice Approaches to Options Risk Measurement

Lattice or tree methods have been used in finance for many years, and are widely used to price derivatives and estimate their Greek parameters. These methods are powerful, versatile and (fairly) reliable. They are useful for problems that lack closed-form solutions, and are ideally suited to handle the early exercise features of American options. They can also be used to estimate options risk measures.

To do so, we construct a lattice to estimate the cdf of the prospective values of our option position at the end of the holding period. Next, we infer the critical end-holding-period value of our option position from the cdf, and we can then obtain the VaR as the present value of the difference between the current and critical values of our option position. We then estimate the other risk measures in the usual way as the probability-weighted average of tail losses.

The accuracy of lattices can (sometimes) be improved by the use of variance-reduction techniques (such as control variates), and by replacing standard lattices with adaptive meshes.[7] Adaptive meshes enable us to refine measurements taken in sensitive areas of the lattice, without reducing step size everywhere else in the tree, which would impose a large increase in the computational burden.

Lattice methods have an advantage over Monte Carlo simulation methods in that the early exercise test for American-style options is much easier to carry out on a lattice, because the option price at each node can be obtained as a risk-neutralised probability-weighted average of the prices at the next two nodes, whereas with a simulation method the option price at each node has to be obtained by applying an option-pricing engine at each node, which is much more cumbersome.

On the other hand, relative to Monte Carlo methods, sophisticated lattices are (usually) not so easy to program, and lattice methods can suffer from severe 'curse of dimensionality' problems.

Example 10.4 (Simulated European call VaR)

Suppose we wish to use simulation to estimate the same option VaR as in Example 10.1. Applying the call option version of Equation (10.7), the VaR is $c_0 - c^*$, where c_0 is the current option price, and c^* is the critical end-holding-period option price. As in Example 10.1, we again use Black–Scholes to estimate c_0 as 0.0306. We then use 5000 simulation trials and the coding set out in the MMR MATLAB function 'example10point4' to estimate c^* as 0.0116. The estimated VaR is therefore $0.0306 - 0.0116 = 0.0190$, which is quite close to the true value of 0.021. Greater accuracy can be obtained if we increase the number of simulation trials.

10.3 DELTA–GAMMA AND RELATED APPROACHES

We turn now to the more traditional approaches to options VaR estimation, the delta–gamma and related approaches. The idea motivating these is to replace an option position with a

[7] For more on adaptive meshes, see Figlewski and Gao (1999).

surrogate position in an option's underlying variable, and then use a first- or (preferably) second-order approximation to estimate the VaR of the surrogate position. Such methods can be used to estimate the risks of any positions that are non-linear functions of an underlying risk factor; they can therefore be applied to options positions that are non-linear functions of an underlying variable, and to fixed-income instruments that are non-linear functions of a bond yield.

10.3.1 Delta–Normal Approaches

The simplest such approaches are delta–normal approaches, in which we replace the 'true' positions with linear approximations and then handle the linearly approximated positions in the same way as genuine linear positions in normal or lognormal risk factors.[8]

Imagine we have a straightforward equity call option of value c. The value of this option depends on a variety of factors (e.g., the price of the underlying stock, the exercise price of the option, the volatility of the underlying stock price, etc.), but in using the delta–normal approach we ignore all factors other than the underlying stock price, and we handle that by taking a first-order Taylor series approximation of the change in the option value:

$$\Delta c \approx \delta \Delta S \tag{10.8}$$

where $\Delta c = c - \bar{c}$ and $\Delta S = S - \bar{S}$, S is the underlying stock price, δ is the option's delta, and the dashes above c and S refer to the current values of these variables. If we are dealing with a very short holding period (i.e., so we can take δ as if it were approximately constant over that period), the option VaR, VaR^{option}, is:

$$VaR^{option} \approx \delta VaR^{S} \tag{10.9}$$

where VaR^{S} is the VaR of a unit of underlying stock.[9] The VaR is approximately δ times the VaR of the underlying stock. If S is normally distributed and the holding period is sufficiently short that we can ignore the expected return on the underlying stock, then the option VaR is:

$$VaR^{option} \approx \delta VaR^{S} \approx \delta S \sigma \sqrt{t} z_{\alpha} \tag{10.10}$$

where σ is the annualised volatility of S.

This approach gives us a tractable way of handling option positions that retains the benefits of linear normality without adding any new risk factors. The new parameter introduced into the calculation, the option δ, is also readily available for any traded option, so the delta–normal approach requires minimal additional data.[10]

However, these first-order approaches are only reliable when our portfolio is close to linear in the first place, since only then can a linear approximation be expected to produce an accurate approximation to a non-linear function. We can therefore get away with delta–normal techniques only if there is very limited non-linearity (i.e., a small amount of optionality or convexity) in our portfolio, but such methods can be very unreliable when positions have considerable optionality or other non-linear features.

---

[8] Any options risk approximation works better with a shorter holding period. The smaller the time period, the smaller the change dS and, hence, the smaller the squared change $(dS)^2$.

[9] We are also assuming that the option position is a long one. If the option position is short, the option VaR would be approximately $-\delta VaR^{S}$. However, these approximations only hold over very short time intervals. Over longer intervals, the long and short VaRs become asymmetric, and the usefulness of these approximations is, to say the least, problematic.

[10] We can handle the non-linearities of bond portfolios in a comparable way, using the duration approximation discussed in Chapter 1, section 1.2.2.

Example 10.5 (Delta normal call VaR)

Suppose we wish to estimate the delta–normal approximate VaR of the same option VaR as in Example 10.1. To carry out this calculation, we calculate the option δ (which turns out to be 0.537) and input this and the other parameter values into the delta–normal equation, Equation (10.10). The delta–normal approximate VaR is therefore $0.537 \times 0.25\sqrt{5/365} \times 1.645 = 0.026$. Given that the true value is 0.021, the delta–normal estimate has an error of about 20%.

10.3.2 Delta–Gamma Approaches

10.3.2.1 The delta–gamma approximation[11]

If a first-order approximation is insufficiently accurate, we can try to accommodate non-linearity by taking a second-order Taylor series (or delta–gamma) approximation. Taking such an approximation for a standard European call option gives us the following:

$$\Delta c \approx \delta \Delta S + \frac{\gamma}{2}(\Delta S)^2 \tag{10.11}$$

This approximation takes account of the gamma risk that the delta–normal approach ignores (cf. Equation (10.8)).[12] The impact of the gamma term is to raise the option price if gamma is positive and to reduce it if gamma is negative, and the correction it makes to the delta–normal estimate is particularly marked when the option has a high (positive or negative) gamma (e.g., as would be the case with at-the-money options that are close to maturity).[13] However, once we get into second-order approximations the problem of estimating VaR becomes much more difficult, as we now have the squared or quadratic terms to deal with. Equation (10.11) then leads to our delta–gamma VaR approximation:

$$VaR^{option} \approx \delta VaR^S - \frac{\gamma}{2}(VaR^S)^2 \approx \delta S\sigma\sqrt{t}z_\alpha - \frac{\gamma}{2}S^2\sigma^2 t z_\alpha^2 \tag{10.12}$$

The delta–gamma VaR estimate is thus a second-order function of the underlying VaR, which we can approximate by a second-order approximation to the underlying itself.

[11] There are many articles on delta–gamma and related approaches, and we only have space to discuss a small number. However, the interested might also look at, e.g., Zangari (1996a, b), Cárdenas *et al*. (1997), Fallon (1996), Studer (1999), Albanese *et al*. (2001), Mina (2001), and Feuerverger and Wong (2000).

[12] However, as is clear from the Black–Scholes equation, both delta–normal and delta–gamma approximations can also run into problems from other sources of risk. Even if the underlying price S does not change, a change in expected volatility will lead to a change in the price of the option and a corresponding change in the option's VaR: this is the infamous problem of vega risk, or the volatility of volatility. Similarly, the option's value will also change in response to a change in the interest rate (the rho effect) and in response to the passing of time (the theta effect). In principle, most of these effects are not too difficult to handle because they do not involve high-order terms, and we can tack these additional terms onto the basic delta–normal or delta–gamma approximations if we wish to, but the volatility of vega is a more difficult problem.

[13] There can be some difficult problems here. (1) The second-order approximation can still be inaccurate even with simple instruments such as vanilla calls. Estrella (1996, p. 360) points out that the power series for the Black–Scholes approximation formula does not always converge, and even when it does, we sometimes need very high-order approximations to obtain results of sufficient accuracy to be useful. However, Mori *et al*. (1996, p. 9) and Schachter (1995) argue on the basis of plausible-parameter simulations that Estrella is unduly pessimistic about the usefulness of Taylor series approximations, but even they do not dispute Estrella's basic point that results based on Taylor series approximations can be unreliable. (2) We might be dealing with instruments with more complex payoff functions than simple calls, and their payoff profiles might make second-order approximations very inaccurate (e.g., as is potentially the case with options such as knockouts or range forwards) or just intractable (as is apparently the case with the mortgage-backed securities considered by Jakobsen (1996)). (3) Especially in multifactor cases, it can be difficult even to establish what the second-order approximation might be: how do we deal with cross-gamma terms, stochastic calculus terms, and so on? For more on these issues, see Wiener (1999).

Note that the impact of a positive gamma is to *reduce* the option VaR: the reason for this is that a positive gamma raises the option price (see Equation (10.11)), and the higher price means that a long position in the option *loses less*; this smaller loss then implies a lower VaR.

Example 10.6 (Delta–gamma call VaR)

Suppose we wish to estimate the delta–normal approximate VaR of the same option VaR as in Example 10.1. To do so, we use the delta–gamma equation, Equation (10.12), with the relevant parameter values, along with the additional (easily obtained) information that δ and γ are 0.537 and 5.542. The delta–gamma approximate VaR is therefore $0.537 \times 0.25\sqrt{5/365} \times 1.645 - (5.542/2) \times 0.25^2 \times (5/365)^2 \times 1.645^2 = 0.022$.

This is very close to the true value (0.021). Thus, the addition of the gamma term leads to a lower VaR estimate which, in this case at least, is much more accurate than the earlier delta normal estimate (0.026). However, for reasons explained already, we cannot assume that a delta–gamma estimate will always be better than a delta one.

Box 10.2 A Duration–Convexity Approximation to Bond Portfolios

The second-order approximation approach used to handle non-linearity in options positions can also be used to handle non-linearity in bonds. Suppose we take a second-order approximation of a bond's price–yield relationship:

$$P(y + \Delta y) \approx P(y) + \frac{dP}{dy}\Delta y + \frac{1}{2}\frac{d^2P}{dy^2}\Delta y^2$$

We know from standard fixed-income theory that

$$\frac{dP}{dy} = -D^m P \quad \text{and} \quad \frac{d^2P}{dy^2} = CP$$

where D^m is the bond's modified duration and C its convexity. The percentage change in bond price is therefore

$$\frac{\Delta P}{P} \approx -D^m \Delta y + \frac{1}{2}C(\Delta y)^2$$

which is the second-order approximation for bond prices corresponding to the delta–gamma approximation for option prices given by Equation (10.11).

10.3.2.2 Rouvinez delta–gamma approach

An alternative approach is suggested by Christophe Rouvinez (1997). Assuming a long position in a single call option, we start by rearranging Equation (10.11) to get:

$$P/L \approx \frac{\gamma}{2}\left(\Delta S + \frac{\delta}{\gamma}\right)^2 - \frac{\delta^2}{2\gamma} \tag{10.13}$$

Since ΔS is normally distributed, it follows that

$$\frac{\gamma}{2}\left(\Delta S + \frac{\delta}{\gamma}\right)^2 \quad \sim \quad \chi_{1,(\delta/\gamma)^2} \tag{10.14}$$

where $\chi_{1,(\delta/\gamma)^2}$ refers to a non-central chi-squared with 1 degree of freedom and non-centrality parameter $(\delta/\gamma)^2$. We now infer the critical value of the term on the left-hand side of (10.14), and unravel the VaR from this.

A nice feature of this approach is that Rouvinez solves for the VaR based on a delta–gamma approximation, and this gives us an exact answer *given* that approximation. This represents an improvement over the earlier approach, because that involves a double approximation, in which we insert one approximation (a delta–normal VaR) into another (the delta–gamma approximation). On the other hand, the Rouvinez approach does not generalise easily to multiple risk factors.

10.3.2.3 The Britten-Jones/Schaefer delta–gamma approach

An approach that does generalise to multiple risk factors is provided by Britten-Jones and Schaefer (1999). Without going into detail here, they show that the m-factor delta–gamma approximation can be written as the sum of m non-central chi-squared variates. They also show that this sum of chi-squareds can be approximated by a single chi-squared, the parameters of which depend on the moments of the chi-squared sum. The delta–gamma VaR can then be inferred from the distribution of a single chi-squared. Thus, Britten-Jones and Schaefer provide an (approximate) solution for VaR based on a delta–gamma approximation, but their approach can also be applied to multi-factor problems. Their results also confirm the (obvious) intuition that the reliability of delta–gamma approaches depends to a large extent on the relative size of the gamma terms: the more important the gamma terms, the less reliable we should expect the approximation to be. However, their approach is not particularly easy to implement and their analysis highlights the subtleties and difficulties involved in such using approximations in a multivariate context.

10.3.2.4 The Bouchaud–Potters dominant factor approach

Another multi-factor approach is to use dominant factors. Suppose that we have m risk factors in our portfolio, and we assume that a large loss would be associated with a large move in a particular factor, say the first factor, which we can regard as a dominant factor. When a large loss occurs, we can also suppose that moves in the other factors would be negligible. (However, we can relax this assumption to allow more than one dominant factor, at the cost of added complexity.) Bouchaud and Potters (2000b) show that for these assumptions to be plausible, we require that the tail of the first factor should decay no faster than an exponential, but this assumption is often reasonable in financial markets. The change in the value of our portfolio is then approximated by

$$\Delta f(e_1, e_2, \ldots, e_m) + \sum_{i=2}^{m} \delta_i e_i + \frac{1}{2} \sum_{i=2, j=2}^{m} \gamma_{i,j} e_i e_j \tag{10.15}$$

where $f(e_1, e_2, \ldots, e_m)$ is the portfolio value for risk-factor realisations $e_1, e_2, \ldots, e_m$, and Δ is the usual difference operator. The VaR is then given by the value of Equation (10.15) when e_1 takes some critical value e_1^* and the others are, say, all zero. If the tail decays exponentially

so that the pdf of e_1 is approximately proportional to $\alpha_1 \exp(-\alpha_1 e_1)$, where α_1 is the exponent index associated with the dominant factor, Bouchaud and Potters go on to show that an approximate value of e_1^* can be solved from

$$\exp(-\alpha_1 e_1^*)\left[1 - \sum_{i=2}^{m} \frac{\delta_i^2 \alpha_1^2 \sigma_i^2}{2\delta_1^2}\right] = \alpha \tag{10.16}$$

where the delta and gamma terms are evaluated at $e_1 = e_1^*$ and $e_i = 0$ for $i > 1$, and the 'raw' α is still our confidence level. We then substitute the solved value of e_1^* and zero values for $e_2, \ldots, e_m$ into Equation (10.15) to obtain the VaR. The Bouchard–Potters approach gives us an alternative multifactor approach based on the dominant factor hypothesis, but as with the previous approach, it is not particularly easy to implement, and becomes even less so as we introduce additional dominant factors.

10.3.2.5 Wilson's delta–gamma approach

One other approach should also be discussed, not only because it has been widely cited, but also because it is closely related to the mechanical stress testing approaches discussed in Chapter 13. This is the quadratic optimisation delta–gamma approach proposed by Tom Wilson (1994b, 1996). Wilson starts with the definition of the VaR as the maximum possible loss with a given level of probability. Wilson suggests that this definition implies that the VaR is the solution to a corresponding optimisation problem, and his proposal is that we estimate VaR by solving this problem.[14] In the case of a single call option, he suggests that the VaR can be formally defined as the solution to the following problem:

$$\begin{aligned} VaR = {} & \text{Max}[-\Delta c], \text{ subject to} (\Delta S)^2 \sigma_S^{-2} \leq z_\alpha^2 \\ & \{\Delta S\} \end{aligned} \tag{10.17}$$

In words, the VaR is the maximum loss (i.e., the maximum value of $-[\Delta c]$ for a long position) subject to the constraint that underlying price changes occur within a certain confidence interval. The bigger is the chosen confidence level, the bigger is z_α and the bigger the permitted maximum price change ΔS.[15] In the present context we also take the option price change Δc to be proxied by its delta–gamma approximation:

$$\Delta c \approx \delta \Delta S + \frac{\gamma}{2}(\Delta S)^2 \tag{10.18}$$

In general, this approach allows for the maximum loss to occur with $(\Delta S)^2$ taking any value in the range permitted by the constraint, i.e.,

$$0 \leq (\Delta S)^2 \leq z_\alpha^2 \sigma_S^2 \tag{10.19}$$

which in turn implies that

$$-z_\alpha \sigma_S \leq \Delta S \leq z_\alpha \sigma_S \tag{10.20}$$

[14] Wilson himself calls his risk measure 'capital at risk' rather than value at risk, but the concepts are similar and I prefer to use the more conventional term. However, there are important differences between the VaR (or whatever else we call it) implied by a quadratic programming approach (of which Wilson's is an example) and conventional or 'true' VaR, and we will come back to these differences a little later in the text.

[15] To avoid further cluttering of notation, I am ignoring the $\sqrt{t}$ terms that go with the σ terms, which we can also regard as the latter terms being rescaled.

However, in this case, we also know that the maximum loss occurs when ΔS takes one or other of its permitted extreme values, i.e., where $\Delta S = z_\alpha \sigma_S$ or $\Delta S = -z_\alpha \sigma_S$. We therefore substitute each of these two values of ΔS into Equation (10.20) and the VaR is the bigger of the two losses.

Wilson also applies his approach to portfolios with more than one instrument, but in doing so it becomes more difficult to implement. In this more general case, the QP VaR is given by the solution to the following quadratic programming (QP) optimisation problem:

$$VaR = \underset{\{\Delta S\}}{\text{Max}} - [\boldsymbol{\delta}^{\mathrm{T}}\mathbf{\Delta S} + \mathbf{\Delta S}^{\mathbf{T}}\boldsymbol{\gamma}\mathbf{\Delta S}/2], \text{ subject to } \mathbf{\Delta S}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{\Delta S} \leq z_\alpha^2 \tag{10.21}$$

where $\boldsymbol{\delta}$ is a vector of deltas, $\boldsymbol{\gamma}$ is a matrix of gamma and cross-gamma terms, the superscript 'T' indicates a transpose and we again use bold face to represent the relevant matrices. This problem is a standard quadratic programming problem, and one way to handle this problem is to rewrite the function to be optimised in Langrangian form:

$$L = -[\boldsymbol{\delta}^{\mathbf{T}}\mathbf{\Delta S} + \mathbf{\Delta S}^{\mathrm{T}}\boldsymbol{\gamma}\mathbf{\Delta S}/2] + \lambda[\mathbf{\Delta S}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{\Delta S} - z_\alpha^2] \tag{10.22}$$

We then differentiate L with respect to each element of $\mathbf{\Delta S}$ to arrive at the following set of Kuhn–Tucker conditions describing the solution:

$$\begin{gathered}[-\boldsymbol{\gamma} - \lambda\boldsymbol{\Sigma}^{-1}]\mathbf{\Delta S} = \boldsymbol{\delta} \\ \mathbf{\Delta S}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{\Delta S} \leq z_\alpha^2 \\ \lambda\mathbf{\Delta S}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{\Delta S} - z_\alpha^2 = 0 \quad \text{and} \quad \lambda \geq 0\end{gathered} \tag{10.23}$$

where λ is the Lagrange multiplier associated with the constraint, which reflects how much the VaR will rise as we increase the confidence level. The solution, $\mathbf{\Delta S}^*$, is then

$$\mathbf{\Delta S}^* = \mathbf{A}(\lambda)^{-1}\boldsymbol{\delta} \tag{10.24}$$

where $\mathbf{A}(\lambda) = -[\boldsymbol{\gamma} + \lambda\boldsymbol{\Sigma}^{-1}]$. Solving for $\mathbf{\Delta S}^*$ requires that we search over each possible λ value and invert the $\mathbf{A}(\lambda)$ matrix for each such value. We also have to check which solutions satisfy our constraint and eliminate those that do not satisfy it. In so doing, we build up a set of potential $\mathbf{\Delta S}^*$ solutions that satisfy our constraint, each contingent on a particular λ-value, and then we plug each of them into Equation (10.22) to find the one that maximises L.[16]

Unfortunately, this QP approach suffers from a major conceptual flaw. Britten-Jones and Schaefer point out that there is a subtle but important difference between the 'true' VaR and the QP VaR: the 'true' VaR is predicated on a confidence region defined over portfolio value changes, while the QP VaR is predicated on a confidence region defined over (typically multidimensional) factor realisations.[17] These are quite different. Furthermore, there is a deeper problem: it is generally not possible to use confidence regions defined over factors to make inferences about functions of those factors. Were that possible, Britten-Jones and Schaefer point out, much of the work in statistics on distributions of functions of random variables would be unnecessary. As they further point out,

[16] However, implementing this procedure is not easy. We have to invert bigger and bigger matrices as the number of risk factors gets larger, and this can lead to computational problems (e.g., matrices failing to invert). That said, we can ameliorate these problems if we are prepared to make some simplifying assumptions, and one useful simplification is to assume that the $\mathbf{A}(\lambda)$ matrix is diagonal. If we make this assumption Equation (10.24) gives us closed-form solutions for $\mathbf{\Delta S}^*$ in terms of λ without any need to worry about matrix inversions. Computations become much faster, but even this improved procedure can be tedious.

[17] Britten-Jones and Schaefer (1999), Appendix A.

> Simply because a point lies within a 95% confidence region does not mean that it has a 95% chance of occurrence. A point may lie within some 95% region, have a negligible chance of occurring and have a massive loss associated with it. The size of this loss does not give any indication of the true VaR. In short the QP approach is conceptually flawed and will give erroneous results under all but special situations where it will happen to coincide with the correct answer.[18]

Britten-Jones and Schaefer go on to prove that the QP VaR will, in general, exceed the true VaR, but the extent of the overstatement will depend on the probability distribution from which P/L is generated.[19]

So, in the end, all we have is a risk measure that generally overestimates the VaR by an amount that varies from one situation to another. It is therefore not too surprising that empirical evidence suggests that the QP approach can give very inaccurate VaR estimates.[20]

10.3.2.6 Some conclusions on delta–gamma approaches

In principle, delta–gamma approaches can be very useful. They can give us approximations to the VaRs of options positions, which we may wish to use for any of a number of reasons:

- They may be the easiest method available.
- We may have reason to believe that they are accurate in some context.
- We may wish to make use of them for mapping purposes (see Chapter 12).
- We may wish to use them to provide an initial starting value for a more sophisticated method (e.g., when using importance sampling to estimate risk measures in Monte Carlo simulation).

On the other hand, delta–gamma methods can also be difficult to implement (e.g., because deltas or gammas might change, the 'right' type of delta–gamma (or other Greek-based) approximation might not be clear, etc.) and unreliable. Delta–gamma methods are definitely to be handled with care.

Box 10.3 Estimating Bounds for VaR

Another response to non-linearity and optionality is to compute bounds for our VaR, and Rouvinez suggests a number of ways to do so.[21] One approach is to start with the well-known Chebyshev inequality. Using obvious notation, this states that:

$$\Pr\{|X - \mu| > s\sigma\} \leq \frac{1}{s^2}$$

for some arbitrary $s > 1$. This can be solved to give

$$VaR \leq \left| \mu - \sqrt{\frac{1}{1-c}}\sigma \right|$$

[18] Britten-Jones and Schaefer (1999), p. 186.

[19] In the simple case where we have positions that are linear positions in m normal risk factors, Studer and Lüthi (1997) show that the ratio of QP 'VaR' (or maximum loss, to use the stress-testing terminology of Chapter 13) to 'true' VaR is equal to the square root of the relevant percentile of chi-squared distribution with m degrees of freedom divided by the corresponding standard normal percentile. This enables us to infer the VaR from its QP equivalent. However, this finding is of relatively limited use: we don't need this formula in the linear normal case, because we can easily estimate the 'true' VaR more directly; and the formula does not generally apply in the more difficult non-linear/non-normal cases where it might have come in useful.

[20] See, e.g., Pritsker (1997), p. 231.

[21] Rouvinez (1997), pp. 58–59.

for $c = 1 - 1/s^2$. However, this bound is unsatisfactory in that it accounts for both tails of the P/L distribution, and we are usually only interested in the lower tail.

It would therefore be better to use an inequality that focuses on the lower tail, and one that does so is the Rohatgi inequality:

$$\Pr\{|X - \mu| < s\} \leq \frac{\mu}{\mu + s^2}$$

for arbitrary $s < 0$. This yields a superior VaR bound:

$$VaR \leq \left| \mu - \sqrt{\frac{c}{1-c}}\sigma \right|$$

If the higher moments exist, there is a second inequality, also due to Rohatgi, which states that

$$\Pr\{|X - \mu| > s\sigma\} \leq \frac{k-1}{s^4 - 2s^2 + k}$$

for $s > 1$ where k is the coefficient of kurtosis. This yields the following VaR bound:

$$VaR \leq \left| \mu - \sqrt{1 + \sqrt{\frac{c(k-1)}{1-c}}}\sigma \right|$$

which, given that we are usually interested in relatively high confidence levels, generally provides a better (i.e., tighter) risk bound than the Chebyshev inequality does.

10.4 CONCLUSIONS

This chapter has looked at the estimation of options risk measures, and a number of conclusions suggest themselves. Where possible, we should use analytical methods on the principle that the simplest adequate method is always the best. Unfortunately, these are few and far between, and where these are not available, we should look for algorithmic (if they are suitable and relatively easy to program) or simulation methods instead. Simulation methods are particularly attractive because they are powerful and flexible, and generally straightforward to program. Depending on the context, we might also look at delta–gamma and related approaches, but these can be cumbersome and their reliability is sometimes highly questionable.

Underlying all this, the real problem in practice has to do with the estimation of risk measures for complex options portfolios (i.e., portfolios of heterogeneous options with large numbers of underlyings and, often, the additional complications of early exercise). This is a pressing problem because many real-world options portfolios are of exactly this nature. Analytical and algorithmic approaches are of very limited use in this context, and although some delta–gamma methods can handle heterogeneity to some extent, they do so unconvincingly: they are difficult to implement, and reliability is a major concern. This leaves us with simulation methods. In theory, simulation methods ought to be able to manage some of these problems much better, but the use of simulation methods to estimate option risk measures is an underdeveloped area, and much more work needs to be done to establish how such methods can be best deployed on such problems. In the meantime, practitioners will simply have to make do as best they can.

11

Incremental and Component Risks

This chapter considers risk decomposition, and we are concerned in particular with the two main types of decomposition:

- *Incremental risks*: These are the changes in risk when some factor changes. For example, we might want to know how the VaR changes when we add a new position to our portfolio, and in this case the incremental VaR or IVaR is the change in VaR associated with the addition of the new position to our portfolio.
- *Component risks*: These are the component or constituent risks that make up a certain total risk. For instance, if we have a portfolio made up of particular positions, the portfolio VaR can be broken down into components, known as component VaRs or CVaRs, that tell us how much each position contributes to the overall portfolio VaR.

Measures of incremental and component risks can be very useful risk management tools. Risk decomposition reports help to gain insight into a portfolio, and they are particularly useful for identifying high sources of risk and their opposite, natural hedges (or positions that reduce overall risk). The information they provide can be used for choosing hedges, making investment decisions, allocating capital, communicating and disclosing risk, and for other risk management purposes.

To keep the discussion straightforward, we will assume for most of this chapter that our benchmark risk measure is the VaR.[1] However, the analysis carries over to coherent risk measures as well, and we will say a little bit more about the decomposition of coherent risk measures at the end.

11.1 INCREMENTAL VaR

11.1.1 Interpreting Incremental VaR

If VaR gives us an indication of portfolio risks, IVaR gives us an indication of how those risks change when we change the portfolio itself. More specifically, the IVaR is the change in portfolio VaR associated with adding the new position to our portfolio. There are three main cases to consider, and these are illustrated in Figure 11.1:

- *High IVaR*: A high positive IVaR means that the new position adds substantially to portfolio risk. Typically, the IVaR not only rises with relative position size, but also rises at an increasing rate. The reason for this is that as the relative position size continues to rise, the new position has an ever-growing influence on the new portfolio VaR, and hence the IVaR, and increasingly drowns out diversification effects.

[1] There is an extensive literature on the decomposition of VaR. For a good taste of it, see Ho *et al*. (1996), Garman (1996a,b,c), Litterman (1996), Dowd (1999b), Hallerbach (1999), Aragonés *et al*. (2001), or Tasche and Tibiletti (2003).

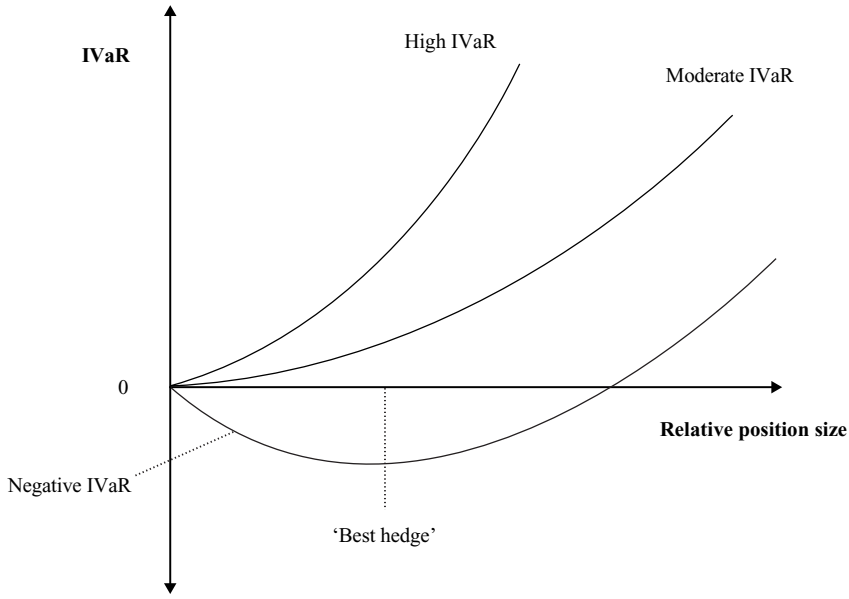

Figure 11.1 Incremental VaR and relative position size

- *Moderate IVaR*: A moderate positive IVaR means that the new position adds moderately to portfolio risk, and once again, the IVaR typically rises at an increasing rate with relative position size.
- *Negative IVaR*: A negative IVaR means that the new position reduces overall portfolio risk VaR, and indicates that the new position is a natural hedge against the existing portfolio. However, as its relative size continues to rise, the IVaR must eventually rise because the IVaR will increasingly reflect the VaR of the new position rather than the old portfolio. This implies that the IVaR must have a shape similar to that shown in the figure – it initially falls, but bottoms out, and then rises at an increasing rate. So any position is only a hedge over a limited range of position sizes, and ceases to be a hedge when the position size gets too large. The point (or relative position) at which the hedge effect is largest is known as the 'best hedge' and is a useful reference point for portfolio risk management.

11.1.2 Estimating IVaR by Brute Force: the 'Before and After' Approach

The most straightforward and least subtle way to estimate IVaR is a brute force, or 'before and after', approach. This approach is illustrated in Figure 11.2. We start with our existing portfolio p, map the portfolio and obtain our portfolio VaR, $VaR(p)$. (We have more to say on mapping in the next chapter.) We then consider the candidate trade, a, construct the hypothetical new portfolio that we would have if we went ahead with the trade, and do the same for that portfolio. This gives us the new portfolio VaR, $VaR(p + a)$, say. The IVaR associated with trade or position a, $IVaR(a)$, is then estimated as the difference between the two VaRs:

$$IVaR = VaR(p + a) - VaR(p) \tag{11.1}$$

Unfortunately, this 'before and after' approach has a fairly obvious drawback. If we have a large number of different positions – and particularly if we have a lot of optionality or other

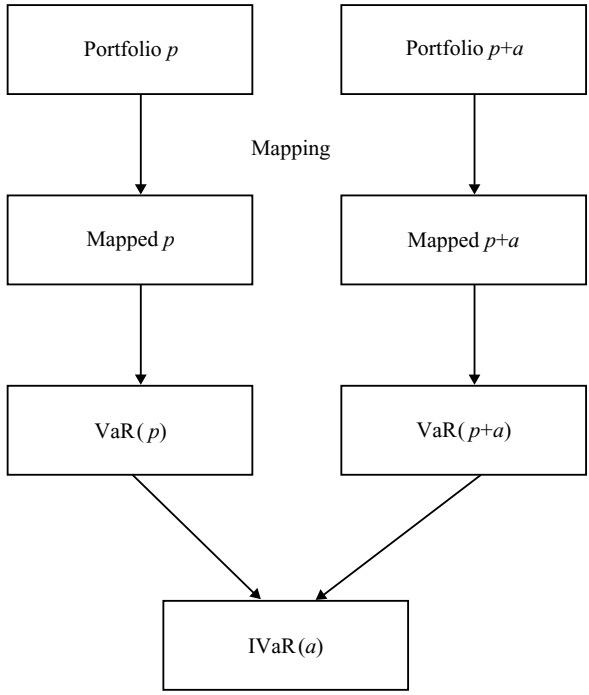

Figure 11.2 The 'before and after' approach to IVaR estimation

forms of non-linearity – then estimating each VaR will take time. Many financial institutions often have tens of thousands of positions, and re-evaluating the whole portfolio VaR can be a time-consuming process. Because of the time they take to obtain, IVaR estimates based on the 'before and after' approach are often of limited practical use in trading and real-time decision-making.

11.1.3 Estimating IVaR using Analytical Solutions

11.1.3.1 Garman's 'delVaR' approach

An elegant way to reduce the computational burden is suggested by Garman (1996a,b,c). His suggestion is that we estimate IVaR using a Taylor series approximation based on marginal VaRs (or, if we like, the mathematical derivatives of our portfolio VaR). Again, suppose we have a portfolio p and wish to estimate the IVaR associated with adding a position a to our existing portfolio. We begin by mapping p and a to a set of n instruments. The portfolio p then has a vector of (mapped) position sizes in these instruments of $[w_1, \ldots, w_n]$ (so w_1 is the size of our mapped position in instrument 1, etc.) and the new portfolio has a corresponding position-size vector of $[w_1 + \Delta w_1, \ldots, w_n + \Delta w_n]$. If a is 'small' relative to p, we can approximate the VaR of our new portfolio (i.e., $VaR(p + a)$) by taking a first-order Taylor series approximation around $VaR(p)$, i.e.,

$$VaR(p + a) \approx VaR(p) + \sum_{i=1}^{n} \frac{\partial VaR}{\partial w_i} dw_i \tag{11.2}$$

where $dw_i \approx \Delta w_i$. Where the changes are infinitesimally small, the partial derivative terms on the right-hand side of Equation (11.2) are sometimes known as marginal VaRs. The IVaR associated with position a, $IVaR(a)$, is then

$$IVaR(a) = VaR(p + a) - VaR(p) \approx \sum_{i=1}^{n} \frac{\partial VaR}{\partial w_i} dw_i \tag{11.3}$$

which gives us the incremental VaR in terms of approximate marginal VaRs. If we wish, we can rewrite Equation (11.3) in matrix notation as:

$$IVaR(a) \approx \nabla \mathbf{VaR(p)dw} \tag{11.4}$$

where **dw** is the transpose of the $1 \times n$ vector $[dw_1, \ldots, dw_n]$ and ∇**VaR(p)**, known as 'delVaR', is the $1 \times n$ vector of partial derivatives of $VaR(p)$ with respect to the w_i. Equation (11.4) gives us an approximation to the IVaR associated with position a given information on the ∇**VaR(p)** and **dw** vectors: the latter is readily obtained from mapping the position, and the former (which depends only on the existing portfolio p) can be estimated at the same time that $VaR(p)$ is estimated. This means that we can approximate the IVaR associated with position a using only one set of initial estimates – those of $VaR(p)$ and ∇**VaR(p)** – relating to the original portfolio, and the only information we need about the position itself is its (readily available) mapped position-size vector $[dw_1, \ldots, dw_n]$. This allows us to estimate as many different IVaRs as we like, given only one set of estimates of *VaR(p)* and ∇**VaR(p)**. This 'delVaR' approach is very useful because it enables us to estimate and use IVaRs in real time – for instance, when assessing investment risks and specifying position limits.

The process of estimating IVaR using the delVaR approach is straightforward. We begin by mapping our portfolio and using market data to estimate the portfolio VaR and delVaRs. Observe, too, that these depend on the portfolio we already have, and not on any candidate trades. Once we have the portfolio VaR and delVaRs, we can then take any candidate trade a, map the trade, and use the mapped trade and delVaRs to estimate the IVaR associated with that candidate trade.

The only question that remains is how to estimate ∇**VaR(p)**, and we can always estimate the terms in this vector by suitable approximations – we can estimate $\partial VaR/\partial w_i$ by estimating the VaR for position sizes w_i and $w_i + \Delta w_i$, and taking $\partial VaR/\partial w_i \approx (VaR(p|w_i + \Delta w_i) - VaR(p|w_i))/\Delta w_i$, where $VaR(p|w_i)$ is the VaR of p with position size i equal to w_i, etc.

In some cases, we can also solve ∇**VaR(p)** algebraically. For example, where P/L is normally distributed with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$, ∇**VaR(p)** is:

$$\nabla \mathbf{VaR(p)} = -\boldsymbol{\mu}\mathbf{dw} + \frac{\boldsymbol{\Sigma}\mathbf{w} z_\alpha}{[\mathbf{w}^\mathbf{T}\boldsymbol{\Sigma}\mathbf{w}]^{1/2}} \tag{11.5}$$

Equation (11.5) allows us to estimate ∇**VaR(p)** making use of information about the position-size vectors for the existing portfolio (**w**) and the new position (**dw**), the mean vector $\boldsymbol{\mu}$ and the variance–covariance matrix $\boldsymbol{\Sigma}$ – all of which are readily available or already known.

The delVaR approach could be implemented on a daily cycle. At the start of each trading day, we would estimate both the VaR and delVaR of our existing portfolio. Estimates of IVaRs could then be obtained using Equations (11.4) and (11.5) based on our initial daily estimates of *VaR(p)* and ∇**VaR(p)**. These estimates of IVaRs could be done extremely quickly without the arduous process of re-estimating portfolio VaRs throughout the day, as we would

have to do using a 'before and after' approach. Experience suggests that this approach works quite well for most institutions most of the time.

Example 11.1 (The delVaR approach to IVaR)

Suppose we have a portfolio, currently worth \$1, invested between two assets 1 and 2 in relative shares of 0.7 and 0.3. The returns on these assets are multivariate standard normal with correlation coefficient ρ. We wish to estimate the IVaR associated with a purchase of \$0.01 in asset 1.

To answer this question using the delVaR approach, we first need to know the partial derivatives of *VaR* with respect to each asset i. We can find them from the multivariate normal *VaR* equation, given by Equation (6.25) in Chapter 6. Plugging $\mu = 0$ and $\sigma = 1$ into this yields

$$VaR = [w_1, w_2] \begin{bmatrix} 1, \rho \\ \rho, 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} z_\alpha = (w_1^2 + 2\rho w_1 w_2 + w_2^2) z_\alpha$$

where w_1 and w_2 are the relative shares of the two assets. It follows that

$$\frac{\partial VaR}{\partial w_1} = 2(w_1 + \rho w_2) z_\alpha = 2\,(0.7 + 0.3\rho) z_\alpha = (1.4 + 0.6\rho) z_\alpha$$
$$\frac{\partial VaR}{\partial w_2} = 2(w_2 + \rho w_1) z_\alpha = 2\,(0.3 + 0.7\rho) z_\alpha = (0.6 + 1.4\rho) z_\alpha$$

The purchase of the additional amount of stock 1 changes the relative position shares by

$$dw_1 = \frac{0.71}{1.01} - 0.7 \approx 0.003 \quad \text{and} \quad dw_2 = \frac{0.30}{1.01} - 0.3 \approx -0.003$$

Plugging all these into Equation (11.3) gives us

$$IVaR \approx \frac{\partial VaR}{\partial w_1} dw_1 + \frac{\partial VaR}{\partial w_2} dw_2 = 0.8(1 - \rho) 0.003 z_\alpha$$

So, for example, if $\rho = 0.5$ and $\alpha = 0.95$, then the *IVaR* is equal to $\approx 0.8 \times (1 - 0.5) \times 0.003 \times 1.645 = 0.002$. Given that the VaR is $(0.7^2 + 2 \times 0.5 \times 0.7 \times 0.3 + 0.3^2) \times 1.645 = 1.2996$, this means that the IVaR is about 0.15% of the original VaR.

Now consider an alternative portfolio, with initial weights $w_1 = 0$ and $w_2 = 1$. If the other parameters are the same, and going through the similar calculations as before, then $IVaR \approx -0.02 \times (1 - 0.5) \times z_\alpha = -0.01 \times z_\alpha$. In this case, the addition of the new asset to our portfolio produces a negative IVaR (i.e., reduces the overall VaR), because it creates diversification effects within the portfolio – and this is exactly as we would expect from basic portfolio theory.

This example also illustrates that the magnitude and even sign of the IVaR estimates can be sensitive to the initial weights, w_1 and w_2.

11.1.3.2 Drawbacks of the delVaR approach

Nonetheless, the delVaR approach only approximates IVaR, and is therefore only as good as the approximation itself. When the position or trade considered is 'small' relative to the size of the original portfolio, the approximation should be a good one and we could expect the delVaR

approach to be reliable. However, there are two circumstances in which this procedure might not be reliable:

- If we are dealing with very large trades, the first-order Taylor series might not give us a good approximation for the VaR of the new portfolio, and in this case the resulting IVaR approximation might be poor.
- If we have a large number of small trades accumulating during the day, the sum of daily trades will cause the intra-day portfolio to drift away from the start-of-day portfolio, and the VaR and delVaRs of the latter will be increasingly poor proxies for the VaR and delVaRs of the former. Inaccurate VaR and delVaR estimates can then lead to inaccurate IVaR estimates due to drift in the portfolio composition, even if individual trades are all 'small'.

Whether these problems are significant will depend on our circumstances, but if we wish to make our IVaR estimates more accurate, we can do so by re-estimating the portfolio VaR and delVaR more frequently: for instance, we can re-estimate VaR and delVaR after a particularly big trade, or after a specified number of trades have taken place, or every so often (e.g., every so many minutes) during the trading day.

11.1.3.3 Approximate IVaR in terms of portfolio VaR

If we continue to assume that our position in an asset is 'small' relative to the size of our overall portfolio, then there is also a simpler approximation for the IVaR:

$$IVaR_i \approx w_i \beta_i VaR(p) \tag{11.6}$$

where w is the relative share of asset i in the portfolio, β_i is the beta coefficient of asset i in the portfolio.[2] This gives us an approximation for the IVaR in terms of the relative weight, the asset's beta (measured against the portfolio), and the overall portfolio VaR. In doing so, it also makes the dependence of the IVaR on the asset's portfolio beta very explicit: clearly, the IVaR rises as the beta gets bigger, and vice versa i and if $VaR(p) > 0$, the IVaR has the same sign as β. It also tells us that the IVaR is approximately proportional to the VaR of the original portfolio: so doubling *VaR(p)* leads to a doubling of the IVaR, and so on.

This expression has the advantage that it requires little information and very little work. Given that the relative share and portfolio VaR are already known, all we need is the beta of the new position, and the IVaR can then be estimated in one simple calculation. And, as with the previous approximation, this approximation has a ready intuition in terms of basic portfolio theory. However, this approximation to the IVaR is likely to be less accurate than the Garman one, because it does not take account of the impact of the trade on risks *within* the existing portfolio, which the Garman approximation does.

Example 11.2 (Obtaining approximate IVaR from the portfolio VaR)

Suppose we have $w_i = 0.01$ and $\beta = 0.8$. If we plug these values into Equation (11.6), then our estimated 95% IVaR is about $0.01 \times 0.8\ VaR(p)$, or about 0.8% of the portfolio VaR.

---

[2] The approximation is derived in Dowd (1999b), p. 32.

11.1.3.4 Approximate IVaR in terms of position VaR

If we now assume any relevant mean terms are zero (or sufficiently close to zero), we can go further and substitute out the β_i and *VaR(p)* terms from Equation (11.6). This gives us a second expression for the IVaR:

$$IVaR_i \approx w_i \beta_i VaR(p) = w_i \rho_i \frac{\sigma_i \sigma_p}{\sigma_p^2} \sigma_p z_\alpha P = \rho_i w_i \sigma_i z_\alpha P = \rho_i VaR(i) \tag{11.7}$$

where *VaR(i)* is the VaR on position *i* considered on its own. Equation (11.7) gives us the IVaR on a position as the product of its stand-alone VaR, *VaR(i)*, and the correlation ρ_i between the return on position *i* and the return on the portfolio. For any given position VaR (denoted by *VaR(i)*), the IVaR is proportional to the correlation coefficient – so a positive correlation implies a positive IVaR, a zero correlation coefficient implies a zero IVaR, and a negative correlation implies a negative IVaR. The IVaR is also proportional to the stand-alone VaR, *VaR(i)*.

Like the previous one, this expression has the advantage that it requires little information (i.e., we only need the position VaR and the single correlation coefficient), very little work and is readily understood. However, it too ignores the impact of the new position on risks within the existing portfolio.

Thus, we have two very straightforward approximations for the IVaR: one in terms of the portfolio VaR, and the other in terms of the stand-alone position VaR. Both these approximations emphasise the importance of correlation in determining the IVaR: in this last case (Equation (11.7)) the dependence of the IVaR on correlation is explicit; in the former case (Equation (11.6)), the dependence is implicit in the beta. Again, all this is consistent with the message of portfolio theory – as we would expect, since these approximations are obtained in a mean-variance (i.e., elliptical) world.

However, these approximations are only reliable for 'small' changes to our portfolio in a world of multivariate ellipticality. If we are faced with non-elliptical distributions, our only practical options are to use the 'before and after' approach (with the calculation problems that entails) or to make use of elliptical IVaR approximations with 'fudge factors' or ad hoc adjustments (e.g., applications of multivariate Cornish–Fisher adjustments, etc.) that we have some reason to think might work in the context of the portfolio we have to deal with.

11.2 COMPONENT VaR

11.2.1 Properties of Component VaR

We turn now to consider the component VaR, CVaR, and we begin by considering the properties that we want CVaR to satisfy. The two main properties we want are:

- *Incrementality*: We want the component VaR to be, at least to a first order of approximation, equal to the IVaR – the increase or decrease in VaR experienced when the relevant component is added to or deleted from the portfolio.
- *Additivity*: We want the arithmetic sum of component VaRs to be equal to the VaR of the total portfolio. This ensures that however we decompose the VaR, all the constituents, the component VaRs, collectively add up to the whole of the VaR.

We can obtain our component VaR as follows. We first select the decomposition criteria – whether we wish to decompose VaR by instrument, asset class, desk, etc. The portfolio VaR will

be a linearly homogeneous function of the positions in the instruments (or asset classes, etc.) concerned.[3] This linear homogeneity allows us to apply Euler's theorem, which tells us that:

$$VaR = \sum_{i=1}^{n} w_i \frac{\partial VaR}{\partial w_i} = \nabla \mathbf{VaR(p)w} \tag{11.8}$$

If we now define the component VaR for instrument i, $CVaR_i$, as:

$$CVaR_i = w_i \frac{\partial VaR}{\partial w_i} \tag{11.9}$$

We now substitute Equation (11.9) into (11.8) to get:

$$VaR = \sum_{i=1}^{n} CVaR_i \tag{11.10}$$

which gives us a breakdown of the VaR into component VaR constituents that satisfies both incrementality and additivity properties.[4] The key to CVaR is thus Equation (11.9), which specifies the $CVaR_i$ in terms of the position sizes (i.e., the w_i) and the marginal VaRs or mathematical first derivatives of the VaR with respect to the w_i.

Component VaRs give us a good idea of the distribution of risks within our portfolio, and they take account of all relevant risk factors, including correlations. This can be seen if we take the CVaR as an approximation to Equation (11.9):

$$CVaR_i \approx w_i \frac{\Delta VaR}{\Delta w_i} = \frac{IVaR_i}{\Delta w_i / w_i} \tag{11.11}$$

This tells us that the CVaR is equal to the IVaR divided by the proportional change in the relative size of position. It also means that the CVaR is approximately proportional to the IVaR, and is therefore affected by correlation or beta in much the same way as the IVaR. And, as with incremental VaR, we can distinguish between three main cases:

- *High contributions to risk*: High CVaRs represent high pockets of risk, which contribute strongly to overall portfolio VaR.
- *Moderate contributions to risk*: Moderate positive CVaRs represent moderate pockets of risk.
- *Negative contributions to risk*: Negative CVaRs represent natural hedges that offset some of the risk of the rest of the portfolio.

As noted already, the CVaR gives us a risk decomposition in nominal (i.e., $) terms: position i contributes so many dollars' risk to our portfolio, and so on, and (as Equation (11.10) shows) the sum of the CVaRs is the portfolio VaR itself.

However, in practice, it is often convenient and transparent to work with the CVaRs in a slightly different, percentage, form: to say that position i contributes $x\%$ of the portfolio VaR, and so on. This representation of CVaR is known as *VaR-beta*, and the VaR-beta is given by:

$$VaR\text{-}beta_i = \frac{CVaR_i}{VaR} \approx \frac{IVaR_i / VaR}{\Delta w_i / w_i} \tag{11.12}$$

[3] A function $y = (x_1, \ldots, x_n)$ is linear homogeneous if multiplying the inputs by some positive constant λ leads to the output multiplying by the same proportion (i.e., $\lambda y = (\lambda x_1, \ldots, \lambda x_n)$).

[4] The latter is obvious; the other is a useful exercise.

and this, as any economist knows, is simply the elasticity of the VaR with respect to the relative position size. The sum of the VaR-betas is of course 1.

We should also keep in mind that the CVaR risk decomposition outlined in Equations (11.8)–(11.12) has an important limitation: it is a linear marginal analysis. The component risks add up to total VaR because of linear homogeneity working through Euler's theorem, but the price we pay for this additivity property is that we have to assume that each component VaR is simply the position size multiplied by the marginal VaR. This is restrictive because it implies that the component VaR is proportional to the position size: if we change the size of the position by $k\%$, then the component VaR will also change by $k\%$. Strictly speaking, this linear proportionality is only guaranteed if each position is very small relative to the total portfolio; and where the position size is significant relative to the total portfolio, then the component VaR estimated in this way is likely, at best, to give only an approximate idea of the impact of the position on the portfolio VaR. If we want a 'true' estimate, we would have to resort to the IVaR, and take the difference between the VaRs of the portfolio with and without the position concerned. The IVaR then gives us an exact estimate of the impact of the portfolio. Unfortunately, this exactness has its price: we lose the additivity property, and the component VaRs no longer add up to equal the total VaR, which makes it difficult to interpret these IVaRs (or CVaRs or whatever else we call them) as true decompositions of the total risk. In short, when positions are significant in size relative to the total portfolio, we can only hope for our CVaRs to give approximate estimates of the effects of the positions concerned on the portfolio VaR.

This can cause problems for risk attribution and capital allocation. If we want to use component VaRs to allocate capital, we want the component VaRs to be accurate and to satisfy additivity, but we can't in general satisfy both conditions. This leaves us with an awkward choice: we can satisfy additivity and base our capital requirements on potentially inaccurate component risk measures; or we can make our IVaR estimates accurate, and then that means that they don't add up properly. In the latter case, we might find that the estimated component VaRs added up to more than the total VaR, in which case a bottom-up approach to capital requirements would leave us excessively capitalised at the firm-wide level; or, alternatively, we might find that the estimated component VaRs added up to less than the total VaR, in which case we would have a capital shortfall at the aggregate level. Either way, we get into overhead allocation problems that have no neat theoretical solution.

If it is essential that our risk decomposition adds up (e.g., as would be the case if we were using decomposed risk estimates to allocate capital), then the only solution in these circumstances is to apply some rule of thumb. For example, we might do an initial 'fair' assessment and discover that there is a shortfall and that the sum of the estimated CVaRs does not add up to the VaR. We could respond by allocating the shortfall across positions using an arbitrary rule (e.g., pro rata to the original CVaR estimates). The resulting risk decomposition would not be exactly fair, but it would still give roughly the right sense of relative risk across the portfolio. There would presumably be some second-order inefficiencies in the risk allocation, but some such effects are more or less unavoidable and should (hopefully) be quite small. We should therefore keep them in perspective, not least because the original risk allocation might be quite inaccurate itself: indeed, it is possible in practice that these second-order effects would get drowned out by the noise in our original risk decompositions, in which case they become academic. Provided the rule of thumb is a reasonable one, we have good reason to hope that adding-up issues might be less of a problem in practice than they appear to be in theory.

Box 11.1 Estimating IVaR and CVaR

We can estimate IVaR (and CVaR) by any of the standard methods: parametric estimation methods, non-parametric (e.g., HS) methods, or simulation methods:

- Parametric methods are appropriate when we can solve for the delVaRs (e.g., as we can for normal VaR).
- We can apply HS methods by using a 'before and after' approach using HS to estimate the 'before' and 'after' portfolio VaRs.
- We can also apply simulation methods using a 'before and after' approach, but this can be inaccurate if the user is not careful. If we run two separate 'before' and 'after' paths, the variance of the IVaR (or CVaR) estimator will behave much like the variance of an option-delta estimator in such circumstances: the variance will get very large as the 'increment' gets small.[5] The solution is to run one set of price paths, and infer the 'before' and 'after' portfolio VaRs from that. The error of this estimator is of order 1, and will therefore get small as the increment gets small. We can also apply simulation methods to estimate the original portfolio VaR and the delVaR terms, and we can then plug these estimates into Equation (11.1) to obtain our estimated IVaRs. These can then be substituted into Equations (11.11) and (11.12) to give us estimates of CVaRs and VaR-betas.

11.2.2 Uses of Component VaR

11.2.2.1 'Drill-down' capability

The additivity of component VaR is, as we have seen, very useful for 'explaining' how VaR can be broken down into constituent components. Yet it also enables us to break down our risks at multiple levels, and at each stage the component risks will correctly add up to the total risk of the unit at the next level up. We can break down the firm-wide risk into component risks associated with large business units (e.g., by country or region); we can break down these in turn to obtain the component risks associated with smaller units (e.g., individual branches); and so on, right down to the level of individual desks or traders. This breakdown is illustrated in Figure 11.3. The key point is that since the component risks correctly add up (even if only approximately and/or we resort to some rule of thumb to make them add up), we can break down our risks to obtain the component VaRs at any level we choose: we can break down our firm-wide VaR into component VaRs at the level of large business units, at the level of smaller units, or at any other level, including the level of the individual desk, the individual trader, or the individual instrument. The additivity of component VaRs therefore gives rise to a 'drill-down' capability – an ability to decompose a risk figure, or identify its components, down to any level we choose. So, for example, an institution might use drill-down to establish how each and every unit, at each and every level – each trader, instrument, asset class, desk, branch, region, or whatever – contributes to overall risk. Drill-down capability is, needless to say, of immense practical usefulness – for determining the positions or units that need attention, identifying hidden sources of risk, setting limits, making investment or trading decisions, determining capital requirements, establishing remuneration schedules, and so on.

[5] See Boyle *et al*. (1997), p. 1304.

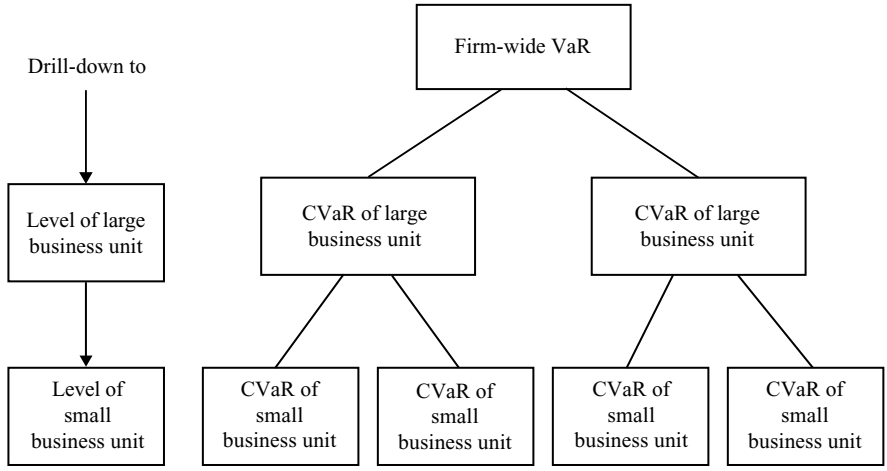

Figure 11.3 Multiple-level risk decomposition and drill-down capability

11.2.2.2 Reporting component VaRs

Given particularly that many component risks are less than obvious, it is very important to report component risks meaningfully, and in ways that interested parties (e.g., senior managers, etc.) can understand without too much difficulty. This suggests that:

- We should 'slice and dice' component VaRs, and report them accordingly, in ways geared to each particular audience, business unit, etc.
- Reports should be as short and straightforward as possible, decomposition information should be as transparent as possible, and reports should avoid unnecessary information that distracts from the key points to be communicated.
- Reports should identify key assumptions and spell out possible consequences if those assumptions are mistaken.

To give an illustration of what such a report might look like, Table 11.1 shows a sample risk decomposition report on a portfolio of stocks invested in each of five sectors – banking, chemicals, energy, utilities and construction.[6] This report is a very basic one, but even so, it conveys a lot of useful information. In this case, the decomposition is by sector, and the report gives the size of the investment in each sector in the second column. Each sector position's stand-alone or undiversified VaR is shown in the third column, and the fourth column gives the component VaRs. In this case, as in many, the component VaRs will be less than the stand-alone VaRs thanks to the diversification of risks within the portfolio. The sum of the component VaRs gives us the VaR of the portfolio taking account of diversification across the positions, and is equal to 30.395. This can be compared to the sum of the undiversified VaRs, which is 36.534, and this latter figure can also be interpreted as the portfolio VaR we would get if there was no diversification and all correlations were 1. The ratio of these two VaRs is 83.19%, and the diversification effect – the extent to which diversification across sectors reduces the portfolio

[6] Aragonés *et al.* (2001) provide some further examples and a good discussion of how they might be interpreted.

Table 11.1 Sample risk decomposition report

Sector	Investment	Undiversified VaR	Component VaR	VaR-beta
Banking	100	5.373	4.343	14.29%
Chemicals	100	10.559	9.463	31.14%
Energy	100	9.063	8.334	27.42%
Utilities	100	5.826	4.350	14.31%
Construction	100	5.714	3.905	12.85%
Sum		36.534	30.395	100.00%
Sum of undiversified VaRs		36.534		
Diversified VaR		30.395		
Ratio		83.19%		
Diversification in portfolio		16.81%		
Key assumption		Risk factors are multivariate normal		
Likely impact if key assumption violated		Risks (VaRs and IVaRs) are underestimated		

VaR – is equal to 100% minus this ratio, or 16.81%. For their part, the VaR-betas vary from 12.85% for construction to 31.14% for chemicals. This tells us that the construction sector contributes the least risk to our portfolio, and the chemicals sector contributes the most. The fact that the VaR-betas are all positive also tells us that none of the investments is a natural hedge. Finally, the report also highlights the key assumption on which this report is based (in this case, multivariate normality) and notes the likely consequences if this assumption is falsified (i.e., in this case, the likely consequence is the underestimation of risks).

Besides decomposing by sector, a CVaR report might decompose by asset class (e.g., equities, commodities, etc.), market risk factors, individual trades or positions, business unit, asset manager or trader, location of desk, tenor (i.e., how risks are allocated over time), types of counterparty (e.g., government counterparties, swap counterparties, etc.), identity of counterparties, and so on. Each of these is good for its own particular purpose.

CVaR and IVaR information can also be presented in the form of 'hot spots', 'best hedges', 'best replicating portfolios' and 'implied views' reports:

- Hot spots reports give the CVaRs ranked in terms of their size – the top-ranking CVaRs are the 'hot spots', or the biggest sources of portfolio risk – and these give a very immediate indication of where the portfolio risks are coming from.
- Best hedges reports give the best hedges – for each instrument or asset, the trade (long or short) that would minimise portfolio VaR. For positions with a negative IVaR, the best hedge would involve a further purchase or investment; for a positive IVaR, the best hedge would involve a sale or short position. Best hedges are very useful benchmarks for portfolio management.[7]
- Best replicating portfolios (BRPs) are those portfolios, made up of small numbers of positions, that best replicate the risks of our 'real' portfolio: we select a small number of assets n, estimate the BRP using regression analysis, and report the BRPs of a range of n-values. Best replicating portfolios are very useful for identifying macro portfolio hedges – hedges against the portfolio as a whole. They also help us to understand the risks we face: if we have a very large portfolio, it can be very difficult to understand what is going on, but if we

[7] See, e.g., Litterman (1996), p. 40.

can replicate the portfolio with one that has only a small number of different assets, we can get a much better picture of the risks involved. BRP reports are therefore especially useful when dealing with very large or very complex portfolios.

- Implied views are the views about future returns that make the current portfolio an optimal one. Comparing implied views about returns with actual views is a useful tool in helping to understand how portfolios can be improved. They are also useful in helping to macro manage a portfolio whose composition is subject to delegated decision-making. A good example, suggested by Litterman, is in big financial institutions whose portfolios are affected by large numbers of traders operating in different markets: at the end of each day, the implied views of the portfolio can be estimated and compared to the actual views of, say, in-house forecasters.[8] Any differences between actual and implied views can then be reconciled by taking positions to bring the implied views into line.

11.3 DECOMPOSITION OF COHERENT RISK MEASURES

In our earlier discussions we assumed for convenience that our risk measure was the VaR. However, we can (and hopefully would) apply similar analysis to coherent risk measures. For example, we can always define incremental coherent risk measures along lines similar to Equation (11.1), and under reasonable conditions a coherent risk measure can be decomposed into its marginal equivalents using Euler's theorem. For example, if we take our risk measure to be the ES, then Euler's theorem tells us that:

$$ES = \sum_{i=1}^{n} w_i \frac{\partial ES}{\partial w_i} \tag{11.13}$$

Following Tasche (2000), we can also show that the marginal ES is:

$$\frac{\partial ES}{\partial w_i} = E[X_i | X > VaR] \tag{11.14}$$

This tells us, in turn, that the component ES, CES_i, is:

$$CES_i = w_i E[X_i | X > VaR] \tag{11.15}$$

Thus, the decomposition of coherent risks is fairly similar to the decomposition of VaR considered earlier.[9]

[8] Litterman (1996), p. 41.

[9] For more on the decomposition of coherent risk measures, see, e.g., Tasche (2000), Yamai and Yoshiba (2001a), or Fischer (2003).

12

Mapping Positions to Risk Factors

Portfolio returns (or P/L) are derived from those on individual positions, and we have tended to take for granted up to now that we are able to model the latter directly: we have assumed that each position i has a return r_i, and that we can directly model the process that generates r_i. However, it is not always possible or even desirable to model each and every asset return in this direct manner. To appreciate the issues involved we must first distinguish between individual *positions* (or instruments) and individual *risk factors*: each position consists of a collection of instruments of the same type, but our risk factors are the stochastic (i.e., risky) variables that determine their returns. So far, we have usually assumed that each position had its 'own' risk factor (or factors), and could be projected onto it (or them). However, in practice, we will often project our positions, not onto 'their' own individualised risk factors, as such, but onto some set of benchmark risk factors. Typically, we would project our n individual instruments onto some smaller number, m, of reference instruments, and then our risk factors are the factors that drive the returns to our reference instruments. This requires that we describe our positions in terms that relate to those reference instruments, i.e., we treat our real positions as approximate combinations of standard building blocks. This process of describing our 'real' positions in terms of these standard building blocks (or underlying reference risk factors) is known as mapping.

There are four main reasons why we might want to engage in mapping. The first is that we might want to map our positions because we do not have enough data on them. We might have an emerging market instrument that has a very short track record, and this means that we don't have enough data on it; or we might have a new type of derivatives instrument that has no track record at all. In such circumstances we might map our position to some comparable position for which we do have sufficient data.

A second reason for mapping is to cut down on the dimensionality of our covariance matrices, and this would be a major consideration if we had a variance–covariance-based risk model. If we have n different instruments in our portfolio, we would need data on n separate volatilities, one for each instrument, plus data on $n(n-1)/2$ correlations – a total altogether of $n(n+1)/2$ pieces of information. As new instruments are added to our portfolio, the additional amount of correlation data needed grows geometrically. As n gets large, the amount of data needed becomes enormous, and it becomes increasingly difficult to collect and process the data involved. For practical purposes, there is therefore a limit on the size of covariance matrix we can handle. In any case, as explained elsewhere (see, e.g., Chapter 4, Appendix 4 and Chapter 5), we would not normally want to work with very high dimension matrices, even if we had the data to do so: we need to keep the dimensionality of our covariance matrix down to reasonable levels to avoid computational problems.

The third reason is closely related to the previous one: if we try to handle risk factors that are closely correlated (or worse, perfectly correlated), there is a danger of running into rank problems with the covariance matrix. Either our algorithms will not work or – if we are really unlucky – they will work but produce pathological estimates without our necessarily realising

it. If we are to avoid these problems, we have to ensure that our risk factors are not too closely related to each other, and this requires that we select an appropriate set of risk factors and map our instruments onto them.

A final reason for engaging in mapping is that it can greatly reduce the time needed to carry out risk calculations: reducing a portfolio with a very large number of diverse positions to a consolidated set of risk-equivalent positions in basic risk factors allows for much faster calculations, albeit at the cost of some loss of precision.

The process of mapping generally involves three stages. The first is to construct a set of benchmark instruments or factors and collect data on their volatilities and correlations. Benchmark instruments might include key bonds, equities, commodities, and so on. Having established a set of benchmark instruments or factors and collected the necessary data, we derive synthetic substitutes for each instrument we hold, made up of positions in the core instruments. This synthetic substitution is the actual mapping. The final stage is to estimate risk measures using the mapped instruments (i.e., the synthetic substitutes) instead of the actual instruments we hold. Put differently, we pretend that we are holding the synthetic portfolio composed only of benchmark instruments, and we estimate its risk measures, which we take to be estimates of the 'true' risk measures we are seeking.

12.1 SELECTING CORE INSTRUMENTS

The usual approach in mapping is to select a set of reference instruments – key money market and equity instruments, key currencies, etc. – that can be regarded as representative of the broad types of instruments actually held. The ideal is to have a rich enough set of reference instruments to be able to provide good proxies for the instruments in our portfolio, while not having so many reference instruments that we run into the high-dimensionality and related problems that we wish to avoid.

The best-known systems – those of the type pioneered by RiskMetrics[1] – use the following set of reference instruments, or something similar:

- Equity positions are represented by equivalent amounts in terms of equity indices in each of the core currencies.
- Fixed-income positions are represented by combinations of cash flows of a limited number of specified maturities in a given currency.[2]
- Foreign exchange (FX) positions are represented by the relevant amounts in terms of a certain number of 'core' currencies, and FX forward positions are mapped as equivalent fixed-income positions in their respective currencies.
- Commodity positions are represented by amounts of selected standardised futures contracts traded on organised exchanges.

RiskMetrics uses a broad set of core instruments to map a correspondingly broad range of different positions. However, most institutions have more specialised portfolios and would therefore work with some subset of the RiskMetrics core instruments to reduce dimensionality

[1] The reader who wants further information on the RiskMetrics approach to mapping is referred to Phelan (1997) or the RiskMetrics *Technical Document* (1996, Chapter 6.2), which both contain extensive discussions of the RiskMetrics mapping system and the issues behind it.

[2] In the RiskMetrics model, positions are also differentiated by their credit standing, i.e., government (which is assumed to be free of default risk) and non-government (which is not). This categorisation obviously fails to do any justice to credit risk, but we can always make more serious adjustments for credit risk if we wish to do so.

problems and speed up calculations. If they wanted, they could also add new cores of their own (e.g., fixed-income instruments with new maturities or sector-specific equity indices).

Box 12.1 Mapping with Principal Components

We can also map using core factors identified by principal components analysis (PCA). As explained in Chapter 4, Appendix 4, principal components are hypothetical variables that are constructed to 'explain' the movements in a group of time series. These series in our case would be a set of prices or returns. We will usually find that a small number of principal components is sufficient to explain a very large proportion of the movement in our price or return series. These procedures can therefore cut down drastically on the dimensionality of our system.

PCA procedures are generally suited to portfolios with a large number of different instruments that tend to be closely correlated with each other, reflecting the presence of common influences. Perhaps the best examples are portfolios of money market instruments and bonds, which typically show very high degrees of correlation with each other. In such cases, PCA offers enormous potential benefits in terms of parameter reduction. For example, if we had, say, 50 different instruments in a bond portfolio and were to work with those instruments directly (i.e., without mapping), we would need to handle a 50×50 covariance matrix with $50(51)/2 = 1275$ separate volatility and correlation parameters. But if we used PCA, we could probably proxy the overwhelming proportion of bond price movements by three principal components, and the only variance–covariance parameters needed would be the volatilities of the three principal components. We would reduce the number of variance–covariance parameters needed from 1275 to only three! Once we have our principal components, each individual instrument can then be mapped to these principal components and treated as if it was a linear combination of them.

12.2 MAPPING POSITIONS AND VaR ESTIMATION

We now consider how to map and estimate VaR for specific types of position. Although there is a huge variety of different financial instruments, the task of mapping them and estimating their VaRs can be simplified tremendously by recognising that most instruments can be decomposed into a small number of more basic, primitive instruments. Instead of trying to map and estimate VaR for each and every specific type of instrument, all we need to do is break down each instrument into its constituent building blocks – a process known as reverse engineering – to give us an equivalent portfolio of primitive instruments. We then map this portfolio of primitive instruments.

12.2.1 Basic Building Blocks

12.2.1.1 Basic FX positions

There are four types of basic building block, and the easiest of these are basic FX positions (e.g., holdings of non-interest bearing foreign currency). FX positions are particularly simple to handle where the currencies involved (i.e., our own and the foreign currency) are included as core currencies in our mapping system. We would then already have the exchange rate

volatilities and correlations that we require for variance–covariance analysis. If the value of our position is x in foreign currency units and the exchange rate is E, the value of the position in domestic currency units – or the mapped position, if you like – is xE. Since there is no foreign interest rate being paid, x is constant, and so the only risk attaches to E.

We can then calculate the VaR (or any other risk measure) in the usual way. For example, if we assume for the sake of argument that the exchange rate is normally distributed with zero mean and standard deviation σ_E over the period concerned, then the VaR is

$$VaR^{FX} = xE\sigma_E z_\alpha \tag{12.1}$$

The VaR is the size of the position in domestic currency units (xE) times the standard deviation of the exchange rate (σ_E) times the standard normal variate (z_α).

Where currencies are not included as core currencies, we need to proxy them by equivalents in terms of core currencies. Typically, non-core currencies would be either minor currencies (e.g., the Hungarian forint) or currencies that are closely tied to some other major currency (e.g., as the Dutch guilder was tied very closely to the mark). Including closely related currencies as separate core instruments would lead to major collinearity problems: the variance–covariance matrix might fail to be positive definite, etc. The mapping of non-core currencies to baskets is much the same in principle as the mapping of individual equities to equity indices as described in the next section.

12.2.1.2 Basic equity positions

The second type of primitive position is equity. Imagine we hold an amount x_A invested in the equity of firm A but lack this particular firm's volatility and correlation data. However, we can reasonably suppose that the firm's return to equity, R_A, is related to the equity market return, R_m, by the following sort of condition:

$$R_A = \alpha_A + \beta_A R_m + \varepsilon_A \tag{12.2}$$

where α_A is a firm-specific constant, β_A is the market-specific component of the firm's return, and ε_A is a firm-specific random element. The variance of the firm's return is then:

$$\sigma_A^2 = \beta_A^2\sigma_m^2 + \sigma_\varepsilon^2 \tag{12.3}$$

where σ_A^2 is the variance of R_A, and so on. The variance of the firm's return therefore consists of a market-based component $\beta_A^2\sigma_m^2$ and a firm-specific component σ_ε^2. Again assuming zero-mean normality for the sake of argument, the VaR of the equity position is then:

$$VaR_A = x_A\sigma_A z_\alpha = x_A\sqrt{\beta_A^2\sigma_m^2 + \sigma_\varepsilon^2}\, z_\alpha \tag{12.4}$$

Estimates of both σ_m^2 and β_A should be publicly available, so we can easily estimate $\beta_A^2\sigma_m^2$. If we also have data on the firm-specific variance σ_ε^2, we can estimate Equation (12.4) directly, and all is well and good.

But what do we do if we don't have information on σ_ε^2? The answer depends on how well diversified our portfolio is but if our portfolio is well diversified, the firm-specific risks will largely net out in the aggregate portfolio and we could estimate VaR as if σ_ε^2 were zero. Our VaR would then be:

$$VaR_A \approx x_\alpha\beta_A\sigma_m z_\alpha \tag{12.5}$$

In short, we map the equity return to the market index, and we use Equation (12.5) to estimate the VaR of the mapped equity return. It is important to note that the only volatility information used is the market volatility, and the only firm-specific information we need is the firm's market beta. This 'market beta' mapping allows us to approximate equity VaRs with very little firm-specific information.

The market beta approach also extends naturally to multi-asset equity portfolios, and enables us to estimate the VaR of such portfolios without needing to bother with covariance matrices. To illustrate, if we have a portfolio of two equities, A and B, and again assume that their expected returns are zero, then the VaR of our mapped portfolio would be:

$$VaR = x_A\beta_A\sigma_m z_\alpha + x_B\beta_B\sigma_m z_\alpha = (x_A\beta_A + x_B\beta_B)\sigma_m z_\alpha \tag{12.6}$$

There is no covariance matrix because both equities are mapped to the same single risk factor, namely, the market risk. This approach is highly convenient – to say the least – when dealing with equity portfolios. Instead of needing volatilities and correlations across all stocks in our portfolio, it means that we can get by with only market beta information for each stock. Thus, the market beta approach represents a huge reduction in both the information and the computations needed to estimate a VaR.

The only real problem with this 'market beta' mapping procedure is that if we hold an undiversified portfolio, then the estimate of VaR given by Equation (12.5) will understate true VaR because it ignores the firm-specific risk. However, if we wish to do so, we can refine the VaR estimate by using an adjustment – explained in Box 12.2 – that takes account of the extent to which the portfolio is imperfectly diversified.

Box 12.2 Adjusting Equity VaRs for Firm-specific Risk

A drawback with the market beta approach to estimating equity VaRs is that it can underestimate VaR because it ignores firm-specific risk. One way to adjust for this bias is to multiply the beta-based VaR (i.e., $x_A\beta_A\sigma_m z_\alpha$) by an adjustment factor that reflects the degree to which the portfolio is *imperfectly* diversified. This adjustment factor is

$$\phi + \frac{(1-\phi)\sigma_p^u}{\beta_A\sigma_m}$$

where σ_p^u is a hypothetical portfolio variance based on the assumption that risks are completely undiversified (i.e., perfectly correlated), and ϕ is a diversification index given by

$$\phi = \frac{\sigma_p^u - \sigma_p}{\sigma_p^u - \sigma_m}$$

where σ_p is an estimate of the portfolio variance as it is. If the portfolio is perfectly diversified, then $\sigma_p = \sigma_m$ and $\phi = 1$. The adjustment factor is therefore also 1, and the VaR is exactly as given in Equation (12.5). At the other extreme, if the portfolio is not diversified at all, then $\sigma_p = \sigma_p^u$ and $\phi = 0$. The adjustment is now $\sigma_p^u/(\beta_A\sigma_m)$ and the VaR is $x_A\sigma_p^u z_\alpha$, which is easily verified as the correct expression in the absence of any diversification. Finally, if the portfolio is imperfectly diversified, ϕ takes a value between 0 and 1 and we get a VaR somewhere between $x_A\beta_A\sigma_m z_\alpha$ and $x_A\sigma_p^u z_\alpha$. Our adjustment factor leads to correct VaRs at the extremes of perfect and zero diversification, and makes some allowance for imperfectly diversified portfolios in the middle.

It only remains to find ways of estimating σ_p^u and σ_p when we have very little information about the volatilities and correlations of specific assets. However, if the portfolio is totally undiversified, then σ_p^u is just the average standard deviation of the individual assets. We can therefore estimate σ_p^u by taking such an average. Estimating σ_p is only slightly more involved. The portfolio variance can be written as

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}^2$$

where w_i is the weight of stock i in the portfolio. If we now assume for convenience that the portfolio is equally weighted in the different assets (i.e., $w_i = 1/n$), then the portfolio variance becomes

$$\sigma_p^2 = \frac{\bar{\sigma}_i^2}{n} + \frac{(n-1)\bar{\sigma}_{ij}}{n}$$

which means that we can approximate the portfolio variance if we have data on the average variance and average covariance of the individual stocks in our portfolio.[3]

12.2.1.3 Zero-coupon bonds

The third type of primitive instrument is a zero-coupon bond, known as a zero. Assuming for convenience that we are dealing with instruments that have no default risk, our task is then to map a default-free zero-coupon bond against a set of default-free reference instruments. Let us also suppose that we are using a set of core or reference assets of the RiskMetrics sort: these might be 1-month zeros, 3-month zeros, and so on. This means that we will have volatility and correlation data on zeros of these particular maturities, but not on zeros of other maturities. We will therefore lack volatility and correlation data for the particular zeros we hold, except in the fortuitous special cases where our instruments happen to be reference ones. We might hold zeros with a maturity of 10 months, say, and yet the nearest reference bonds might be 3-month and 12-month zeros. How then do we estimate the VaR of a 10-month zero?

The answer is that we estimate the VaR of the 10-month zero by taking it to be the same as the VaR of its mapped equivalent – that is to say, from its equivalent in terms of some combination of 3-month and 12-month zeros – and we can estimate the latter because we have the volatility and correlation parameters for these bonds.

There are various approaches we could take to map our 10-month bond. For example, the 4th edition of the *RiskMetrics Technical Document* (1996) suggests as mapping criteria that the mapped position should have the same value and same variance as the old one, and should have cash flows of the same sign.[4] Alternatively, we might map our bond using the criteria that the mapped position should have the same value and same duration as the original one. A closely related set of criteria is that the mapped bond has the same value and same PV01

[3] There are also other ways to reduce the degree of understatement of VaR that arises from a portfolio being imperfectly diversified. One solution is to use more than one price index. Instead of using the overall stock market index, we might use more specific indices such as those for energy stocks, manufacturing stocks and so on. These sector-specific indices would pick up more of the movement in each stock return, and thereby reduce the amount of risk left unaccounted for as firm specific. σ_ε^2 would then be lower relative to σ_m^2 and so result in a smaller understatement of VaR.

[4] See *RiskMetrics Technical Document* (1996), pp. 117–121.

(i.e., present value of a basis point change) as the old one. To illustrate the mechanics of fixed-income mapping, let us suppose that we adopt this latter set of mapping criteria, and demonstrate how the mapping takes place.[5]

We begin by taking the 10-month spot rate, r_{10}, to be the linear interpretation of the 3-month and 12-month spot rates, r_3 and r_{12}, with the weights given by the relative maturities. For example, if $r_3 = 4.5\%$ and $r_{12} = 5\%$, then r_{10} will be

$$r_{10} = \left(\frac{12-10}{12-3}\right) \times 4.5\% + \left(\frac{10-3}{12-3}\right) \times 5\% = 4.88889\% \tag{12.7}$$

We now shock r_3 and r_{12} by increasing them each in turn by 1 basis point, and infer the impact of each shock on r_{10}. This gives us two sets of changes to r_{10}, one corresponding to the change in the left-reference rate (r_3) and the other corresponding to the change in the right-reference rate (r_{12}). Using obvious notation, these turn out to be:

$$\Delta r_{10}^{L} = 0.00222\% \text{ and } \Delta r_{10}^{R} = 0.00778\% \tag{12.8}$$

Let us suppose for the sake of argument that our original cash flow is \$1 million at 10 months. The impact of each of the changes on the present value of the \$1 million cash flow is then:

$$PV01_3 = 1\,000\,000\{\exp\left[\left(r_{10} + \Delta r_{10}^{L}\right) 3/12\right] - \exp[-r_{10}3/12]\} = -17.78 \tag{12.9a}$$

$$PV01_{12} = 1\,000\,000\{\exp\left[\left(r_{10} + \Delta r_{10}^{L}\right) 10/12\right] - \exp[-r_{10}10/12]\} = -62.23 \tag{12.9b}$$

We now want to find the reference point cash flows C_3 and C_{12} that have the same PV01s as Equation (12.9). Using basic fixed-income algebra, these are:

$$C_3 = -\frac{-17.78}{\exp[-(r_3 + 0.01)3/12] - \exp[-r_3 3/12)]} = \$719\,217 \tag{12.10a}$$

$$C_{10} = -\frac{-62.23}{\exp[-(r_{10} + 0.01)10/12] - \exp[-r_{10}10/12)]} = \$654\,189 \tag{12.10b}$$

Thus, \$1 million at 10 months has the same sensitivity to changes in the 3-month and 12-month spot rates as two cash flows of \$719 217 at 3 months and \$654 189 at 12 months. In other words, the original cash flow of \$1 million in 10 months' time maps to these latter two cash flows, which can therefore be considered its mapped equivalent.

The steps involved in the calculation process are summarised in Table 12.1, which are based on the Excel spreadsheet 'Mapping and VaR of zero coupon bond' which is included in the book's CD.

We now use our volatility and correlation information for the 3-month and 12-month cash flows to obtain the VaR. For example, suppose that the daily 3-month and 12-month volatilities are, say, 1.25% and 1%, and that their correlation is 0.8. If we continue to assume zero-mean (but in this case multivariate) normality, then straightforward calculations show that the VaR is \$650.88. Again, the calculations are shown in the Excel spreadsheet.

[5] The following discussion is based on that in Dowd and Rowe (2004), which also discusses mapping in more detail and provides a number of additional explicit examples. I would like to thank Dave Rowe and the editors of the volume in which that chapter appeared, Carol Alexander and Elizabeth Sheedy, for their contributions to it.

Table 12.1 Mapping a zero to reference cash flow points

Reference or payment horizon	Left reference horizon	Payment horizon	Right reference horizon
Time forward in months	3	10	12
Time forward in years	0.25	0.83	1.00
Zero rates			
Observed zero rates	4.50%		5.00%
Interpolated zero rate for cash flow date		4.88889%	
Impact of 1 bp change in reference rates			
Impact of change in left-reference rate	0.01000%	0.00222%	
Impact of change in right-reference rate		0.00778%	0.01000%
Bond cash flows			
Principal		$1 000 000	
Present value of total bond cash flow		$960 078.01	
Impact of 1 bp change on present values			
Impact of change in left-reference rate		−$17.7791	
Impact of change in right-reference rate		−$62.2253	
Impact on cash flows mapped to reference dates	−$17.7791		−$62.2253
Mapped cash flows	**$719 217**		**$654 189**
Actual cash flows for comparison		$1 000 000	

12.2.1.4 Basic forwards/futures

The fourth building block is a forward/futures position. As any financial engineering textbook will explain, a forward contract is an agreement to buy a particular commodity or asset at a specified future date at a price agreed now, with the price being paid when the commodity/asset is delivered; and a futures contract is a standardised forward contract traded on an organised exchange. There are a number of differences between these contracts, but for our purposes these differences are seldom important. We can therefore run the two contracts together and speak of a generic forward/futures contract.

To illustrate what is involved for VaR estimation, suppose we have a forward/futures position that gives us a daily return dependent on the movement of the end-of-day forward/futures price. If we have x contracts each worth F, the value of our position is xF. If F is normal with standard deviation σ_F, the VaR of our position is approximately:[6]

$$VaR \approx xF\sigma_F z_\alpha \tag{12.11}$$

assuming again for convenience that the expected return is zero.

In this case, the mapping boils down to the allocation of F to reference points, on the one hand, and the estimation of σ_F, on the other. Typically, we would solve these problems using some linear interpolation.[7]

[6] One reason for the approximation is that, with either contract, the investor is likely to face collateral or margin requirements, and the cost of maintaining these margin positions will usually be interest sensitive. With forward markets, a second source of approximation is the illiquidity of secondary forward markets. A forward VaR is based on a price in a thin market, and any estimated forward price/VaR is subject to considerable liquidity risk.

[7] The forward/futures position is easier to map than the fixed-income one because we have assumed here that our generic forward/futures position does not involve spot-rate (or time-value) issues. This may not always be appropriate, but we can always modify it to take such issues into account. To do so, we would carry out the mapping in much the same way as with fixed-income positions.

12.2.2 More Complex Positions

Having set out our building blocks, we can now map more complex positions. We do so by reverse engineering them – we produce synthetic equivalents for them, in terms of positions in our primitive building blocks.

12.2.2.1 Coupon-paying bonds

A coupon bond can be regarded as a portfolio of zero-coupon bonds, each maturing on a different maturity date. We can therefore map coupon bonds by regarding them as portfolios of zero-coupon bonds and mapping each individual zero-coupon bond separately. The VaR of our coupon bond is then equal to the VaR of its mapped equivalent portfolio in zero-coupon bonds. In other words, we map each coupon payment to adjacent reference points (as with the single zero just considered), and then regard the coupon bond as a portfolio of zeros, each of which has a different maturity.

An example is given in Table 12.2, based on the Excel workbook 'Mapping and VaR of coupon bond' which is also included in the book's CD. In this example we have a bond which makes payments at horizons of 4, 10, 16, ..., 40 and 46 months hence. We have reference horizons of 3, 12, 24, 36 and 48 months. The bond has a face value of $1 million and a 6% annualised coupon rate: this means that it makes payments of $30 000 for each payment date except the last, when it makes a final payment of $1 030 000 as it repays the principal as well. The payments at 4 and 10 months must therefore be mapped to the reference points at 3 and 12 months, the payments at 16 and 22 months must be mapped to reference payments at 12 and 24 months, and so on. Given a set of reference horizon spot rates, the table shows how the mapped payments are obtained. These turn out to be: $56 592 at 3 months, $77 127 at 12 months, $48 088 at 24 months, $258 256 at 36 months, and $843 749 at 48 months.

We then use a variance–covariance approach to estimate the VaR of the portfolio of zeros along standard portfolio-theory lines.

12.2.2.2 Other standard instruments

Other instruments can be mapped using the same principles, which draw on established results in financial-engineering theory:

- *Forward rate agreements*: An FRA is equivalent to a portfolio long in a zero-coupon bond of one maturity and short in a zero-coupon bond of a different maturity. We can therefore map an FRA and estimate its VaR by treating it as a long–short combination of two zeros of different maturities.
- *Floating-rate instruments*: Since a floating-rate note reprices at par with every coupon payment, we can think of it as equivalent to a zero-coupon bond whose maturity is equal to the period until the next coupon payment. We can therefore map a floating-rate instrument by treating it as a zero-coupon bond that matures on the next coupon date.
- *Vanilla interest-rate swaps*: A vanilla interest-rate swap is equivalent to a portfolio that is long a fixed-coupon bond and short a floating-rate bond, or vice versa, and we already know how to map these instruments.
- *Structured notes*: These can be regarded as a combination of interest-rate swaps and conventional floating-rate notes, which we can already map.[8]

[8] There is a very small amount of embedded optionality in basic structured note instruments, which we can ignore. However, more sophisticated structured notes can have a great deal of optionality, and such positions can only be dealt with properly by coming to terms with their embedded options.

Table 12.2 Mapping of coupon bond

Reference or payment horizon	Reference horizon	Payment horizon	Payment horizon	Reference horizon	Payment horizon	Payment horizon	Reference horizon	Payment horizon	Payment horizon	Reference horizon	Payment horizon	Payment horizon	Reference horizon
Time to cash flow in months	3	4	10	12	16	22	24	28	34	36	40	46	48
Time to cash flow in years	0.25	0.33	0.83	1	1.33	1.83	2	2.33	2.83	3	3.33	3.83	4
Zero rates													
Observed zero rates	4.50%			5.00%			6.00%			6.50%			7.00%
Interpolated zero rates for cash flow dates		4.5555%	4.8889%		5.3333%	5.8333%		6.1667%	6.4167%		6.6667%	6.9167%	
Impact of 1 bp change in reference rates													
Impact of a 01 change in left-reference rate	0.0100%	0.0089%	0.0022%	0.0100%	0.0067%	0.0017%	0.0100%	0.0067%	0.0017%	0.0100%	0.0067%	0.0017%	0.0100%
Impact of a 01 change in right-reference rate		0.0011%	0.0078%		0.0033%	0.0083%		0.0033%	0.0083%		0.0033%	0.0083%	

Bond cash flows ($)												
Coupon		30 000	30 000		30 000	30 000		30 000	30 000		30 000	30 000
Principal												1 000 000
Total		30 000	30 000		30 000	30 000		30 000	30 000		30 000	1 030 000
Present value of cash flow		29 547.9	28 802.3		27 940.8	26 957.3		25 979.5	25 012.9		24 022.1	790 112.4
Impact of 1 bp change on present values												
Impact of change in left-reference rate		−0.8755	−0.5334		−1.2418	−4.1182		−4.0409	−1.1811		−5.3377	−50.4778
Impact of change in right-reference rate		−0.1094	−1.8668		−2.4835	−0.8237		−2.0206	−5.9051		−2.6690	−252.3567
Impact on cash flows mapped to reference dates	−1.41			−7.34			−8.53			−63.74			−255.03
Mapped cash flows	**56 992**			**77 127**			**48 088**			**258 256**			**843 749**
Actual cash flows for comparison		30 000	30 000		30 000	30 000		30 000	30 000		30 000	1 030 000

- *FX forwards*: A foreign-exchange forward is the equivalent of a long position in a foreign-currency zero-coupon bond and a short position in a domestic currency zero-coupon bond, or vice versa.
- *Commodity, equity and FX swaps*: These can be broken down into some form of forward/futures contract on the one hand, and some other forward/futures contract or bond contract on the other.

We could extend this list ad nauseam, but the main point is very clear: subject to a reservation to be discussed below, we can map a great many different instruments using a small number of building blocks and some elementary financial engineering theory.

And what is the reservation? Optionality. The instruments just covered all have in common that their returns are linear or nearly linear functions of the underlying risk factors. Mapping is then fairly straightforward, and our results should be fairly reasonable. Of course, we should always recognise that mapping involves approximation and hence error, but in the presence of linear risk factors there is no reason to be unduly concerned about such errors provided we are reasonably conscientious. But once we have significant optionality in any of our positions, our positions can (and typically will) become highly non-linear functions of the underlying risk factors, and this non-linearity can seriously undermine the accuracy of any standard (i.e., linear) mapping procedure.

To map options positions more accurately, we could use non-linear or delta–gamma approximations, which are discussed further in Chapter 10. In principle, such methods can be used whenever an option value can reasonably be approximated by a second-order or delta–gamma approximation. However, in practice, such approximations can be difficult to implement, and their accuracy can be problematic. Furthermore, the various approximations available can give quite different results, so when using them, we need to ensure that the approximation(s) we use are suitable for the positions we have.

13

Stress Testing

This chapter examines stress testing – procedures that attempt to gauge the vulnerability of our portfolio to hypothetical events. Financial institutions have used stress testing in some form or other for many years, particularly for gauging their exposure to interest-rate risk. Early stress tests were often little more than 'back of the envelope' exercises, but the methodology has improved over the years, thanks in large part to improvements in spreadsheet technology and computing power. Though still limited in many respects, modern stress testing is much more sophisticated than its predecessors.

There has in the past often been a tendency to see stress testing as secondary to other methods of risk estimation, such as those based on Greek parameter estimation in the derivatives field or VaR more generally. This is due in part to the fact that the methodology of stress testing is inherently primitive. In addition, prior to 1996 most stress testing was done at the desk level, and relatively few firms carried out stress testing at the corporate-wide level. However, since then, corporate-wide stress testing has become much more common and more sophisticated, and stress tests are now routinely applied to credit and liquidity shocks, as well as market ones. As Schachter points out,

> The events of October 1997 represent a watershed of sorts for stress testing. The attention given to stress tests by regulators and banks has increased dramatically since that event. In some respects, the event has kindled a love affair with stress testing. Yet the theory behind stress testing is still ill developed, more so than value at risk, which itself is an immature risk management tool.[1]

So stress testing is now getting much more attention, fuelled in large part by the belated recognition that good stress testing might have helped institutions to avoid some of the painful losses of recent years.

Stress testing has also received strong endorsements from other sources. One of these is the regulatory system, which has long recognised the importance of stress testing. An example is the Amended Basel Accord (1996), which required that banks that seek to have their capital requirements based on their internal models should have in place a 'rigorous and comprehensive' stress testing programme, and this programme should include tests of the portfolio against past significant disturbances and a bank's own stress tests based on the characteristics of its portfolio. Stress tests have also been used by organised exchanges to set their margin requirements, a good example of which is the SPAN stress testing system used by the Chicago Mercantile Exchange (see Chapter 2).

A little more recently, stress testing has also received a major boost from a very unexpected direction – the theory of coherent risk measures. As we discussed in Chapter 2, this theory tells us that the outcomes of stress tests can be regarded as coherent risk measures, and this means that stress-test outcomes are now 'respectable' risk measures. Since the same theory also tells us that VaR is *not* a 'respectable' risk measure, this signals a remarkable reversal in the standings of the two measures. Ten years ago, VaR ruled the roost and stress tests were regarded

[1] Schachter (1998), p. 5F-10.

as very much secondary. The outcomes of stress tests were useful, but had no real intellectual foundation, whereas VaR was a 'proper' risk measure. Now it transpires that all this was the wrong way round: it turns out that VaR has no right to be regarded as a true risk measure at all, and stress-test outcomes are revealed to have been 'respectable' risk measures all along. VaR is thus revealed as a fake risk measure, and stress tests at last receive the recognition they deserve.

Stress testing is particularly good for quantifying what we might lose in crisis situations where 'normal' market relationships break down and VaR and ES risk measures can be very misleading. Stress tests can identify our vulnerability to a number of different crisis phenomena:

- *Breakdowns in 'normal' correlation relationships*: In crises, correlations often swing to extreme values, and losses can be much greater than suggested by VaR estimates based on 'normal' correlation assumptions.
- *Sudden decreases in liquidity*: Markets can suddenly become very illiquid in crisis situations, bid–ask spreads and order-execution times can increase dramatically, and risk management strategies (e.g., such as those based on dynamic trading) can become unhinged, leading to much bigger losses than anticipated.
- *Concentration risks*: Stress tests can sometimes reveal that we might have a much larger exposure to a single counterparty or risk factor than we had realised, taking into account the unusual conditions of a crisis. VaR or ES measures can overlook such concentration because they tend not to pay much attention to crisis conditions.
- *Macroeconomic risks*: Stress tests are also better suited for gauging our exposure to macroeconomic factors such as the state of the business cycle, the economic condition of a particular country, and so on.

Although the principles behind stress testing are straightforward, there is a huge variety of different categories of stress test depending on the *type of event* (i.e., normal, extreme, contingent, sea change, liquidity, etc.), the *type of risk* involved (market risk, liquidity risk, credit risk, and combinations of these risks), the *risk factors* (e.g., equity risks, yield curve risks, FX risks, default risks, etc.), the *country or region* (e.g., N. America, Japan, etc.), the stress test *methodology* (i.e., scenario analysis, factor push, maximum loss optimisation, etc.), the *model assumptions* (e.g., relating to yield curves, the distributions of risk factors, default parameters, etc.), the *book* (i.e., trading book, banking book, off-balance-sheet), the *instruments* concerned (e.g., basic equities or bonds, futures, options, etc.), the *level* of the test (desk level, business unit level, or corporate level), data requirements (e.g., desk-level data, corporate-wide data, etc.) and the *complexity of our portfolio*. Stress testing is thus simple in principle but complex in practice.

In some ways stress testing is also a natural complement to probability-based risk measures such as VaR and ES. Recall that the VaR gives us the maximum likely loss at a certain probability, but gives us no idea of the loss we might suffer if we experience a loss in excess of VaR. ES is a little better because it gives us the expected value of a loss in excess of VaR, but even ES tells us nothing else about the distribution of 'tail losses' other than its expected value. By contrast, stress testing can give us a lot of information about bad states – and, indeed, stress testing is explicitly *designed* to give us information about losses in bad states. However, stress testing does not (usually) tell us, and is not as such designed to tell us, the *likelihood* of these bad states. So VaR and ES are good on the probability side, but poor on the 'what if' side, whereas stress tests are good for 'what if' questions and poor on probability questions. The two approaches (i.e., the probabilistic risk measures, the VaR and ES, on the one side, and stress tests, on the other) are natural complements, each highlighting what the other tends to miss.

Broadly speaking, we can distinguish between two main approaches to stress testing:

- Scenario (or 'what if') analyses, in which we evaluate the impact of specified scenarios (e.g., such as a particular fall in the stock market) on our portfolio. The emphasis is on specifying the scenario and working out its ramifications.
- Mechanical stress tests, in which we evaluate a number (and often a large number) of mathematically or statistically defined possibilities (e.g., such as increases and decreases of market risk factors by a certain number of standard deviations) to determine the most damaging combination of events and the loss it would produce.

We will consider these presently, but we begin by looking at the benefits and difficulties of stress testing.

Box 13.1 Uses of Stress Tests

In the right hands, stress testing can be a very useful risk management tool, and stress tests can be used for risk management in various ways. The first is as a source of information, with the results of stress tests disseminated to all interested parties. Stress test results can be a particularly effective means of communicating risk information because the underlying conceptual experiment – i.e., what if . . . happens? – is easy to understand and free of any dependence on the probability notions that are inescapable when using probabilistic risk measures. However, it is important not to swamp recipients with unnecessary data, so it is best to give each level of manager or decision-maker only the stress test information relevant to them. When used in this way, stress tests can help to assess risks in the context of the firm's risk appetite, as well as identify major contributors to the firm's overall exposure and reveal hidden sources of risk that might not otherwise be apparent. If they are to provide up-to-date information, stress tests also need to be carried out on a reasonably frequent basis (e.g., every week or month).

Stress tests can also guide decision-making and, in particular, help with setting position limits, allocating capital, and managing funding risks. The usefulness of stress tests for setting positions and allocating capital is self-evident, but stress tests can help the management of funding risks by identifying the circumstances in which firms might run into funding problems, so that managers can take appropriate pre-emptive action.

Finally, stress testing can also help firms design systems to protect against bad events – for example, they can provide a check on modelling assumptions, help design systems to protect against stress events, and help in contingency planning.

13.1 BENEFITS AND DIFFICULTIES OF STRESS TESTING

13.1.1 Benefits of Stress Testing

Stress testing (ST) is ideal for showing up the vulnerability of our portfolio (and of our VaR calculations) to otherwise hidden risks or sources of error. Schachter (1998) suggests five different ways in which ST can provide valuable information to risk managers that may not otherwise be available to them:

- Since stress events are (usually) unlikely, the chances are that the data used to estimate VaR (or ES) will not reveal much about them.
- The short holding period often used for VaR will often be too short to reveal the full impact of a stress event, so it is important to carry out stress events on longer holding periods.

- If stress events are rare, they are likely to fall in the VaR tail region, and VaR will tell us nothing about them. As noted earlier ES fares a little better because it would tell us the expected (i.e., average) value of tail losses, but even ES does not really tell us a great deal about prospective bad events.
- Assumptions that help to value non-linear positions in normal times might be wide of the mark in a stress situation, so a stress test with full revaluation could reveal considerably more than, say, a second-order approximation VaR.
- A stress test could take account of the unusual features of a stress scenario (e.g., such as radicalised correlations, etc.) and so help reveal exposures that a VaR procedure would often overlook.

We can give many examples where stress testing highlights exposures that other probabilistic or other approaches to risk measurement might easily overlook. An important example is in helping to identify an institution's breaking point – it helps to identify those types of scenario (in terms of severity and conjunction of events, etc.) that would force the institution into insolvency. To quote the Federal Reserve chairman, Alan Greenspan:

> In estimating necessary levels of risk capital, the primary concern should be to address those disturbances that occasionally do stress institutional solvency – the negative tail of the loss distribution that is so central to modern risk management. As such, the incorporation of stress scenarios into formal risk modelling would seem to be of first-order importance.[2]

A stress test could identify these scenarios much more explicitly than other methods, and so give management a much clearer idea of the scenarios that they had to worry about. Once these scenarios are identified, it becomes much easier to develop hedging or other risk management strategies to protect against them.

A stress test is also very good for identifying and quantifying liquidity exposures: a stress test can identify liquidity risk factors that might not otherwise be apparent. Liquidity effects – such as connections between interest rates and collateral requirements or credit triggers, the impacts of widening bid–ask spreads and increasing execution times, etc. – can be quite subtle. VaR systems cannot really do them justice, but they are quite amenable to well-designed stress tests. As with solvency tests, the information provided by liquidity stress tests can be crucial in determining how to deal with the risks concerned.

A stress test can be useful in identifying the consequences of large market moves. For example, given the leverage involved in options positions, a firm that delta hedges could be covered against a very small market move and destroyed by a very large one, and the only way to detect this sort of exposure is to run stress tests based on very large hypothesised market moves (e.g., moves of 5–10 standard deviations, or more). We might also use stress tests to examine some of the other potential consequences of a large market move, including the consequences of a drying up of market liquidity, or the possible funding consequences if positive-value derivatives positions suddenly become major liabilities and force us to put up collateral or meet margin calls.

Stress testing is also good for examining the consequences of changes in volatility. Estimates of volatility based on historical data can be unreliable, and reliance on them can, on occasion, lead to much bigger losses than might have been expected. To illustrate the point, if we take a reasonable period prior to any major exchange rate crisis, any historically based estimate of the

[2] Greenspan (2000), p. 2.

VaR or ES of a relevant cross-currency portfolio would have indicated relatively little exchange rate risk: the exchange rate would have been stable for a considerable time, so no historical approach would have had any reason to indicate major exchange rate risk. The exchange rate then changes very abruptly and anyone on the wrong side of the market would have taken major losses, and yet this vulnerability could easily be picked up by a simple stress test. Volatility can also change suddenly in other markets as well, particularly in equity and commodities markets.

Similarly, we can also use stress tests to highlight dependence on correlation assumptions. Since the risk of a portfolio depends on the expected correlations of the various positions included in it, a major change in correlation could leave our portfolio much more exposed than we thought it was going to be. Historical correlations can themselves be very volatile and the most drastic changes in correlations tend to occur in crises such as market crashes. If we wish to survive such events, it is important that we not only examine our exposure to large market moves, but also examine what we stand to lose if 'normal' correlations break down and markets all move against us, and the only way to gauge this sort of exposure is to carry out scenario analyses.

Lastly, but not least, stress tests can be very useful for highlighting other weaknesses in our risk management set-up. The process of actually going through a stress testing exercise should force risk managers and senior managers to think through the ramifications of bad scenarios as well as help them to pinpoint weaknesses that they might have underestimated or overlooked. If it is done well, it should not only give some indication of where the institution is vulnerable, but also show up flaws in contingency planning. In fact, what risk managers learn about these hidden weaknesses is often as valuable for risk management purposes as the loss figures that the exercise finally produces.[3]

13.1.2 Difficulties with Stress Tests

Stress testing is generally much less straightforward than it looks. Stress tests are based on large numbers of decisions about the choice of scenarios and/or risk factors to stress, how risk factors should be combined, the range of values to be considered, the choice of time frame, and so on.

Stress testing is also dependent on the chosen scenarios and, hence, on the judgement and experience of the people who carry out the stress tests. This is a serious drawback because, as we all know, the negative events that we want to guard against can often be very hard to predict. Choosing the 'right' scenarios is therefore an important but sometimes difficult task. There have been many cases in the last few years of large companies being severely embarrassed or bankrupted by events that their management did not see coming (and, in some cases, by events that they clearly *should* have seen coming). When portfolios are complex, it can also be very difficult even to identify the risk factors to look at. The usefulness of stress testing therefore boils down to the skill, good sense and intuition of those who carry out the stress tests – and, in the final analysis, this is why good risk management is still as much craft as science.

Another problem with stress testing is the sheer difficulty of working through scenarios in a consistent, sensible way, *without* being overwhelmed by a mass of different possibilities.

[3] In order to make best use of stress tests, a good practice is to specify a threshold beyond which the loss would be regarded as a serious problem. This threshold would be set in terms of the institution's capital or in terms of the capital allocated to the business unit concerned. If a stress test threw up a loss that exceeded this threshold, the institution would respond with a formal review to examine the circumstances under which a very high loss could occur. This process would look closely at the co-movements leading to the loss and assess how likely the outcome is. An informed decision can then be made as to whether and, if so, how, to cover the risk.

There are three main issues here:

- We need to be able to follow through scenarios, and the consequences of some scenarios can be very complex: a trigger event occurs, and affects a number of variables; each of these affected variables then impacts on a number of others, and each other; these affect other variables; and so on. A trigger event can rapidly lead to a plethora of possibilities, and if we are not careful, the number of possibilities can become unmanageable.
- In working through scenarios, we will often (though not necessarily always) want to take account of the interactions of different risks. While it is sometimes useful to carry out scenario analyses in which all correlations are assumed to move in the most damaging ways, the fact is that we will not always want to make such assumptions and, on the contrary, will often want to take account of the interrelationships between different variables. Our stress test might indicate that the maximum loss could occur when one price rises and the other falls, and yet the prices of the two assets might be very strongly positively correlated. The stress test then ignores the likelihood that the two prices will move up or down together, and may produce a loss estimate much higher than any loss that could plausibly occur. In using stress tests, we must therefore decide when and, if so, how, to allow for correlations.
- In designing our scenarios, we must also recognise that there are often situations where prices cannot move independently of each other because doing so would violate a zero-arbitrage condition. To carry out stress testing sensibly, we need to eliminate all co-movements that are inconsistent with zero arbitrage.

Stress tests can also run into various computational problems. (1) The first of these is the need to take account of the differing sensitivities of instrument prices to underlying risk factors. The point here is that pushing all prices by the same multiple of their standard deviation ignores the sensitivity of each position to the underlying risk factors: for example, the price of a deeply out-of-the-money option is insensitive (in absolute terms) to a change in the underlying price, but the price of an in-the-money option could be very sensitive to it. The probability of an option price change that is α times the volatility is therefore much higher for a deeply in-the-money option than for a deeply out-of-the-money option. Consequently, it does not make much sense to push all prices by the same number of standard deviations, when the probability of such a change varies very considerably from one position to another. The solution is not to push the individual prices by any particular multiple, but to push the underlying risk factors instead. (2) Stress tests can be computationally expensive, and computational considerations impose a limit on how frequently they can be carried out. This is often the case where options positions are fully revalued during stress tests using intensive procedures such as simulation methods. Many firms also face computational problems because of systems incompatibilities of one sort or another. (3) There are serious difficulties integrating market and credit risk factors in stress analysis, and a recent BIS survey of stress testing in financial institutions reported that none of the surveyed firms had systems that fully integrated market and credit risk in stress testing.[4] Much of the time, integration appears to have gone little further than taking account of the impact of credit-related changes in the prices of traded instruments.

There is also the issue of probability: since stress tests as such do not give any indication of likelihood, we always face the problem of judging the importance of stress test results. Suppose a stress test suggests that a particular event would drive our firm into insolvency. Does this matter? The answer is that we cannot say without more information. If the event concerned

[4] BIS (2000), p. 15.

could occur with a significant probability, then clearly the stress test result is important and should (presumably) be taken seriously. But if the probability of occurrence was negligible, there is no real point paying much attention to it: rational people don't waste time and resources dealing with dangers that are too improbable to worry about. In order to use stress tests meaningfully, we need to form some idea, even a very loose and informal one, of the likelihood of the events concerned.[5]

Box 13.2 Integrating Stress Tests into Probabilistic Risk Estimation

A risk manager typically faces two separate types of risk estimate – probabilistic estimates such as VaR or ES, and the loss estimates produced by stress tests – with no obvious way of combining them. So how can we combine a probabilistic risk estimate with an estimate that such-and-such a loss will occur if such-and-such happens? The traditional answer is that we can't: we have to work with these estimates separately, and the best we can do is use one set of estimates to check for possible problems with the other.

Berkowitz (2000a) offers a solution to this problem: he suggests that we integrate stress testing into formal risk modelling by assigning probabilities to stress test scenarios. The resulting risk estimates then incorporate both traditional market risk estimates and the outcomes of stress tests, as well as the probabilities of each, and so give risk managers a single, integrated set of risk estimates to work with. This suggests the following risk modelling process: (1) We go through our stress testing in the usual way, and the outputs of this process will be a set of realised profits/losses associated with each scenario. (2) We go through a second, judgemental, process and assign probabilities to each of our scenarios. (3) We then go through a formal risk modelling process of the traditional kind, and model our risks using appropriate risk measurement techniques, to produce a set of P/L figures and their associated probabilities. (4) We bring together our two sets of P/L figures and associated probabilities, and estimate our desired risk measure (VaR, ES, etc.) in the usual way.

Naturally, these estimates are dependent on the judgemental factors that go into stress testing and into the evaluation of scenario probabilities, but – especially since the advent of coherent risk measures – there are good arguments that it is better to incorporate our judgements of stress test events into risk modelling than to ignore them completely.

13.2 SCENARIO ANALYSIS

We now turn to the first of our two main approaches to stress testing – scenario analysis.

13.2.1 Choosing Scenarios

The first step in scenario analysis is to choose the scenarios to be considered, and the scenarios can come in three main forms.

[5] Evaluating the plausibility (or, more formally, the probability) of stress scenarios is not difficult, at least in principle, and one straightforward way to do so is suggested by Breuer and Krenn (2000, p. 16). If we identify our stress scenario in terms of an n-dimensional vector of stress factors $\mathbf{r}_{\mathbf{stress}}$, this vector and the factor variance–covariance matrix $\boldsymbol{\Sigma}$ define an n-dimensional ellipsoid of scenarios $\mathbf{r}$ that satisfies $(\mathbf{r} - \mathbf{r}_{\mathbf{stress}})^{\mathbf{T}} \boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{r} - \mathbf{r}_{\mathbf{stress}})$. If $\mathbf{r}$ is normally distributed, the mass of the normal distribution contained in this ellipsoid will give us the probability of the stress scenario $\mathbf{r}_{\mathbf{stress}}$; but even if $\mathbf{r}$ is not normally distributed, we can often interpret the normal mass contained in this ellipsoid as an informal measure of plausibility.

13.2.1.1 Stylised scenarios

One type of scenario is a stylised scenario – a simulated movement in one or more major interest rates, exchange rates, stock prices or commodity prices. These scenarios can range from relatively moderate changes to quite extreme ones, and the movements considered can be expressed in terms of absolute changes, percentage changes, or in standard deviation units (i.e., the price change divided by the historical standard deviation of the relevant price). Some possible scenarios have been suggested by the Derivatives Policy Group (1995), and include parallel yield curve shifts of plus or minus 100 basis points, yield curve shifts of plus or minus 25 basis points, stock index changes of plus or minus 10%, currency changes of plus or minus 6%, and volatility changes of plus or minus 20%. If the institution is concerned about more extreme events, it might also want to consider such relatively rare events as 5- or 10-standard deviation changes in the relevant underlying price. We might also want to consider the impact of other factors too, such as changes in the slope or shape of the yield curve, a change in correlations, and a change in credit spreads (e.g., a jump or fall in the TED spread).

Stylised scenarios have been used for a long time in asset-liability management where they are suited to handling portfolios that are exposed to a small number of risk factors. The usual idea is to imagine hypothetical changes in the value of each risk factor and then use pricing equations (e.g., simple linear equations for straightforward positions, duration or duration–convexity approximations for bonds, or delta or delta–gamma approximations for options) to determine the change in the portfolio value resulting from the market factor change. We might assume that the exchange rate rises by $x\%$, interest rates fall by $y\%$, and so on. Each particular combination of risk factor movements leads to a particular new portfolio value and hence a particular profit or loss. If we can combine the analysis with some assessment of the likelihood of the changes, even an informal one, these computations can give a good picture of the risks confronting our portfolio. However, the main limitation of this approach is that it easily becomes unmanageable when there is more than a small number of risk factors. If there are too many risk factors or too many different scenarios for each factor, then the risk manager can easily end up with thousands of loss figures, each for a different combination of risk factor movements. The information can be overwhelming, and the risk manager can have great difficulty getting any overall sense of portfolio risk.

13.2.1.2 Actual historical events

We can also choose our scenarios from actual historical events. Historical scenarios can be based on relatively moderate market changes, which presumably have a reasonable chance of repeating themselves, or more extreme market changes, which are much less likely but more significant if they do, and they can also be based on bootstrap exercises from historical data. Historical scenarios have two advantages relative to other scenarios:

- The fact that historical scenarios have actually occurred reduces their arbitrariness and gives them a certain plausibility that other scenarios lack. It is also hard(er) to dismiss historical scenarios on the grounds that they couldn't happen.
- They are readily understood. A statement like 'the firm would lose $X million if there were a repeat tomorrow of the October 1987 stock market crash' is easy to understand, and this type of clarity is very useful in communicating risk information effectively.

While the precise choice of historical scenario – the data period used, the prices or price indices considered, whether and how to bootstrap, etc. – is inevitably subjective, we can make

the selection process a little more systematic by using a well-produced scenario catalogue, rather than just a handful of ad hoc scenarios pulled out of thin air. Such a catalogue might include:

- Moderate market scenarios such as bootstrapped recent market-return scenarios, changes in market volatility, a bond market squeeze due to fiscal surpluses, changes in the euro, a widening or falling TED spread, and others from recent market experience.
- More extreme market scenarios such as repeats of major stock market crises (e.g., the 23% fall in the Dow-Jones on October 19, 1987, etc.) or exchange rate crises (e.g., the fall in the peso in December 1994, the east Asian devaluations in 1997, the 40% fall in the rouble in August 1998, etc.), a bond market crash (e.g., the near doubling of US interest rates in 1994), major country shocks (e.g., the Latin American crisis in 1995, the Asian crisis in 1997, Russia in August 1998, and Brazil in 1999), or the failure or near failure of a large financial institution (e.g., LTCM, Enron, Parmalat, etc.).

A good guide is to choose scenarios of much the same order of magnitude as the worst-case events in our historical (or bootstrapped historical) data sets. In doing so, we should obviously keep in mind that these events are notoriously difficult to predict in advance. We should also keep in mind that market experience suggests that maximum price falls vary enormously from one market to another and, within any given market, are often very much bigger than the next largest price fall.

Box 13.3 Guidelines for Successful Scenario Analysis[6]

Many firms could improve their stress testing by ensuring that stress testers: allow for large changes in risk factors, so that stress situations are genuinely stressful; take proper account of the speed and duration of stress events; identify key assumptions, and gauge firms' vulnerability to them, to avoid the danger that important assumptions remain hidden; take account of linkages between risk factors (such as connections between market, credit and liquidity risks); and carry out stress tests frequently enough to ensure that results are up to date and relevant. It is also important to strike the right balance between hypothetical plausible and historical scenarios, and between stress test exercises and probabilistic risk measurement.

Successful stress testing also requires that firms avoid or at least mitigate a number of common pitfalls: senior management might not 'buy into' the stress test exercise, and so ignore stress test results; results might be evaluated by managers who lack the authority to take remedial action; and users of stress test results might confuse 'what if' scenarios and worst-case scenarios. Senior managers should also try to avoid the danger that stress tests might become a political weapon to be used or ignored by interested parties within the firm in the pursuit of other objectives, thus compromising the integrity and credibility of the stress testing process. Institutions can try to avoid these pitfalls by doing what they can to ensure that senior managers buy into stress testing exercises, that all interested parties are involved in selecting scenarios, and that results are reported to interested parties in appropriate detail.

[6] For more on these points, see Wee and Lee (1999), pp. 16–17.

13.2.1.3 Hypothetical one-off events

Scenarios can also come from plausible hypothetical scenarios that have no direct historical precedents. These scenarios would not be replays of past historical events, as such, although they would have some similarity with past events. These scenarios might be natural (e.g., a major earthquake in California), political (e.g., the outbreak of a war or a sovereign default), legal (e.g., a ruling on the legality of a derivatives contract), economic or financial (e.g., the default of a major financial institution or a new financial crisis), credit related (e.g., the downgrading of a major counterparty), or liquidity related (e.g., a major hike in credit spreads). We can often formulate such scenarios by looking at historical experience and asking what might have been.

We can also look to the historical record to give us an indication of what such an event might look like. A good example highlighted in the recent BIS report on stress testing is a 'flight to quality'.[7] Such scenarios involve shocks to credit spreads, such as a jump in the TED spread. However, the flight to quality experienced in late 1998 also involved close connections between credit spreads and liquidity factors, so the flight-to-quality scenarios now used by financial institutions have been refined further to take more account of liquidity considerations, with more emphasis on liquidity-related changes in spreads, such as changes in the spread between on- and off-the-run US Treasury instruments.

13.2.2 Evaluating the Effects of Scenarios

Having specified each scenario as fully as we can, we need to consider the effect of each scenario on the prices of all instruments in our portfolio. The key task is to get an idea of the sensitivities of our various positions to the underlying risk factors whose hypothetical changes we are considering. This is very easy for some positions. Thus, a straight FX position changes one-for-one in value with changes in the exchange rate, and the value of a diversified stock portfolio changes (roughly) one-for-one with changes in the stock market index. Many other positions also change one-for-one (or thereabouts) with changes in the underlying market risk factor. Some other positions have less straightforward sensitivities, but we can usually handle them by using approximations. For example, we could obtain the approximate sensitivities of option prices to changes in underlying risk factors from estimates of their deltas, gammas, vegas and other risk parameters, all of which should be readily available; and where bonds are concerned, we might proxy their sensitivities to changes in market interest rates by taking duration or duration–convexity approximations.

Once we have determined the effect of each scenario on all relevant prices, we can infer the effect of each scenario on the portfolio value as a whole. The portfolio loss is then found by subtracting the portfolio's existing value from its post-scenario value.

In evaluating the effects of scenarios on our portfolio, we should also consider the impact of our hypothesised events on the markets in which we operate. In particular, it is very unwise to assume that markets will continue to function 'normally' when subjected to extreme stress. To illustrate, under normal stock market conditions we could expect to see sell orders executed within a matter of minutes; yet, on October 19, 1987, stock markets were so overwhelmed that it could take hours to get orders executed. Sell orders either expired because of time or price limits or else were executed at much lower prices than the sellers had expected. Market

[7] BIS (2001), p. 12.

liquidity consequently dried up just when sellers were most dependent on it. Firms whose risk management strategies are based on dynamic hedging or an assumed ability to rebalance portfolios quickly should therefore pay considerable attention to the impact of extreme events on market liquidity. They should also watch out that volatility and correlation assumptions that may appear reasonable in 'normal' times do not break down when markets are stressed and leave them with much bigger exposures than they thought they would have.

Companies that use futures contracts to hedge illiquid positions should also take into account the funding implications of their hedge positions. Gains or losses in futures positions must be settled on a daily basis, while changes in other positions (e.g., forward ones) will not be settled until the position is finally closed out. Hence, even otherwise well-designed hedges can lead to mismatches between the timing of receipts and the timing of the payments that theoretically cover them. If the hedges are large, these interim funding requirements can also be large. Indeed, it was the failure to consider just this point that played a key factor in bringing the German industrial giant Metallgesellschaft to its knees in 1993–94.

Box 13.4 Stress Testing in a VaR Framework

A major problem with traditional stress testing is that it throws away valuable information, particularly about volatilities and correlations. To remedy this drawback, Kupiec (1999) proposes a new approach – a form of conditional stress testing or stress-VaR approach – that seeks to make use of this information in stress testing. Suppose we have a set of risk factors that are, say, normally distributed. We partition these into two sets – a set of k factors, $\tilde{\mathbf{R}}_{\mathbf{1t}}$, that are to be stressed to take values $\mathbf{R}_{\mathbf{1t}}$, and those that are not, $\tilde{\mathbf{R}}_{\mathbf{2t}}$. If the variance–covariance matrix $\boldsymbol{\Sigma}$ is unaltered in the stress test,[8] the unstressed factors $\tilde{\mathbf{R}}_{\mathbf{2t}}$ are conditionally distributed as:

$$\tilde{\mathbf{R}}_{\mathbf{2t}}|_{\tilde{\mathbf{R}}_{\mathbf{1t}}=\mathbf{R}_{\mathbf{1t}}} \sim \mathbf{N}(\boldsymbol{\mu}_{\mathbf{c}}, \boldsymbol{\Sigma}_{\mathbf{c}})$$

where $\boldsymbol{\mu}_c = \boldsymbol{\Sigma}_{\mathbf{21}}\boldsymbol{\Sigma}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{R}_{\mathbf{1t}}$, $\boldsymbol{\Sigma}_{\mathbf{c}} = \boldsymbol{\Sigma}_{\mathbf{22}} - (\boldsymbol{\Sigma}_{\mathbf{21}}\boldsymbol{\Sigma}_{\mathbf{11}}^{-\mathbf{1}}\boldsymbol{\Sigma}_{\mathbf{12}})$, and the $\boldsymbol{\Sigma}_{11}$ and so on are the partitioning submatrices of $\boldsymbol{\Sigma}$, given the values $\boldsymbol{R}_{\mathbf{1}t}$ of the stressed factors. Given this conditional density function, the stress-scenario change in portfolio value is a normally distributed random variable with mean $\mathbf{X}_{\mathbf{1t}}\mathbf{R}_{\mathbf{1t}} + \mathbf{X}_{\mathbf{2t}}\boldsymbol{\mu}_{\mathbf{c}}$ and variance $\mathbf{X}_{\mathbf{2t}}\boldsymbol{\Sigma}_{\mathbf{c}}\mathbf{X}_{\mathbf{2t}}^{\mathbf{T}}$, where $\mathbf{X}_{\mathbf{1t}}$ and $\mathbf{X}_{\mathbf{2t}}$ are the position vectors corresponding to the two types of risk factors. Once the joint distribution of the risk factors is taken into account, our stress test produces a distribution of scenario loss values, not just a single loss value. If we wish, we can then focus on one of the percentile points of the scenario loss distribution, in which case our output can be interpreted as a stress VaR: the likely worst outcome at a chosen confidence level.

Alternatively, we can focus on the mean of the conditional loss distribution, in which case our stress loss is $\mathbf{X}_{\mathbf{1t}}\mathbf{R}_{\mathbf{1t}} + \mathbf{X}_{\mathbf{2t}}\boldsymbol{\Sigma}_{\mathbf{21}}\boldsymbol{\Sigma}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{R}_{\mathbf{1t}}$. This expected loss differs from the expected loss we get under traditional stress testing because it takes account of the correlations between different risk factors: under a traditional stress test, we would stress the stressed risk factors and take the other risk factors to be unaltered, and therefore get an expected loss of $\mathbf{X}_{\mathbf{1t}}\mathbf{R}_{\mathbf{1t}}$. Kupiec's results indicate that his proposed measure fares better than the traditional expected loss which ignores correlations between risk factors.

[8] This assumption and the earlier assumption of normality are only made for convenience: they are not essential, and we can relax them to some extent if we are prepared to make the analysis a little more difficult.

Example 13.1 (Impact of volatility change on VaR)

Suppose we have a normal VaR defined over some holding period, with P/L parameters $\mu = 0.1$ and $\sigma = 0.25$. What happens to our estimate of the 95% VaR if we change our volatility forecast to 0.5?

To answer this question, we first apply the usual normal VaR formula, which tells us that VaR $= -\mu + \sigma z_\alpha$. This tells us that our 95% VaR is $-0.1 + 0.25 \times 1.645 = 0.311$. If the volatility doubles to 0.5, our VaR changes to $-0.1 + 0.5 \times 1.645 = 0.723$. More generally, if volatility increases by $\Delta\sigma$, then our VaR increases by $\Delta\sigma z_\alpha$. This illustrates the sensitivity of VaR estimates to volatility assumptions.

Example 13.2 (Impact of change in correlation on VaR)

Suppose we have a bivariate normal VaR position with P/L parameters

$$\boldsymbol{\mu} = [0, \quad 0], \textit{ volatilities } \boldsymbol{\sigma} = \begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}, \textit{ correlations } \mathbf{C} = \begin{bmatrix} 1, & 0.1 \\ 0.1, & 1 \end{bmatrix}$$

(which implies a correlation of 0.1) and position sizes $\mathbf{x} = [0.5, \quad 0.5]$. What happens to our estimate of the 95% confidence-level VaR if we change our correlation forecast from 0.1 to 0.9?

We first note that we can write our VaR as equal to

$$-\boldsymbol{\mu}\mathbf{x}^{\mathbf{T}} + \mathbf{x}\boldsymbol{\sigma}\boldsymbol{\rho}\boldsymbol{\sigma}^{\mathbf{T}}\mathbf{x}^{\mathbf{T}} z_\alpha = -[0, \quad 0]\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + [0.5, \quad 0.5]\begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}\begin{bmatrix} 1, & 0.1 \\ 0.1, & 1 \end{bmatrix} \times \begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \times 1.645 = 0.9048$$

Changing the cross-correlation coefficient from 0.1 to, say, 0.9, changes our VaR to

$$-[0, \quad 0]\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + [0.5, \quad 0.5]\begin{bmatrix} 1, & 0 \\ 0, & 11 \end{bmatrix}\begin{bmatrix} 1, & 0.9 \\ 0.9, & 1 \end{bmatrix}\begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \times 1.645 = 1.5628$$

The VaR thus rises by 0.6058 or 72.72%, which indicates how sensitive VaR estimates can be to correlation assumptions.

Example 13.3 (Stress testing the values of option positions)

Suppose we have \$1 invested in Black–Scholes call options, and the option price parameters are S (underlying price) $= 1.2$, X (strike) $= 1$, r (risk-free rate) $= 0.04$, σ (volatility rate) $= 0.25$, and T (term to expiry in years) $= 0.5$. Now:

(a) How would the value of our option position change if the underlying price were to fall 25% by the end of the next trading day?
(b) How would the results differ if the current underlying price was 0.8 instead of 1.2?

(a): We begin by using the Black–Scholes call option formula to tell us that a single option has a price of 0.232. The number of options in our position is therefore $1/0.232 = 4.314$. If the underlying price falls by 25% by the end of the next trading day, the option price will fall to 0.034, so our option position will fall in value by $4.314 \times (0.232 - 0.034) = 0.854$ – a fall of over 85%.

(b): We now repeat the analysis for an initial underlying price of 0.8. Using Black–Scholes, the initial option price is 0.010, and the number of options in our position is $1/0.010 = 100$. If the underlying price falls by 25% by the end of the next trading day, the option price will fall to 0.0001, so our option position will fall in value by $100 \times (0.010 - 0.0001) = 0.99$ – a fall of 99%.

The lessons? First, that the sensitivity of the option position can be very different from that of the underlying position. Second, that the sensitivity of the value of an option position to a shock in the underlying price is dependent on the extent to which the option is in- or out-of-the-money.

13.3 MECHANICAL STRESS TESTING

We turn now to mechanical stress tests. These procedures attempt to avoid the subjectivity of scenario analyses and put the mechanics of stress testing on a firmer methodological foundation: instead of choosing scenarios subjectively, we generate them from a set of mathematically or statistically defined possibilities. We then work through these to determine the most damaging combination of events and the loss it would produce. Mechanical approaches are therefore more thorough and more systematic than traditional scenario analysis, but can also be computationally more intensive. Some mechanical stress testing procedures also differ from scenario analysis in that they are able to give some indication of the likelihood of different outcomes, and this information can be useful when trying to decide how seriously to take them.

13.3.1 Factor Push Analysis

The simplest of these procedures is factor push analysis, in which we 'push' the price of each individual security or (preferably) the relevant underlying risk factor in the most disadvantageous direction and work out the combined effect of all such changes on the value of the portfolio.[9] We have already met this type of approach in the shape of Wilson's delta–gamma approach to VaR, which was discussed in Chapter 10. We start by specifying a level of confidence, which gives us a confidence-level parameter α. We then consider each risk factor on its own, 'push' it by α times its standard deviation, and revalue the portfolio at the new risk factor value; we do the same for all risk factors, and select that set of risk factor movements that has the worst effect on the portfolio value. Collecting these worst price movements for each instrument in our portfolio gives us our worst-case scenario, and the maximum loss (ML) is equal to the current value of our portfolio minus the portfolio value under this worst-case scenario.[10]

Factor push analysis is relatively easy to program, at least for simple positions, and is good for showing up where and how we are most vulnerable. It does not require particularly

[9] For more on this approach, see, e.g., Rouvinez (1997) and Studer (1999).

[10] To do factor push analysis properly, we should also take account of relevant constraints, such as zero-arbitrage conditions, and we might also want to work with mapped positions, delta–gamma approximations and so on.

restrictive assumptions and can be used in conjunction with a non-linear P/L function (e.g., such as a P/L function that is quadratic in underlying risk factors), and can be modified to accommodate whatever correlation assumptions we care to make (e.g., using a Choleski decomposition). Furthermore, in common with measures such as those generated by the SPAN risk measurement system or by worst-case scenario analysis, not to mention the ES, the ML risk measure also has the major attractions of being coherent.[11]

If we are prepared to make certain additional assumptions, factor push can also tell us something about the likelihood of the losses concerned. If we have just one risk factor and make appropriate parametric assumptions (e.g., such as normality), α enables us to infer the tail probability, and the maximum loss on the boundary of the confidence region is our VaR. α can still tell us something about the probabilities when we have multiple risk factors and make appropriate parametric assumptions (e.g., multivariate normality, given a set of correlation assumptions), but the analysis is more complicated and we find that the ML on the boundary of the confidence region is a conservative estimate (i.e., an overestimate) of our VaR.[12] However, we can also adjust the α value to make the ML equal to our VaR, and if we make this adjustment, we can interpret ML as a VaR and use the factor push approach as an alternative way of estimating VaR.[13]

However, FP rests on the not-always-appropriate assumption that the maximum loss occurs at extreme values of the underlying risk variables (i.e., it assumes that the maximum loss occurs when the underlying factor moves up or down by α times its standard deviation).[14] Yet this assumption is only appropriate for certain relatively simple types of portfolio (e.g., uncomplicated equity or FX positions) in which the position value is a monotonic function of a (typically, single) risk factor, and there are many other instruments for which this assumption does not hold. A good example is a long straddle – a combination of a long call and a long put written against the same underlying asset. The profit on a straddle depends on movements in the underlying variable, either up or down – the greater the movement, the bigger the profit – and the maximum loss on a straddle actually occurs when the underlying price does not move at all. A naïve factor push methodology applied to a straddle position would then give a misleading picture of maximum loss, since it would assume that the maximum loss occurred in exactly those circumstances where it would in fact make its maximum profit! There is also good reason to believe that this type of problem is quite serious in practice. To quote Tanya Beder:

> In our experience, portfolios do not necessarily produce their greatest losses during extreme market moves . . . portfolios often possess Achilles' heels that require only small moves or changes between instruments or markets to produce significant losses. Stress testing extreme market moves will do little to reveal the greatest risk of loss for such portfolios. Furthermore, a review of a portfolio's

[11] SPAN and worst-case scenario analysis approaches were discussed further in Chapter 2.

[12] See Studer (1999), p. 38. This is the same problem raised by Britten-Jones and Schaefer (1999) in their critique of Wilson's QP approach (e.g., as in Wilson (1996)) discussed earlier – namely, that identifying outcomes as being inside or outside a confidence region does not tell us the probability of those outcomes, with the result that Wilson's ML is an overestimate of VaR.

[13] An alternative method of obtaining VaR estimates from a factor push methodology is provided by Rouvinez (1997). Suppose we start by assuming that the changes in the risk factors are independent normal. The sum of the squares of these changes is then distributed as a χ^2, and (assuming a zero mean) the VaR is equal to the relevant quantile of the χ^2 distribution, say β, times the portfolio standard deviation. Since β is generally bigger than the standard normal quantile α, this approach generally leads to bigger VaR estimates than, say, a delta–normal approach. For more on this method, see Rouvinez (1997, pp. 60–62).

[14] One other problem with mechanical stress tests is that the largest losses might come from conjunctions of events that will not in fact occur. For example, we might find that the maximum loss occurs when there is a large fall in the stock market associated with a large fall in stock market volatility. Since such combinations cannot plausibly occur, the losses associated with them are not really worth worrying about. In carrying out mechanical stress tests, we therefore need to screen for such implausible combinations and remove them from consideration.

> expected behavior over time often reveals that the same stress test that indicates a small impact today indicates embedded land mines with a large impact during future periods. This trait is particularly true of options-based portfolios that change characteristics because of time rather than because of changes in the components of the portfolio.[15]

When using factor push, we first need to satisfy ourselves that our portfolio suffers its maximum loss when the risk factors make their biggest moves.

13.3.2 Maximum Loss Optimisation

The solution to this latter problem is to search over the losses that occur for intermediate as well as extreme values of the risk variables. This procedure is known as maximum loss optimisation (MLO), and is essentially the same as factor push analysis, except that it also searches over intermediate as well as extreme values of the risk variables. There are therefore more computations involved, and MLO will take longer if there are many risk factors involved and a lot of intermediate values to search over. Consequently, the choice between FP and MLO depends on the payoff characteristics of our portfolio. If the portfolio is made up of straightforward positions, each of which takes its maximum loss at extreme values of the underlying risk factors, then FP and MLO will deliver exactly the same results and we may as well use the computationally simpler FP approach. However, if the portfolio has less straightforward payoff characteristics (e.g., as with some options positions), it may make sense to use MLO instead. MLO can also help pick up interactions between different risks that we might otherwise have overlooked, and this can be useful for more complex portfolios whose risks might interact in unexpected ways. As a general rule, if the portfolio is complex or has significant non-linear derivatives positions, it is best to play safe and go for MLO.[16]

13.3.3 CrashMetrics

CrashMetrics is a form of maximum loss optimisation that is designed to estimate worst-case losses.[17] The idea is that we estimate the plausible worst-case outcome using some reasonable ad hoc method. To illustrate, if we have a long position in a single option, the option P/L can be approximated by a delta–gamma approximation:

$$\Pi = \delta \Delta S + \frac{\gamma}{2}(\Delta S)^2 \tag{13.1}$$

where ΔS is the change in the stock price, and δ and γ are the option's delta and gamma. The maximum loss on this position occurs when $\partial \Pi / \partial \Delta S = 0$ and $\partial^2 \Pi / \partial \Delta S^2 > 0$, and therefore occurs when $\Delta S = -\delta / \gamma$. This means that the maximum loss is equal to

$$Loss^{\max} = -\Pi^{\min} = \frac{\delta^2}{2\gamma} \tag{13.2}$$

[15] Beder (1995), p. 18.

[16] The actual calculations can be done using a variety of alternative approaches. The most obvious approach is a grid search, in which we discretise the possible movements in risk factors and search over the relevant n-dimensional grid to find that combination of risk-factor changes that maximises our loss. However, we can also use simulation methods, or numerical methods such as a multidimensional simplex method. These methods are also capable of considerable refinement to increase accuracy and/or reduce computation time.

[17] Hua and Wilmott (1997).

We can get comparable expressions for multi-option portfolios provided we can model the relationship, and the approach can be extended to deal with the other Greek factors, changes in bid–offer spreads, and so on. This particular type of application is open to criticism on the ground that it relies heavily on ad hoc Greek approximations in circumstances where those approximations are not likely to be good. However, more generally, CrashMetrics relies on some ad hoc MLO algorithm, and has the advantage of being relatively transparent.

Example 13.4 (CrashMetrics)

Suppose we wish to apply the CrashMetrics approach to estimate the worst-case margin or cash outflow on a vanilla call position with the following parameters: current stock price $S_0 = 1$, strike $X = 1$, risk-free rate $r_f = 0.05$, volatility rate $\sigma = 0.25$, and period-to-expiry (in days) $T = 30$.

To proceed, we estimate the option's delta and gamma using the standard formulas, which turn out to be 0.537 and 5.542. The worst-case cash outflow is therefore $\delta^2/(2\gamma) = 0.540^2/(2 \times 5.542) = 0.026$.

13.4 CONCLUSIONS

Stress tests have many attractions. First and foremost, they can give us a lot of information about what we stand to lose in bad states – and, indeed, stress testing is explicitly designed to give us information about losses in bad states. The information provided by stress testing is a natural complement to that provided by probabilistic risk measures, most particularly VaR. Second, stress test results can be very useful in risk management decision-making – in setting capital requirements and position limits, and so on. Finally, stress tests can highlight weaknesses in our risk management systems (such as awkward assumptions in risk measurement models or failures in contingency plans). If we do not engage in stress tests, it is only a matter of time before we become seriously unstuck by something or other: we will delta hedge, say, and take a big negative gamma hit when the underlying price crashes, or correlations will alter suddenly and leave us much more exposed than we thought we would be. Stress testing is essential for sound risk measurement and management.

But stress testing also has its limitations, and the most important of these revolve around its subjectivity. As Schachter puts it, stress testing

> represent[s] only a limited number of scenarios, the likelihood of any one being impossible to estimate. As a result neither the completeness nor the reliability of the information provided can be scientifically assessed. In addition, hypothetical stress scenarios cannot be 'validated' based on actual market events. That is, even when the events specified in a hypothetical scenario actually occur, there is usually no way to apply what was 'right' or 'wrong' in the scenario to other hypothetical scenarios to improve them. These limitations are not shared by value-at-risk models, which are statistically based. In these, it is possible to construct statistical confidence intervals around the VaR estimate and to conduct meaningful 'backtests' of the VaR model's predictions.[18]

Thus, Schachter suggests that stress tests are subjective and that their outcomes cannot be validated or falsified. This seems true enough.

---

[18] Schachter (1998), p. 5F-8.

However, even though they are subjective, the theory of coherent risk measures also tells us that the loss estimates they generate can still be regarded as bona fide risk measures. They are certainly subjective, and validation may be difficult if not impossible, but they are still risk measures in a respectable sense of the term – and, to rub the point in yet again, the VaR is not. No doubt this conclusion will make many risk managers uncomfortable, particularly those from quantitative backgrounds. But then again, risk measurement always did rest on a qualitative, judgemental, foundation, and the newly recognised 'respectability' of stress test outcomes as risk measures merely confirms that point.

14

Estimating Liquidity Risks

We have usually assumed so far that markets are liquid – that is, that we can costlessly liquidate or unwind positions at going market prices, typically taken as the mean of bid and ask prices. This assumption is very convenient and provides a nice justification for the practice of marking positions to market. However, it is often empirically questionable, and where it does not hold, we need to revise the way we estimate market risks to allow for the effects of illiquidity.

This chapter looks at liquidity issues and how they affect the estimation of market risk measures. Liquidity issues affect market risk measurement not just through their impact on our standard measures of market risk, but also because effective market risk management involves an ability to estimate and manage liquidity risk itself. We therefore need to be able to estimate liquidity risk – or liquidity at risk, if you will. Furthermore, since liquidity problems are particularly prominent in market crises, we also need to address how to estimate crisis-related liquidity risks. In short, the main themes of this chapter are:

- The nature of market liquidity and illiquidity, and their associated costs and risks.
- The estimation of VaR in illiquid or partially liquid markets – liquidity-adjusted VaR (or LVaR).
- Estimating liquidity at risk (LaR).
- Estimating crisis-related liquidity risks.

Of course, although we focus on VaR in this chapter, we should remember throughout that these same methods can also be used to estimate coherent risk measures as well (e.g., by applying the 'weighted average quantile' approach outlined in Chapter 3).

14.1 LIQUIDITY AND LIQUIDITY RISKS

The notion of liquidity refers to the ability of a trader to execute a trade or liquidate a position with little or no cost, risk or inconvenience. Liquidity is a function of the market, and depends on such factors as the number of traders in the market, the frequency and size of trades, the time it takes to carry out a trade, and the cost (and sometimes risk) of transacting. It also depends on the commodity or instrument traded, and more standardised instruments (e.g., such as FX or equities) tend to have more liquid markets than non-standardised or tailor-made instruments (e.g., such as over-the-counter (OTC) derivatives). Markets vary greatly in their liquidity: markets such as the FX market and the big stock markets are (generally) highly liquid; but other markets are less so, particularly those for many OTC instruments and instruments that are usually held to maturity and, hence, are rarely traded once initially bought. However, even the 'big' standardised markets are not perfectly liquid – their liquidity fluctuates over time,[1] and can fall dramatically in a crisis – so we cannot take their liquidity for granted.

[1] We can estimate market liquidity if we have good transactions data, and such data are now becoming available. Using such data, Froot *et al.* (2001) regress returns on cross-border purchases and sales and use the coefficients from these exercises to estimate a 'liquidity index', which gives us an estimate of the price impact of trades. Results suggest that market liquidity was very hard hit in

Imperfect liquidity also implies that there is no such thing as 'the' going market price. Instead, there are two going market prices – an ask price, which is the price at which a trader sells, and a (lower) bid price, which is the price at which a trader buys. The 'market' price often quoted is just an average of the bid and ask prices, and this price is fictional because no one actually trades at this price. The difference between the bid and ask prices is a cost of liquidity, and in principle we should allow for this cost in estimating market risk measures.

The bid–ask spread also has an associated risk, because the spread itself is a random variable. This means there is some risk associated with the price we can obtain, even if the fictional mid-spread price is given. Other things being equal, if the spread rises, the costs of closing out our position will rise, so the risk that the spread will rise should be factored into our risk measures along with the usual 'market' price risk.

We should also take account of a further distinction. If our position is 'small' relative to the size of the market (e.g., because we are a very small player in a very large market), then our trading should have a negligible impact on the market price. In such circumstances we can regard the bid–ask spread as exogenous to us, and we can assume that the spread is determined by the market beyond our control. However, if our position is large relative to the market, our activities will have a noticeable effect on the market itself, and can affect both the 'market' price and the bid–ask spread. For example, if we suddenly unload a large position, we should expect the 'market' price to fall and the bid–ask spread to widen.[2] In these circumstances the 'market' price and the bid–ask spread are to some extent endogenous (i.e., responsive to our trading activities) and we should take account of how the market might react to us when estimating liquidity costs and risks. Other things again being equal, the bigger our trade, the bigger the impact we should expect it to have on market prices.

In sum, we are concerned with both liquidity costs and liquidity risks, and we need to take account of the difference between exogenous and endogenous liquidity. We now consider some of the approaches available to adjust our estimates of VaR to take account of these factors.

14.2 ESTIMATING LIQUIDITY-ADJUSTED VaR

There are many ways we could estimate liquidity-adjusted VaR. These vary in their degrees of sophistication and in their ease (or otherwise) of implementation, and there is no single 'best' method. However, sophisticated approaches are not necessarily more useful than more basic ones, and the 'best' method, even if we could establish what it is, is not necessarily better than a collection of 'inferior' ones. Instead, what we really seek are *simple-to-implement* (i.e., spreadsheet-executable) approaches that are *transparent in terms of their underlying assumptions*: in effect, we are looking for liquidity 'add-ons' that allow us to modify original VaR estimates that were obtained without any consideration for liquidity. We can then easily assess the impact of our assumptions on our estimates of VaR. Moreover, there is a premium on *compatibility*, because different methods look at different aspects of illiquidity and it can be helpful to combine them to get some sense of an 'overall' liquidity adjustment. Because of

the summer of 1998, and thereafter took a long time to recover: for example, cross-border equity liquidity was still less in 2000 than it was in 1997, and liquidity in some emerging markets was even lower in 2000 than during the turbulence of 1998.

[2] There is an extensive financial economics literature on this subject, but broadly speaking, the literature suggests two reasons why market prices might move against the trader. The first is the liquidity effect already alluded to, namely, that there is a limited market, and prices must move to induce other traders to buy. The other reason is a little more subtle: large trades often reveal information, and the perception that they do will cause other traders to revise their views. A large sale may encourage other traders to revise downwards their assessment of the prospects for the instrument concerned, and this will further depress the price.

this, a really good method might not always be as useful as two 'inferior' methods that actually work well together.

Whichever models we used, we also need to check their sensitivities – how does the liquidity adjustment change as we change the confidence level, holding period or any other parameters? A priori, we should have some idea of what these should be (e.g., that the liquidity adjustment should fall as the holding period rises, etc.) and we need to satisfy ourselves that the models we use have sensitivities of the 'right' sign and approximate magnitude. Going further, we should also try to ensure that models are calibrated against real data (e.g., bid–ask spread parameters should be empirically plausible, etc.), and be properly stress tested and backtested. In addition, we should keep in mind that different approaches are often suited to different problems, and we should not seek a single 'best' approach to the exclusion of any others. In the final analysis, liquidity issues are much more subtle than they look, and there is no established consensus on how we should deal with them. So perhaps the best advice is for risk measurers to hedge their bets, and use different approaches to highlight different liquidity concerns.

14.2.1 The Constant Spread Approach

A good place to start is with the bid–ask spread,[3] and the simplest way to incorporate liquidity risk into a VaR calculation is in terms of a spread that is assumed to be constant. If we make this assumption, the liquidity cost is then equal to half the spread times the size of the position liquidated. Using obvious notation, this means that we add the following liquidity cost (LC) to a 'standard' VaR:

$$LC = \frac{1}{2} spread^{*} P \tag{14.1}$$

where *spread* is expressed as actual spread divided by the midpoint. For the sake of comparison, let us compare this to a benchmark conventional lognormal VaR with no adjustment for liquidity risk:

$$VaR = P[1 - \exp(\mu_R - \sigma_R z_\alpha)] \tag{14.2}$$

where the returns have been calculated using prices that are the midpoints of the bid–ask spread. The liquidity-adjusted VaR, *LVaR*, is then given by:

$$LVaR = VaR + LC = P[1 - \exp(\mu_R - \sigma_R z_\alpha) + \frac{1}{2} spread\,] \tag{14.3}$$

Setting $\mu_R = 0$ to clarify matters, the ratio of *LVaR* to *VaR* is then

$$\frac{LVaR}{VaR} = 1 + \frac{spread}{2[1 - \exp(-\sigma_R z_\alpha)]} \tag{14.4}$$

It is easy to show that the liquidity adjustment (a) rises in proportion with the assumed spread, (b) falls as the confidence level increases, and (c) falls as the holding period increases. The

[3] Of course, if we had actual transaction prices rather than the means of bid and ask prices, we could infer the actual returns obtained by traders, in which case conventional VaR methods would take account of spread liquidity factors without the need for any further adjustment. Thus, we would model actual returns (taking account of how they depend on market volume, etc.), infer a relevant conditional distribution (e.g., a *t*), and plug in the values of the parameters concerned into the appropriate parametric VaR equation. For more on how this might be done, see Giot and Grammig (2003).

first and third of these are obviously 'correct', but the second implication is one that may or may not be compatible with one's prior expectations.

This approach is easy to implement and requires only minimal information, but the assumption of a constant spread is highly implausible, and it takes no account of any other liquidity factors.

Example 14.1 (Constant spread approach)

Suppose $\mu = 0$, $\sigma = 0.25/\sqrt{250}$, $spread = 0.02$ and $\alpha = 0.95$. Using Equation (14.4), the ratio of *LVaR* to *VaR* is

$$\frac{LVaR}{VaR} = 1 + \frac{0.02}{2[1 - \exp(-0.25 \times 1.645/\sqrt{250})]} = 1.3895$$

The constant spread liquidity adjustment therefore raises the VaR by almost 40%. Thus, even a small spread can translate into a surprisingly large liquidity adjustment to our VaR.

14.2.2 The Exogenous Spread Approach

A superior alternative is to assume that traders face random spreads. If our position is sufficiently small relative to the market, we can regard our spread risk as exogenous to us (i.e., independent of our own trading), for any given holding period. We could assume any process for the spread that we believe to be empirically plausible.[4] For example, we might believe that the spread is normally distributed:

$$spread \sim N\left(\mu_{spread}, \sigma^2_{spread}\right) \tag{14.5}$$

where μ_{spread} is the mean spread and σ_{spread} is the spread volatility. Alternatively, we might use some heavy-tailed distribution to accommodate excess kurtosis in the spread.

We could now estimate the LVaR using Monte Carlo simulation: we could simulate both P and the spread, incorporate the spread into P to get liquidity-adjusted prices, and then infer the liquidity-adjusted VaR from the distribution of simulated liquidity-adjusted prices.

However, in practice, we might take a short-cut suggested by Bangia *et al.* (1999).[5] They suggest that we specify the liquidity cost (LC) as:

$$LC = \frac{P}{2}(\mu_{spread} + k\sigma_{spread}) \tag{14.6}$$

where k is some parameter whose value is to be determined. The value of k could be determined by a suitably calibrated Monte Carlo exercise, but they suggest that a particular value ($k = 3$) is plausible (e.g., because it reflects the empirical facts that spreads appear to have excess kurtosis and are negatively correlated with returns, etc.). The liquidity-adjusted VaR, *LVaR*, is

[4] For more on the bid–ask spread and some empirical results on its behaviour, see Muranaga and Ohsawa (1996).

[5] Their framework was further developed by Le Saout (2002), who also presented some empirical results that are similar to the illustrative ones presented here in suggesting that the liquidity adjustment can make a big difference to our VaR estimates.

then equal to the conventional VaR plus the liquidity adjustment (Equation (14.6)):

$$LVaR = VaR + LC = P[1 - \exp(\mu_R - \sigma_R z_\alpha) + \frac{P}{2}(\mu_{spread} + 3\sigma_{spread}] \tag{14.7}$$

Observe that this LVaR incorporates Equation (14.3) as a special case when $\sigma_{spread} = 0$. It therefore retains many of the properties of Equation (14.3), but generalises Equation (14.3) in allowing for the spread volatility as well. The ratio of LVaR to VaR is then:

$$\frac{LVaR}{VaR} = 1 + \frac{LC}{VaR} = 1 + \frac{1}{2}\frac{(\mu_{spread} + 3\sigma_{spread})}{[1 - \exp(-\sigma_R z_\alpha)]} \tag{14.8}$$

This immediately tells us that the spread volatility σ_{spread} serves to increase the liquidity adjustment relative to the earlier case.

Example 14.2 (Exogenous spread VaR)

Suppose we have the same parameters as in the previous example, but also have $\mu_{spread} = 0.02$ and $\sigma_{spread} = 0.005$. The ratio of *LVaR* to *VaR* is then:

$$\frac{LVaR}{VaR} = 1 + \frac{1}{2}\frac{0.02 + 3 \times 0.005}{[1 - \exp(-0.25 \times 1.645/\sqrt{250})]} = 1.6816$$

Allowing for the randomness of the spread increases the liquidity adjustment in this case from almost 40% to nearly 70%.

Box 14.1 Liquidation Strategies

A trader who wishes to liquidate a position over a certain period has a number of ways to do so. Typically, a strategy of selling quickly will involve high transactions costs – the more rapid the sale, the more pressure the trader puts on the market, and the worse the price he/she gets, the less time he/she has to shop around for a good deal, and so on – but also means that the trader rapidly reduces his/her exposure to loss from adverse price movements. On the other hand, a more leisurely strategy generally involves lower transactions costs, but a greater exposure over a longer period. There is therefore a trade-off between transactions costs and exposure costs.

How should we trade off transactions costs against exposure costs? One solution, suggested by Almgren and Chriss,[6] is that we begin by identifying this trade-off and estimating the set of efficient trading strategies that produce the minimum remaining risk exposure at any given point in time, for any given expected cost. Once we have identified the efficient trading strategies, we choose the one that best fits our risk-aversion. If we are risk-averse, we would choose a strategy that rapidly reduces our exposure, but at the cost of accepting a higher expected cost; and if we are less risk-averse, we would choose a strategy that leaves us more exposed, but is not likely to be so expensive.

[6] Almgren and Chriss (1999). See also Lawrence and Robinson (1995b,c) and Bertsimas and Lo (1998).

14.2.3 Endogenous-price Approaches

The previous approaches assume that prices are exogenous and therefore ignore the possibility of the market price responding to our trading. However, this is often unreasonable, and we may wish to make a liquidity adjustment that reflects the response of the market to our trading. If we sell, and the act of selling reduces the price, then this market-price response creates an additional loss relative to the case where the market price is exogenous, and we need to add this extra loss to our VaR. The liquidity adjustment will also depend on the responsiveness of market prices to our trade: the more responsive the market price, the bigger the loss.

We can estimate this extra loss in various ways,[7] but the simplest is to make use of some elementary economic theory. We begin with the notion of the price elasticity of demand, η, defined as the ratio of the proportional change in price divided by the proportional change in quantity demanded:

$$\eta = \frac{\Delta P/P}{\Delta N/N} < 0; \Delta N/N > 0 \tag{14.9}$$

where in this context N is the size of the market and ΔN is the size of our trade. $\Delta N/N$ is therefore the size of our trade relative to the size of the market. The impact of the trade on the price is therefore

$$\frac{\Delta P}{P} = \eta \frac{\Delta N}{N} \tag{14.10}$$

We can therefore estimate $\Delta P/P$ on the basis of information about η and $\Delta N/N$, and both of these can be readily guessed at using a combination of economic and market judgement. The LVaR is then:

$$LVaR = VaR\left(1 - \frac{\Delta P}{P}\right) = VaR\left(1 - \eta \frac{\Delta N}{N}\right) \tag{14.11}$$

bearing in mind that the change in price is negative. The ratio of LVaR to VaR is therefore:

$$\frac{LVaR}{VaR} = 1 - \eta \frac{\Delta N}{N} \tag{14.12}$$

This gives us a very simple liquidity adjustment that depends on two easily calibrated parameters. It is even independent of the VaR itself: the adjustment is the same regardless of whether the VaR is normal, lognormal, etc.

The ratio of *LVaR* to *VaR* thus depends entirely on the elasticity of demand η and the size of our trade relative to the size of the market ($\Delta N/N$).[8]

This approach is easy to implement, and is of considerable use in situations where we are concerned about the impact on VaR of endogenous market responses to our trading activity, as might be the case where we have large portfolios in thin markets. However, it is also narrow in focus and entirely ignores bid–ask spreads and transactions costs.

On the other hand, the fact that this approach focuses only on endogenous liquidity and the earlier ones focus on exogenous liquidity means that this last approach can easily be combined with one of the others: in effect, we can add one adjustment to the other. Thus, two very simple approaches can be added to produce an adjustment that addresses both exogenous and

[7] Other variants of this type of approach are suggested by Berkowitz (2000a) and Cosandey (2001).

[8] The ratio is also independent of the confidence level, which may or may not be desirable depending on our prior views of how the liquidity adjustment 'should' relate to the confidence level.

endogenous liquidity risk. This combined adjustment is given by

$$\left.\frac{LVaR}{VaR}\right|_{combined} = \left.\frac{LVaR}{VaR}\right|_{exogenous} \left.\frac{LVaR}{VaR}\right|_{endogenous} \tag{14.13}$$

Example 14.3 (Elasticity-based liquidity adjustment)

Suppose that we have a position worth 20% of the size of the market, so $\Delta N/N$ is 0.2, and we assume that the price elasticity of demand η is -0.5. It follows that $\Delta P/P = \eta \Delta N/N = -0.5 \times 0.2 = -0.1$. Using Equation (14.2), the ratio of *LVaR* to *VaR* is

$$\frac{LVaR}{VaR} = 1 - \frac{\Delta P}{P} = 1 + 0.1 = 1.1$$

In this case, the liquidity adjustment increases VaR by 10%.

Using the parameters of this example and the previous one, we can also estimate the combined adjustment factor:

$$\left.\frac{LVaR}{VaR}\right|_{combined} = 1.6815 \times 1.1 = 1.8487$$

14.2.4 The Liquidity Discount Approach

A more sophisticated approach is suggested by Jarrow and Subramanian (1997). They consider a trader who faces an optimal liquidation problem – the trader must liquidate his or her position within a certain period of time to maximise expected utility, and seeks the best way to do so. Their approach is impressive, as it encompasses exogenous and endogenous market liquidity, spread cost, spread risk, an endogenous holding period and an optimal liquidation policy.

Their analysis suggests that we should modify the traditional VaR in three ways. First, instead of using some arbitrary holding period, we should use an optimal holding period determined by the solution to the trader's expected-utility optimisation problem, which takes into account liquidity considerations and the possible impact of the trader's own trading strategy on the prices obtained. We should also add the average liquidity discount to the trader's losses (or subtract it from our prices) to take account of the expected loss from the selling process. Finally, their analysis also suggests that the volatility term should take account of the volatility of the time to liquidation and the volatility of the liquidity discount factor, as well as the volatility of the underlying market price.

To spell out their approach more formally, assume that prices between trades follow a geometric Brownian motion with parameters μ and σ. The current time is 0 and the price at time t is $p(t)$, so that geometric returns $\log(p(t)/p(0))$ are normally distributed. However, the prices actually obtained from trading are discounted from $p(t)$: more specifically, the prices obtained are $p(t)c(s)$, where $c(s)$ is a random quantity-dependent proportional discount factor, s is the amount traded, $0 \leq c(s) \leq 1$ and, other things being equal, $c(s)$ falls as s rises. Any order placed at time t will also be subject to a random execution lag $\Delta(S)$ and therefore take place at time $t + \Delta(s)$. Other things again being equal, the execution lag $\Delta(S)$ rises with s: bigger orders usually take longer to carry out. Our trader has S shares and wishes to maximise the present value of his/her current position, assuming that it is liquidated by the end of some horizon t, taking account of all relevant factors, including both the quantity discount $c(s)$ and

the execution lag $\Delta(s)$. After solving for this problem, they produce the following expression for the liquidity-adjusted VaR:

$$\begin{aligned} LVaR &= P\{E[\ln(p(\Delta(S))c(S)/p(0))] + std[\ln(p(\Delta(S))c(S)/p(0))]z_\alpha\} \\ &= P\left\{\left(\mu - \frac{\sigma^2}{2}\right)\mu_{\Delta(S)} + \mu_{\ln c(S)} + \left[\sigma\sqrt{\mu_{\Delta(S)}} + \left(\mu - \frac{\sigma^2}{2}\right)\sigma_{\Delta(S)} + \sigma_{\ln c(S)}\right]z_\alpha\right\} \end{aligned} \tag{14.14}$$

where all parameters have the obvious interpretations. This expression differs from the conventional VaR in three ways. First, the liquidation horizon t in the conventional VaR is replaced by the expected execution lag $\mu_{\Delta(S)}$ in selling S shares. Clearly, the bigger is S, the longer the expected execution lag. Second, the LVaR takes account of the expected discount $\mu_{\ln c(s)}$ on the shares to be sold. And, third, the volatility σ in the conventional VaR is supplemented by additional terms related to $\sigma_{\Delta(s)}$ and $\sigma_{\ln c(s)}$, which reflect the volatilities of the execution time and the quantity discount. Note, too, that if our liquidity imperfections disappear, then $\mu_{\Delta(S)} = t$, $\sigma_{\Delta(S)} = 0$, and $c(S) = 1$ (which in turn implies $\mu_{\ln c(s)} = \sigma_{\ln c(s)} = 0$) and our LVaR (Equation (14.14)) collapses to a conventional VaR as a special case – which is exactly as it should be.

To use this LVaR expression requires estimates of the usual Brownian motion parameters μ and σ, as well as estimates of the liquidity parameters $\mu_{\Delta(S)}$, $\sigma_{\Delta(S)}$, $\mu_{\ln c(S)}$, and $\sigma_{\ln c(s)}$, all of which are fairly easily obtained. The approach is therefore not too difficult to implement: all we have to do is then plug these parameters into Equation (14.13) to obtain our LVaR.[9]

Example 14.4 (Liquidity discount VaR)

To illustrate, suppose we are interested in using the Jarrow–Subramanian approach to estimate the LVaR at the 95% confidence level over a horizon of 1 day. We take the parameter values involved to be: $P = 1$, $\mu = 0$, $\sigma = 0.20/\sqrt{250}$, $\mu_{\Delta(S)} = 5/250$, $\sigma_{\Delta(S)} = 1/250$, $\mu_{\ln c(S)} = -0.125$ and $\sigma_{\ln c(S)} = 0.10$. To estimate *LVaR*, we plug these parameters into Equation (14.14) to obtain:

$$\begin{aligned} &\left(-\frac{0.20^2}{2 \times 250}\right)\frac{5}{250} - 0.125 + \left[\frac{0.20}{\sqrt{250}}\sqrt{\frac{5}{250}} - \left(\frac{0.20^2}{2 \times 250}\right)\frac{1}{250} + 0.10\right]1.645 \\ &= 0.0432 \end{aligned}$$

Using Equation (14.2), the conventional VaR is 0.0206, so the ratio of *LVaR* to *VaR* is $0.0401/0.0206 = 1.9466$.[10]

14.3 ESTIMATING LIQUIDITY AT RISK (LaR)

We turn now to liquidity at risk (LaR), sometimes also known as cash Flow at Risk (CFaR). LaR (or CFaR) relates to the risk attached to prospective cash flows over a defined horizon period, and can be defined in terms analogous to the VaR. Thus, the LaR is the maximum likely cash outflow over the horizon period at a specified confidence level: for example, the 1-day

[9] There are also other approaches not considered in the text. For example, Krakovsky (1999), Frey (2000) and Cherubini and Della Lunga (2001) look at how derivatives risk measures are affected by illiquidity in the underlying market. Hisata and Yamai (2000) look at optimal liquidation in a context of endogenous liquidity risk, and there is also the Giot and Gramming study mentioned in note 3.

[10] With these parameters, at least, the ratio of *LVaR* to *VaR* also rises with the confidence level.

LaR at the 95% confidence level is the maximum likely cash outflow over the next day, at the 95% confidence level, and so on. A positive LaR means that the likely 'worst' outcome, from a cash flow perspective, is an outflow of cash; and a negative LaR means that the likely worst outcome is an inflow of cash. The LaR is the cash flow equivalent to the VaR, but whereas VaR deals with the risk of losses (or profits), LaR deals with the risk of cash outflows (or inflows).

These cash flow risks are quite different from the risks of liquidity-related losses.[11] Nonetheless, they are closely related to these latter risks, and we might use LaR analysis as an input to evaluate them. Indeed, the use of LaR for such purposes is an important liquidity management tool.[12]

An important point to appreciate about LaR is that the amounts involved can be very different from the amounts involved with VaR. Suppose for the sake of illustration that we have a large market risk position that we hedge with a futures hedge of much the same amount. If the hedge is a good one, the basis or net risk remaining should be fairly small, and our VaR estimates should reflect that low basis risk and be relatively small themselves. However, the futures hedge leaves us exposed to the possibility of margin calls, and our exposure to margin calls will be related to the size of the futures position, which corresponds to the gross size of our original position. Thus, the VaR depends largely on the netted or hedged position, while the LaR depends on the larger gross position. If the hedge is a good one, the basis risk (or the VaR) will be low relative to the gross risk of the hedge position (or the LaR), and so the LaR can easily be an order of magnitude greater than the VaR. On the other hand, there are also many market risk positions that have positive VaR, but little or no cash flow risk (e.g., a portfolio of long European option positions, which generates no cash flows until the position is sold or the options expire), and in such cases the VaR will dwarf the LaR. So the LaR can be much greater than the VaR or much less than it, depending on the circumstances.

As we might expect, the LaR is potentially sensitive to any factors or activities, risky or otherwise, that might affect future cash flows. These include:

- Borrowing or lending, the impact of which on future cash flows is obvious.
- Margin requirements on market risk positions that are subject to daily marking-to-market.
- Collateral obligations, such as those on swaps, which can generate inflows or outflows of cash depending on the way the market moves. Collateral obligations can also change when counterparties like brokers alter them in response to changes in volatility, and collateral requirements on credit-sensitive positions (e.g., such as default-risky debt or credit derivatives) can change in response to credit events such as credit downgrades.
- Unexpected cash flows can be triggered by the exercise of options, including the exercise of convertibility features on convertible debt and call features on callable debt.
- Changes in risk management policy: for instance, a switch from a futures hedge to an options hedge can have a major impact on cash flow risks, because the futures position is subject to margin requirements and marking-to-market while a (long) option position is not.

[11] The link between cash flow risks and risks of loss associated with cash flow risks is important, and anyone who has any doubts on this needs to re-examine the Metallgesellschaft debacle of 1993. In the early 1990s, a US subsidiary of MG, MG Refining and Marketing, had sold a large number of long-term guarantees on the oil price, and it hedged the resulting oil price risk using futures and swaps. However, when oil prices fell in 1993, its hedge positions lost a lot of value, and MGRM faced large margin and collateral calls on them. These created a huge cash flow drain, and the firm ended up making a loss of about $1.3bn. The MG case shows very clearly that cash flow problems can easily lead to 'real' losses – and potentially very large ones too.

[12] Estimating these risks is critical if we are to manage them: it gives us an indication of our potential liquidity needs, so we can then arrange to meet them (e.g., by arranging lines of credit, etc.). With liquidity risks, it is also very important to have estimates of LaR over the whole of the foreseeable business horizon period – over the next day, week, month, etc. Failing to anticipate cash flow needs is one of the most serious (and, in many cases, also most elementary) errors that firms can make. A good LaR (or equivalent cash flow risk) system is therefore an essential risk management tool.

Two other points are also worth emphasising here. The first is that obligations to make cash payments often come at bad times for the firms concerned, because they are often triggered by bad events. The standard example is where a firm suffers a credit downgrade that leads to an increase in its funding costs, and yet this same event also triggers a higher collateral requirement on some existing (e.g., swap) position and so generates an obligation to make a cash payment. It is axiomatic in many markets that firms get hit when they are most vulnerable. The second point is that positions that might be similar from a market risk perspective (e.g., such as a futures hedge and an options hedge) might have very different cash flow risks. The difference in cash flow risks arises, not so much because of differences in market risk characteristics, but because the positions have different *credit* risk characteristics, and it is the measures taken to manage the credit risk – the margin and collateral requirements, etc. – that generate the differences in cash flow risks.

We can estimate LaR using many of the same methods used to estimate VaR and other measures of market risk.[13] One strategy, suggested by Singer (1997), is to use our existing VaR estimation tools to estimate the VaRs of marginable securities only (i.e., those where P/L translates directly into cash flows), thus allowing us to infer a LaR directly from the VaR. We could then combine this LaR estimate with comparable figures from other sources of liquidity risk within the organisation (e.g., such as estimates of LaR arising from the corporate treasury) to produce an integrated measure of firm-wide liquidity risk. The beauty of this strategy is that it makes the best of the risk measurement capabilities that already exist within the firm, and effectively tweaks them to estimate liquidity risks.[14]

However, this strategy is also fairly rough and ready, and cannot be relied upon when the firm faces particularly complex liquidity risks. In such circumstances, it is often better to build a liquidity risk measurement model from scratch. To do this we can start by setting out the basic types of cash flow to be considered. These might include:

- Known certain (or near certain) cash flows (e.g., income from government bonds, etc.): these are very easy to handle because we know them in advance.
- Unconditional uncertain cash flows (e.g., income from default-risky bonds, etc.): these are uncertain cash flows, which we model in terms of the pdfs (i.e., we choose appropriate distributions, assign parameter values, etc.).
- Conditional uncertain cash flows: these are uncertain cash flows that depend on other variables (e.g., a cash flow might depend on whether we proceeded with a certain investment, and so we would model the cash flow in terms of a pdf, conditional on that investment); other conditioning variables that might trigger cash flows could be interest rates, exchange rates, decisions about major projects, and so on.

Once we specify these factors, we can then construct an appropriate engine to carry out our estimations. The choice of engine would depend on the types of cash flow risks we have to deal with. For instance, if we had fairly uncomplicated cash flows we might use an HS or

[13] We can also estimate liquidity risks using the old spreadsheet methods (e.g., such as gap analysis) originally developed to look at bank interest-rate risk in the days before we had any substantial computer power. Such methods are useful for giving ballpark figures (e.g., much like duration figures can give us ballpark figures for interest-rate risk), but are certainly not a substitute for more sophisticated approaches.

[14] Another alternative is to use scenario analysis. We can specify liquidity scenarios, such as those arising from large changes in interest rates, default by counterparties, the redemption of putable debt, calls for collateral on repos and derivatives, margin calls on swaps or futures positions, and so on. We would then (as best we could) work through the likely/possible ramifications of each scenario, and so get an idea of the liquidity consequences associated with each scenario. Such exercises can be very useful, but, as with all scenario analyses, they might give us an indication of what could happen if the scenario occurs, but don't as such tell us anything about the probabilities associated with those scenarios or the LaR itself.

variance–covariance approach, or some specially designed term-structure model; however, since some cash flows are likely to be dependent on other factors such as discrete random variables (e.g., such as downgrades or defaults), it might be not be easy 'tweaking' such methods to estimate LaRs with sufficient accuracy. In such circumstances, it might be better to resort to simulation methods, which are much better suited to handling discrete variables and the potential complexities of cash flows in larger firms.

14.4 ESTIMATING LIQUIDITY IN CRISES

We now consider liquidity in crisis situations. As we all know, financial markets occasionally experience major crises – these include, for example, the stock market crash of 1987, the ERM crisis of 1992, and the Russian default crisis of the summer of 1998. Typically, some event occurs that leads to a large price fall. This event triggers a huge number of sell orders, traders become reluctant to buy, and the bid–ask spread rises dramatically. At the same time, the flood of sell orders can overwhelm the market and drastically slow down the time it takes to get orders executed. Selling orders that would take minutes to execute in normal times instead take hours, and the prices eventually obtained are often much lower than sellers had anticipated. Market liquidity dries up, and does so at the very time market operators need it most. Assumptions about the market – and in particular about market liquidity – that hold in 'normal' market conditions can therefore break down when markets experience crises. This means that estimating crisis liquidity is more than just a process of extrapolation from LaR under more normal market conditions: we need to estimate crisis-liquidity risks using methods that take into account the distinctive features of a crisis – large losses, high bid–ask spreads, and so on.[15]

One way to carry out such an exercise is by applying 'CrashMetrics'.[16] As we discussed in Chapter 13, we might have a position in a single derivatives instrument, and the profit/loss Π on this instrument is given by a delta–gamma approximation:

$$\Pi = \delta \Delta S + \frac{\gamma}{2}(\Delta S)^2 \tag{14.15}$$

where ΔS is the change in the stock price, and so on. The maximum loss occurs when $dS = -\delta/\gamma$ and is equal to:

$$L^{\max} = -\Pi^{\min} = \frac{\delta^2}{2\gamma}$$

The worst-case cash outflow is therefore $m\delta^2/(2\gamma)$, where m is the margin or collateral requirement. This approach can also be extended to handle the other Greek parameters (the vegas, thetas, rhos, etc.), multi-option portfolios, counterparty risk, and so on. The basic idea – of identifying worst-case outcomes and then evaluating their liquidity consequences – can also be implemented in other ways. For example, we might identify the worst-case outcome as the expected outcome at a chosen confidence level, and we could estimate this (e.g., using extreme-value methods) as the ES at that confidence level. The cash outflow would then be m times this ES.[17]

[15] For more on these methods, see BIS (2000).

[16] Wilmott (2000), Chapter 58.

[17] There are also other ways we can estimate crisis LaR. Instead of focusing only on the high losses associated with crises, we can also take account of the high bid–ask spreads and/or the high bid–ask spread risks associated with crises. We can do so, for example, by estimating these spreads (or spread risks), and inputting these estimates into the relevant liquidity-adjusted VaR models discussed in section 14.2.

However, these suggestions (i.e., Greek and ES based) are still rather simplistic, and with complicated risk factors – such as often arise with credit-related risks – we might want a more sophisticated model that was able to take account of the complications involved, such as:

- The discreteness of credit events.
- The interdependency of credit events.
- The interaction of credit and market risk factors (e.g., the ways in which credit events depend in part on market risk factors).
- Complications arising from the use of credit-enhancement methods such as netting arrangements, periodic settlement, credit derivatives, credit guarantees, and credit triggers.[18]

These complicating factors are best handled using simulation methods tailor-made for the problems concerned.

The obvious alternative to probabilistic approaches to the estimation of crisis liquidity is to use crisis-scenario analyses, discussed in the last chapter. We would imagine a big liquidity event – a major market crash, the default of a major financial institution or government, the outbreak of a war, or whatever – and work through the ramifications for the liquidity of the institution concerned. One attraction of scenario analysis in this context is that we can work through scenarios in as much detail as we wish, and so take proper account of complicated interactions such as those mentioned in the last paragraph. This is harder to do using probabilistic approaches, which are by definition unable to focus on any specific scenarios. However, as with all scenario analysis, the results of these exercises are highly subjective, and the value of the results is critically dependent on the quality of the assumptions made.

[18] For more on these methods and their liquidity implications, see, e.g., Wakeman (1998).

15

Backtesting Market Risk Models

Once a risk model is constructed, it is important that it be carefully validated before being put to practical use, and that its performance be regularly evaluated after that. Risk models need to be regularly validated, and a key feature of model validation is backtesting – the application of quantitative methods to determine whether the forecasts of a VaR forecasting model are consistent with the assumptions on which the model is based, or to rank a group of such models against each other. We will have more to say on the more general or qualitative issues involved with model validation in the next chapter, but in this chapter we focus on the quantitative methods (i.e., formal backtesting).

This chapter covers six main topics:

- Preliminary data issues: the collection of data suitable for backtesting purposes, preliminary analysis, and so on.
- Statistical backtests based on the frequency of exceedances, the number of times we get losses exceeding VaR.
- Statistical backtests based on the whole of the distribution of exceedances, or the whole of the distribution of P/L. These tests use more information than the frequency-based tests, and are generally more reliable.
- Comparing and ranking alternative models.
- Backtesting with alternative positions and data.
- Assessing the accuracy of backtest exercises: the use of simulation methods (including the bootstrap) to evaluate the precision of backtest results.

The focus of the discussion in this chapter is on the application of quantitative (and mainly statistical) methods to backtesting problems. The statistical tests used are discussed in more detail in the appendix to this chapter.

15.1 PRELIMINARY DATA ISSUES

The first requirement in backtesting is to obtain suitable data. This is a more difficult problem than it first appears to be, because P/L data are typically calculated according to standard principles of accounting prudence, which often means that assets are understated in value and fluctuations in their values are smoothed over. However, for risk measurement purposes it is more important that we use P/L data that reflect underlying volatility rather than accounting prudence.

Our P/L data also need cleaning to get rid of components that are not directly related to current or recent market risk-taking. Such components include fee income, hidden profits/losses from trades carried out at prices different from the mid-bid–ask spread, P/L earned from other forms of risk-taking (e.g., high yields on bonds with high credit risks), and unrealised P/L and provisions against future losses. We also need to take account of the impact of the internal

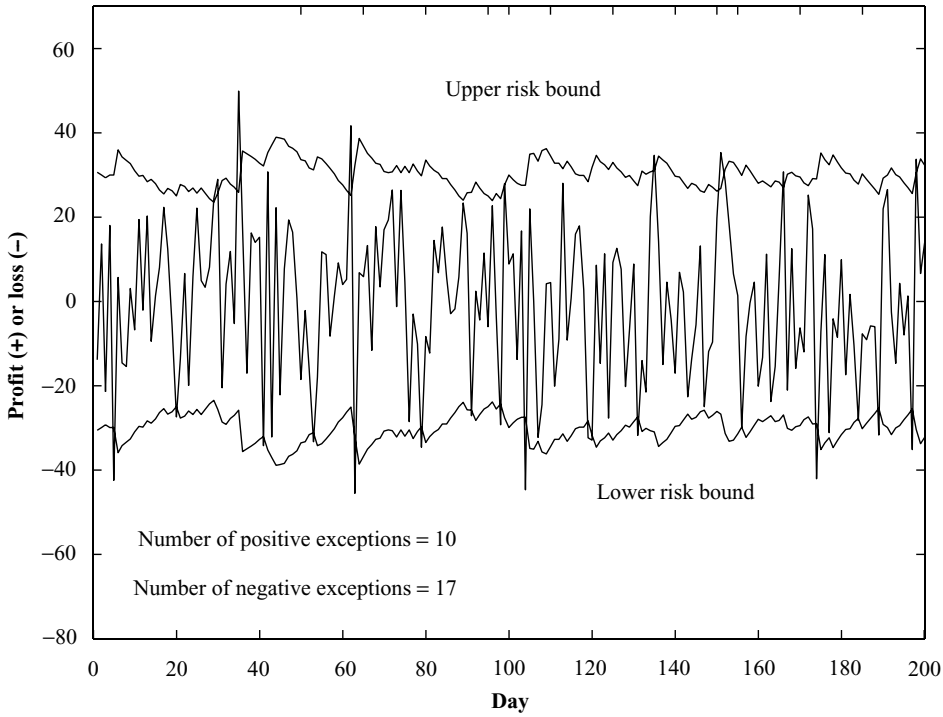

Figure 15.1 A backtesting chart

funding regime that underlies the institution's trading activity, and of the impact of intra-day trading on both P/L and risk measures. To compare P/L against market risk, we should therefore either clean the P/L data so that they (as much as possible) reflect the P/L on end-of-day market risk positions, or we should use hypothetical P/L data obtained by revaluing trading positions from one day to the next.

Having obtained (reasonably) clean data, it can be very useful to draw up a chart like the one shown in Figure 15.1. This chart can be drawn up at institutional or business-unit level, and shows the time series (i.e., sequential values) of both daily P/L and risk measures (e.g., VaRs at 5% and 95% confidence levels) delineating 'regular' profits or losses from more extreme ones. This chart shows how these series have behaved over time, and gives a good visual indication of the behaviour of the outliers or exceptions – the very high profits above the upper risk bound and the very large losses lying below the lower risk bound in the chart. It also shows how many exceptions there were, how big they were, and whether they show any pattern. Such a chart gives a good indication of possible underlying causes:

- A relatively large number of extreme observations indicates that our risk measures are probably too low.
- A relatively small number of tail observations, or none at all, indicates that our risk measures are probably too high.

- If there are major differences between high and low exceptions, then our P/L measures might be biased.
- If the risk lines show flatness, or excessive smoothness, risk measures are not being updated sufficiently quickly.
- If P/L is close to zero much of the time, then there is relatively little trading taking place, and this suggests that positions are illiquid.
- Abrupt changes in risk lines suggest changes in volatility or changes in the way the risks are estimated.

This particular figure shows a backtesting chart for a hypothetical portfolio daily P/L series and the associated 5% VaRs and 95% VaRs, which generally lie on either side of the P/L series. All three series are fairly stable, and show no obvious deformities. Given the number of observations (200) and the VaR confidence levels, we would expect 10 positive and 10 negative exceptions, and we actually get 10 positive and 17 negative exceptions. The number of negative exceptions (or tail losses) is well above what we would expect. In any practical situation, the risk practitioner faced with these results would be advised to look into this further.

It is good practice to supplement backtesting charts with P/L histograms, which sometimes give a clearer indication of the empirical distribution of P/L, and (perhaps) QQ or similar charts, which can help to give a broad indication of whether the empirical P/L distribution is consistent with the risk model. It is also a good idea to examine summary P/L statistics, including the obvious statistics of mean, standard deviation, skewness, kurtosis, volatility, range, etc. and basic statistics on the number and size of extreme observations. Such information can be very useful in helping practitioners to get to know their data and get a feel for any problems they might encounter.

15.2 BACKTESTS BASED ON FREQUENCY TESTS

Having completed our preliminary data analysis, we now proceed to formal statistical backtesting. This is based on a standard hypothesis testing paradigm. We first specify the null hypothesis that we wish to test, and also select an alternative hypothesis to be accepted if the null is rejected. We then select a significance level and estimate the probability associated with the null hypothesis being 'true'. Typically, we would accept the null hypothesis if the estimated value of this probability, the estimated prob-value, exceeds the chosen significance level, and we would reject it otherwise. The higher the significance level, the more likely we are to accept the null hypothesis, and the less likely we are to incorrectly reject a true model (i.e., to make a Type I error). Unfortunately, it also means that we are more likely to incorrectly accept a false model (i.e., to make a Type II error). Any test therefore involves a trade-off between these two types of possible error. Ideally, we should select a significance level that takes account of the likelihoods of these errors (and, in theory, their costs as well) and strikes an appropriate balance between them. However, in practice, it is common to select some arbitrary significance level such as 5% and apply that level in all our tests. A significance level of this magnitude gives the model a certain benefit of the doubt, and implies that we would reject the model only if the evidence against it is reasonably strong.

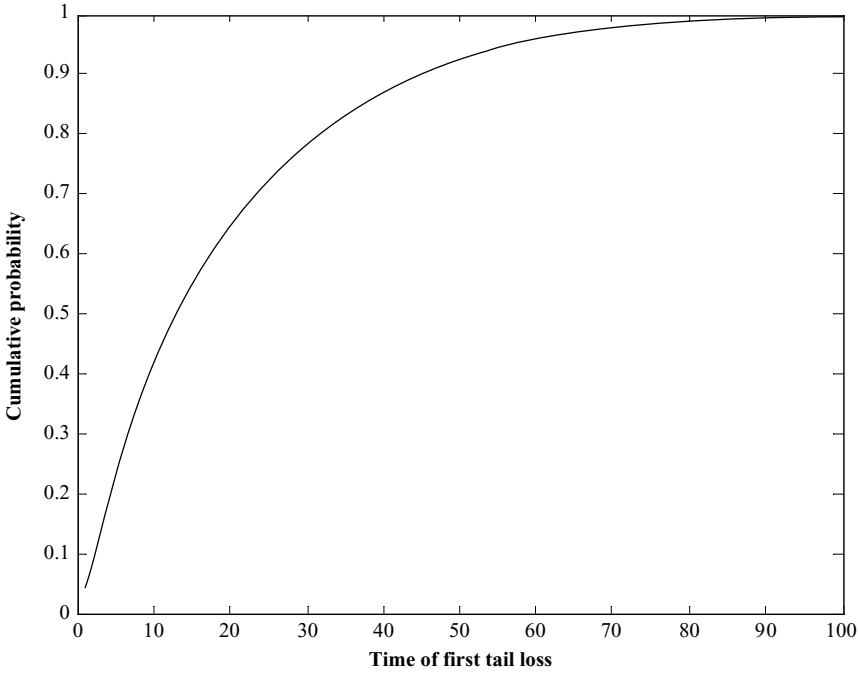

Figure 15.2 Probabilities for the time of first exceedance
Note: Estimated for an assumed p-value of 5%.

Box 15.1 The Time-to-first-exceedance Test

One simple backtest focuses on the time when the first exceedance occurs. This can be useful when long runs of data are not available, such as in the period shortly after we have introduced a new risk model or made major changes to an existing one. If the probability of an exceedance is p, the probability of observing the first exceedance in period T is $p(1-p)^{T-1}$, and the probability of observing the first exceedance by period T is $1-(1-p)^T$, which obeys a geometric distribution.

These probabilities are easily calculated, and Figure 15.2 shows a plot of the probability of observing our first exceedance by period T, for a p-value of 5%. The figure shows that the probability of observing our first exceedance by time T rises with T itself – for example, at $T=5$, the probability of having observed a first exceedance is 22.6%; but by $T=50$, the same probability is 92.3%.

However, this test is inferior to the basic frequency test because it uses less information: it only uses information since the previous tail loss, and so ignores other information we might have.

15.2.1 The Basic Frequency Backtest

Perhaps the most widely used test is the basic frequency (or binomial) test suggested by Kupiec (1995). The idea is to test whether the observed frequency of tail losses (or frequency of losses

that exceed VaR) is consistent with the frequency of tail losses predicted by the model. In particular, under the null hypothesis that the model is 'good' (i.e., consistent with the data), the number of tail losses x follows a binomial distribution. Given n P/L observations and a predicted frequency of tail losses equal to p, this tells us that the probability of x tail losses is:

$$\Pr(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{15.1}$$

where p is of course equal to 1 minus the confidence level.

The above equation also tells us that the only information required to implement a binomial test is information about the values of n, p, and x. Of these, the sample size n is obtained by simply counting the number of observations, p is apparent from the confidence level, and x is obtained by counting the number of observations in which a loss exceeds the VaR. The test statistic is then calculated using a suitable calculation engine (e.g., using the 'binomdist' function in Excel or the 'binocdf' function in MATLAB).

To illustrate, suppose we have a random sample of $n = 1000$ P/L observations drawn from a portfolio. We take the confidence level α to be 0.95, so our model predicts that $p = 1 - \alpha = 0.05$. Our null hypothesis is then $H_0 : p = 0.05$ which predicts $np = 50$ exceedances in our sample. Now suppose that with this sample the number of exceedances, x, is 55, say. This corresponds to an empirical frequency, $\hat{p}$, equal to 0.055. Since $\hat{p}$ exceeds p, it is plausible in this situation to specify a one-sided alternative hypothesis $H_1 : p > 0.05$. The prob-value of the test is then the probability under the null that $x \geq 55$. This is most easily calculated as $1 - \Pr[x \leq 54]$, which equals 0.25 given our values for n and p. At a standard significance level such as 5%, we would then have no hesitation in 'passing' the model as acceptable.[1]

The relationship between x and the binomial test prob-value is shown in Figure 15.3. As we can see, the prob-value declines as x gets bigger, and there comes a point where the number of tail losses, x, is so high that it is no longer credible to maintain that the model is sound. The cut-off point is determined by our significance level, here set at 5%: values of x less than or equal to 62 have a prob-value that exceeds the significance level, and values of x greater than 62 have a prob-value less than 5%. Hence, we accept the model if $x \leq 62$ and we reject it if $x \geq 63$.

The above discussion was based on the premise that x exceeds the expected number of exceedances. Had x been less than np, the most natural alternative hypothesis would have been $H_1 : p < 0.05$. The prob-value associated with the null would then be the probability that x is less than or equal to its observed value. Proceeding as before, we would end up accepting the null if $x \geq 39$ and rejecting it if $x \leq 38$.[2]

If we wish to, we can also apply binomial tests using a two-sided alternative hypothesis. We could do so by estimating a confidence interval for the number of tail losses based on the model's prediction, and then checking whether the observed number of exceedances lies within this interval.[3] We can construct a confidence interval for the number of exceedances using the

[1] The text implements the binomial test in terms of the number of exceedances x, but we could equally well have chosen to work with the frequency of exceedances, x/n. The former is more common in the risk management area, but the latter is more common in statistical textbooks. However, the two approaches are equivalent because n is a constant. Had we adopted the latter approach, we would have tested whether the empirical frequency of exceedances (0.055) was significantly different from the predicted frequency (0.05), but these would have produced exactly the same test statistics as in the text.

[2] The critical cut-off point is $x = 39$ because $\Pr[x \leq 39] = 5.98\%$, and $\Pr[x \leq 38] = 4.33\%$.

[3] We can also do the test the other way round, in which case we would calculate the confidence interval around the observed number of tail losses, and see if the interval included the predicted number. However, either way we would get the same prob-value for our null.

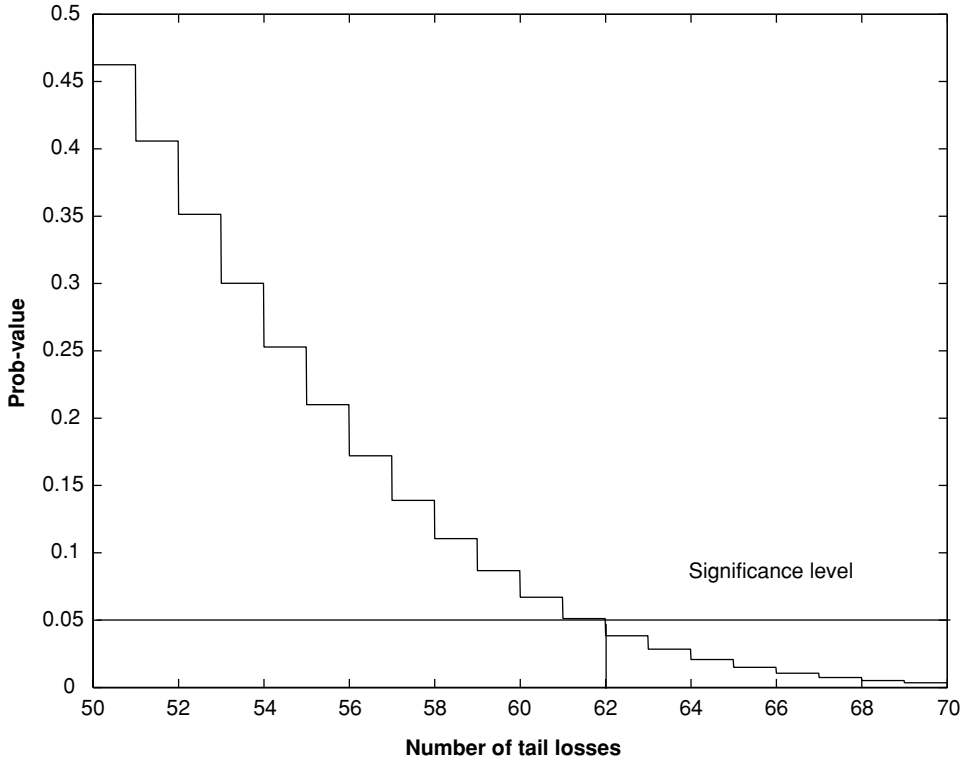

Figure 15.3 The binomial test

inverse of the tail-loss binomial distribution (e.g., using the 'binofit' function in MATLAB). Given $n = 1000$ observations and $p = 0.05$, the predicted 95% confidence interval for the number of exceedances turns out to be [37, 65]. We would therefore accept the null if x falls within this range, and reject it otherwise.

These examples also highlight the importance of the alternative hypothesis. The three cases are summarised in Table 15.1. In the first case, our alternative hypothesis is $H_1 : p > 0.05$ and our acceptable x range is [51, 62]. In choosing this alternative, we implicitly use the observation that x exceeds its expected value to presume that the true value of p was not less than 0.05. In the second case, our alternative hypothesis is $H_1 : p < 0.05$ and our acceptable x range is [39, 49]; this reflects the converse presumption that the true value of p is not more than

Table 15.1 The alternative hypothesis and the acceptance/rejection regions of the binomial test

H_1	Acceptance region	Rejection region	Comment
$p > 0.05$	x in [51, 62]	$x \geq 63$	Given that $x > 50$, we rule out $p < 0.05$.
$p < 0.05$	x in [39, 49]	$x \leq 38$	Given that $x < 50$, we rule out $p > 0.05$.
$p \neq 0.05$	x in [37, 65]	$x \leq 36; x \geq 66$	We allow unrestricted possibilities for p.

0.05. In the third case, we retain a more open mind about the true value of p, our alternative hypothesis is $H_1 : p \neq 0.05$, and we get an acceptable x range of [37, 65]. Thus, the range of x-values considered acceptable depends on the alternative hypothesis (i.e., how we specify the alternative hypothesis can influence the outcome of the test). For example, if we have an observed x-value in the range from 63 to 65, then we would have rejected the null had we specified the one-sided alternative hypothesis $H_1 : p > 0.05$, but we would have accepted it had we specified the two-sided alternative $H_1 : p \neq 0.05$. The reason for this difference is that a one-sided test has a single rejection region represented by (depending on the chosen alternative hypothesis) the upper or lower 5% tail, whereas the two-sided test has two rejection regions, the lower 2.5% tail and the upper 2.5% tail.

These basic frequency tests have a simple intuition, are easy to apply and do not require a great deal of information. However, they lack power (i.e., the ability to identify bad models) except with very large sample sizes, because they throw potentially valuable information away. This loss of information is the downside of the otherwise attractive fact that the binomial test can be implemented using knowledge of only n, p, and x. The information discarded is of two kinds:

- Since they focus exclusively on the frequency of exceedances over the sample period, these tests throw away information about the temporal pattern of exceedances. However, this can be important information, because many risk models predict that exceedances should be independently and identically distributed – that is to say, many models predict that the probability of a tail loss is constant and independent of whether or not an exceedance occurred the previous period.[4]
- Frequency tests throw away (often useful) information on the sizes of tail losses predicted by risk forecasting models. This has the unfortunate implication that a 'bad' risk model will pass a frequency test if it generates an acceptably accurate frequency of exceedances, even if its forecasts of losses larger than VaR are very poor.

Hence, we would expect that basic frequency tests should be less reliable than tests that take account of the information they throw away, and this expectation turns out to be broadly correct.

Example 15.1 (One-sided frequency test)

Suppose that we have 500 P/L observations and a VaR confidence level α of 95%. Hence, $n = 500$ and $p = 1 - \alpha = 0.05$. The expected number of losses is $np = 25$. Using the 'binocdf' function in MATLAB, the probability of 30 or more exceedances is $1 - binocdf(29,500,0.05) = 0.1765$, and the probability of 40 or more exceedances is $1 - binocdf(39,500,0.05) = 0.0027$. Thus, 30 exceedances is plausible, but 40 is not.

Similarly, the probability of 20 or fewer exceedances is $binocdf(20,500,0.05) = 0.1789$, and the probability of 10 or fewer is $binocdf(10,500,0.05) = 0.0005$. In this case, 20 exceedances is plausible, but 10 is not.

---

[4] The iid prediction for exceedance events arises naturally in circumstances where the forecast horizon period coincides with the data observation period (e.g., where we have a VaR holding period equal to a day, and we observe P/L daily). However, if consecutive horizon periods overlap with each other, then the iid prediction does not hold. We discuss this further in section 15.3.3.

Example 15.2 (Two-sided frequency test)

Given the same n and p as before, what is the 95% confidence interval for the number of excess losses, given an observed number of excess losses, x, equal to 30?

Using the 'binofit' function in MATLAB with $x = 30$ and $n = 500$, the 95% confidence interval for the probability of an exceedance is [0.0408, 0.0845]. As this includes the value of $p = 0.05$ on which the model is predicated, we can accept the model at the 5% significance level. The confidence interval for the number of exceedances is therefore $500 \times [0.0408, 0.0845] \approx [20, 42]$. This interval includes the number predicted by the model, which again tells us that the model is acceptable at the 5% significance level.

Box 15.2 Regulatory Backtesting Requirements

Commercial banks in the G-10 countries are obliged to carry out a set of standardised backtests prescribed by the 1996 Amendment to the 1988 Basel Accord, which lays down capital adequacy standards for commercial banks. The main features of these regulations are:

- Banks must calibrate daily VaR measures to daily P/L observations, and these VaRs are predicated on a confidence level of 99%.
- Banks are required to use two different P/L series: actual net trading P/L for the next day; and the theoretical P/L that would have occurred had the position at the close of the previous day been carried forward to the next day.
- Backtesting must be performed daily.
- Banks must identify the number of days when trading losses, if any, exceed the VaR.

The results of these backtests are used by supervisors to assess the risk models, and to determine the multiplier (or hysteria) factor to be applied: if the number of exceptions during the previous 250 days is less than 5, the multiplier is 3; if the number of exceptions is 5, the multiplier is 3.40, and so on; and 10 or more exceptions warrant a multiplier of 4.

Leaving aside problems relating to the capital standards themselves, these backtesting rules have a number of weaknesses:

- The frequency test used has known deficiencies (e.g., it ignores the sizes and patterns of exceptions, is unreliable except with very large samples, etc.).
- Models can fail the regulatory backtests in abnormal situations (e.g., such as a market crash or natural disaster) and lead banks to incur unwarranted penalties.
- The rules relating the capital multiplier to the number of exceptions are arbitrary, and there are concerns that the high scaling factor could discourage banks from developing and implementing best practice.

However, even if these problems were dealt with or at least ameliorated, there would always remain deeper problems: any regulatory prescriptions would inevitably be crude, inflexible, possibly counterproductive, and certainly behind best market practice.[5]

[5] For more on regulatory backtesting, see, e.g., Crouhy *et al.* (1998).

15.2.2 The Conditional Testing (Christoffersen) Backtest

Assuming that we have a risk forecasting model that predicts that exceedances are iid, one way to take account of this prediction is the conditional backtesting procedure suggested by Christoffersen (1998). His idea is to separate out the particular predictions being tested, and then test each prediction separately. The first of these is the prediction just examined, namely, that the model generates the 'correct' frequency of exceedances, which in this context we would describe as the prediction of correct unconditional coverage. The other prediction is that exceedances are independent of each other. This latter prediction is important insofar as it suggests that exceedances should not be clustered over time. Evidence of exceedance clustering would suggest that the model is misspecified, even if the model passes the prediction of correct unconditional coverage.

To appreciate the Christoffersen approach, we begin by rephrasing the earlier basic frequency or coverage test in likelihood ratio (LR) form. If x is the number of exceedances in our sample, and n is the number of observations, then the observed frequency of exceedances is x/n. Given that the predicted probability of exceedances is p, the earlier test can also be expressed in terms of an LR test. More particularly, under the hypothesis/prediction of correct unconditional coverage, the test statistic

$$LR_{uc} = -2\ln[(1-p)^{n-x}p^x] + 2\ln[(1-x/n)^{n-x}(x/n)^x] \tag{15.2}$$

is distributed as a $\chi^2(1)$. As we can see from Equation (15.2), this boils down to a test of whether the empirical frequency x/n is sufficiently close to the predicted frequency p.

Turning to the independence prediction, let n_{ij} be the number of days that state j occurred after state i occurred the previous day, where the states refer to exceedances/non-exceedances, and let π_{ij} be the probability of state j in any given day, given that the previous day's state was i. Under the hypothesis of independence, the test statistic

$$LR_{ind} = -2\ln[(1-\hat{\pi}_2)^{n_{00}+n_{11}}\hat{\pi}_2^{n_{01}+n_{11}}] + 2\ln[(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}{}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\pi_{11}{}^{n_{11}}] \tag{15.3}$$

is also distributed as a $\chi^2(1)$, and noting that we can recover estimates of the probability from:

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}}, \quad \hat{\pi}_{11}\frac{n_{11}}{n_{10}+n_{11}}, \quad \hat{\pi}_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}} \tag{15.4}$$

It follows that under the combined hypothesis of correct coverage and independence – the hypothesis of correct conditional coverage – the test statistic

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{15.5}$$

is distributed as $\chi^2(2)$. The Christoffersen approach enables us to test both coverage and independence hypotheses at the same time. Moreover, if the model fails a test of both hypotheses combined, his approach enables us to test each hypothesis separately, and so establish where the model failure arises (e.g., does the model fail because of incorrect coverage, or does it fail because of lack of independence?).

Example 15.3 (LR frequency of excess losses test)

Suppose that we have $n = 500$, $p = 1-\alpha = 0.05$, and $x = 30$. We wish to evaluate the hypothesis of correct unconditional coverage (i.e., that the model generates the 'true' frequency of exceedances) using an LR test. To carry out this test, we plug these parameter values into

Equation (15.2) to get our test statistic:

$$\begin{aligned} LR_{uc} &= -2\ln[(1-p)^{n-x}p^x] + 2\ln[(1-x/n)^{n-x}(x/n)^x] \\ &= -2\ln[(1-0.05)^{500-30}0.05^{30}] + 2\ln[(1-30/500)^{500-30}(30/500)^{30}] \\ &= 0.9921 \end{aligned}$$

The test statistic is distributed as $\chi^2(1)$, and the probability of a test value of 0.9921 or more under a $\chi^2(1)$ is 0.3192, which indicates that the null hypothesis is (easily) acceptable: we cannot reject the hypothesis that the model generates the correct frequency of tail losses.

Example 15.4 (LR independence test)

Suppose we have the same parameters as in Example 15.3, but we are now also told that all our excess losses occur consecutively in the middle of our sample period, representing a very extreme case of clustering. This information tells us that $n_{00} = 469$, $n_{01} = 1$, $n_{10} = 1$ and $n_{11} = 29$. Equations (15.4) then imply that

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}} = \frac{1}{470} = 0.0021, \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}} = \frac{29}{30} = 0.9667,$$
$$\hat{\pi}_2 = \frac{n_{11}}{n_{10}+n_{11}} = \frac{1+29}{469+1+1+30} = \frac{30}{500}$$

Substituting these values into Equation (15.3) gives us the value of the LR test statistic

$$\begin{aligned} LR_{ind} = &-2\ln[(1-30/500)^{469+29}(30/500)^{1+29}] \\ &+ 2\ln[(1-0.0021)^{469}0.0021^1(1-.9667)^1 0.9667^{29}] = 207.36 \end{aligned}$$

Given that this test statistic is distributed as $\chi^2(1)$, the prob-value associated with this test value is negligible, which indicates a strong rejection of the model. This is the correct answer given the extreme clustering of excess losses in the middle of the sample period.

Box 15.3 Tests of Independence

Many risk models predict that excess losses should be independently and identically distributed, and there are various ways in which this prediction can be tested. One way is through an LR test, as suggested by Christoffersen. We can also estimate the autocorrelation structure of our z*t* observations (see section 15.3.2), which should be (jointly) insignificant if the observations are independent, and then test for independence by testing the joint significance of the empirical autocorrelations (e.g., using Box–Pierce Q statistics, which are very easy to calculate; cf. section 15.3.2.2). Other possibilities are to apply a runs test (see the appendix to this chapter) or to estimate a binary regression model – a model whose dependent variable takes the value 1 if there is an exceedance or 0 if there is no exceedance – and then test for the joint insignificance of the explanatory variables (e.g., using F or LR statistics).[6] A binary regression approach is more powerful than a basic runs test because it can take account of the impact of conditioning variables, which a runs test does not. A final possibility, if we have enough data, is to test independence using a BDS test (Brock *et al.* (1987)): a BDS test is very powerful, but also data intensive.

[6] A good example of how this might be done is provided by Engle and Manganelli (1999).

15.3 BACKTESTS BASED ON TESTS OF DISTRIBUTION EQUALITY

The tests considered so far all share one common feature: they focus exclusively on the frequency of exceedances, and therefore throw away information about their sizes. Yet information about the sizes of exceedances is potentially useful for assessing model adequacy, and we would expect tests that use such information to be more reliable than tests that use only frequency information. We also have to wonder whether it is helpful to throw away the non-tail information, in which case we might wish to consider the whole P/L distribution, and not just its tails. This suggests that we should compare predicted distributions (of complete P/L or just tail losses) against predicted distributions.

We need to consider carefully what information such a test would require. Where the VaR is parametric, it would require the full specification of the VaR forecasting model (i.e., information about the nature of the parametric form assumed by the model and about the values of the parameters involved for each set of forecasts). Where VaR is non-parametric, it would require the complete set of (P/L or tail-loss) forecasts. However, this means that we need, in principle, a different set of forecasts for each day in our sample: the set of forecasts for day 1 will be different from that for day 2, and so on. It then becomes apparent that we have a rather awkward problem. In effect, we have a whole set of forecasted distributions – the first for day 1, the second for day 2, and so on – and we have a set of realised P/L values, each drawn (under the null) from its own contemporaneous forecast distribution. So, with n observations in our sample, we have n different forecasted distributions, and a single realised value drawn from each of these distributions. Except in the implausible special case where the forecasted distributions happened to be constant, this is most definitely *not* a single forecasted distribution and a set of realised losses drawn from that distribution.

How then do we compare the P/Ls against the forecasts, when the forecasts themselves (typically) change from one day to the next?

15.3.1 Tests Based on the Rosenblatt Transformation

The answer is to transform our data so that we are dealing with them in terms of their forecast percentiles or forecast cumulative probability values. Thus, if x_t is the day-t P/L value, and this observation is associated with a forecasted cumulative density function $F_t(.)$, and which in principle might change from one day to the next, then the transformed observation is the value of $F_t(.)$ evaluated at x_t:

$$p_t = F_t(x_t) \tag{15.6}$$

For example, if our risk model is standard normal, then a P/L of 1.645 would be mapped to 0.95; a loss of −1.645 would be mapped to 0.05; and so on. This transformation is known as the Rosenblatt transformation, and ensures that our data become standardised in units of the same metric (i.e., in terms of forecasted cdfs). This standardisation makes our observations directly comparable to each other, which they otherwise would not be.

To illustrate, suppose our risk model assumes that P/L is normal and we have the observations and forecasts shown in Table 15.2. The first column in the table refers to the date to which the P/L and forecast parameters refer. The second column gives realised P/L, and the next two columns give forecasts of the mean and standard deviation of P/L. The fifth column

Table 15.2 An example of the Rosenblatt transformation

Observation date	Realised P/L	Forecast mean	Forecast std	Normal cdf value
1	12.507	10	20	0.550
2	15.041	9	21	0.613
3	−14.222	11	22	0.126
4	36.391	9	23	0.883
5	9.172	10	22	0.485

then shows the values of the normal cdf associated with the P/L realisations and the associated parameter forecasts, and these cdf-values are our Rosenblatt-transformed observations. So, for example, the first observation has a realised P/L of 12.507, a forecast mean of 10 and a forecast std of 20, and this maps to a cdf value of 0.550. The other observations are comparable. The transformation takes each realised P/L through the relevant parameterised normal cdf function to give us the cdf value. Naturally, if our forecasting model does not assume normality, we would replace the normal cdf with the cdf appropriate to our model (e.g., a *t*-cdf if our model is based on a *t*-distribution, empirical cdfs if our model is non-parametric, etc.).

The application of the Rosenblatt transformation paves the way for us to apply distribution-equality tests to assess model adequacy. In particular, under the null hypothesis that the model is adequate, we would expect the lowest 10% of transformed observations to fall in the region between 0 and 0.10, the next lowest 10% of observations to fall between 0.10 and 0.20, and so on. So, under the null hypothesis that the model is adequate, the Rosenblatt transformed data are predicted to be distributed as standard uniform (i.e., U(0,1)).

The same approach can also be used if we are interested in the distribution of tail losses rather than the complete P/L distribution. Let us suppose that we have put all our P/L data through a Rosenblatt transformation. The mapped observations would be predicted to be distributed U(0,1) under the null, but we now delete all observations except those associated with losses exceeding VaR. If the model is good, the empirical percentiles left should be close to uniformly distributed over the interval $[0,1-\alpha]$. We can now map the tail observations to *their* percentile points, rather than to percentile points of the P/L distribution as a whole, and the resulting distribution of tail losses will be U(0,1) under the null. For example, suppose $\alpha = 0.95$, and our first P/L exceedance observation turns out to be −28 (i.e., a loss of 28) and is drawn (under the null) from a normal with mean 10 and std 20. This observation has a cdf (or Rosenblatt) value of 0.0287. The other observations will also lie between 0 and 0.05, and will be uniformly distributed over that interval under the null. However, a percentile of 0.0287 on the original P/L distribution is equivalent to a percentile $1 - 0.0287/(1-\alpha) = 1 - 0.0287/0.05 = 0.426$ on the distribution of (Rosenblatt-transformed) *tail loss observations*, which are distributed as U(0,1) under the null. In short, regardless of whether we are dealing with the complete P/L distribution or just the distribution of tail losses, we can always obtain a Rosenblatt-transformed series that is distributed as standard uniform under the null.[7]

[7] There remains the question of whether one should work with the whole P/L distribution, or just the tail losses. In principle, the answer depends on the problem at hand. However, for many risk management problems, we are more interested in the tails than in the central mass observations, and in such circumstances it will often be more helpful to focus on the tails and therefore work with the distribution of tail losses.

We can then evaluate our model by testing this prediction, which we can do by applying a distribution-equality test suitable for testing whether an empirical distribution is distributed as standard uniform. We can test this prediction using an appropriate distribution-equality test, as discussed further in the appendix.[8]

15.3.2 Tests using the Berkowitz Transformation

We can improve our testing procedure further by putting our series through a second transformation. Instead of testing the prediction that the forecast percentiles, p_t, should be U(0,1) distributed under the null, Berkowitz (2001) suggests that we transform the p_t to make them standard normal under the null. We can do so by applying an inverse normal transformation, making use of the result that if p_t is U(0,1), then $z_t = \Phi^{-1}(p_t)$ is N(0,1). The Berkowitz transformation converts a uniform series into a standard normal series, and therefore enables us to test model adequacy by means of tests for standard normality. One attraction of this suggestion is that once the data are 'transformed to normal', we can apply much more powerful statistical tools than we can apply directly to uniform data. In particular, it now becomes possible to apply a battery of powerful tests and identify more clearly the possible sources of model failure. It also enables us to test any independence prediction, where it arises, within the same framework used to test the rest of the null.

Assume for the time being that there is an iid prediction, in which case the full null prediction is that z_t is iid N(0,1). Berkowitz suggests that we can test this by nesting the null hypothesis within a fairly general first-order autoregressive process with a possibly different mean and variance. If we write this process as:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t \tag{15.7}$$

then the null hypothesis predicts that $\mu = 0$, $\rho = 0$, and σ^2, the variance of ε_t should equal 1. The log-likelihood function associated with Equation (15.7) is known to be:

$$\begin{aligned} &-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln[\sigma^2/(1-\rho)^2] - \frac{[z_1 - \mu/(1-\rho)]^2}{2\sigma^2/(1-\rho)^2} \\ &-\frac{(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln(\sigma^2) - \left[z_{t-1} - \sum_{t=2}^{T}\frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}\right] \end{aligned} \tag{15.8}$$

The likelihood ratio test statistic for the null hypothesis is then:

$$LR = -2[L(0,1,0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})] \tag{15.9}$$

where $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\rho}$ are maximum likelihood estimates of the parameters concerned. The LR statistic is distributed under the null hypothesis as a $\chi^2(3)$, a chi-squared with 3 degrees of freedom. We can therefore test the null hypothesis against this alternative hypothesis by

[8] However, neither of these tests addresses any 'iid' prediction, where such predictions arise. Such predictions need to be tested using other methods, such as those discussed in Box 15.3 or in section 15.3.2.

obtaining maximum likelihood estimates of the parameters, deriving the value of the LR statistic, and comparing this value against the critical value for a $\chi^2(3)$. This is a powerful test because the alternative hypothesis is quite a general one.

However, this testing procedure has a weakness: it does not address the normality of z_t (or ε_t) as such. It merely tests the rather less demanding null hypothesis that $\rho = \mu = 0$ and $\sigma = 1$. This is a problem because we could easily have a situation where $\sigma = 1$, but ε_t is *not* $N(0,1)$, and this might be the case if we had a model that generated forecasts that are skewed or that exhibited insufficient kurtosis, both of which are common problems with risk measurement models. In such cases, this procedure would be likely to miss the skewness or excess kurtosis because it tests for $\sigma = 1$ rather than for $\varepsilon_t \sim N(0,1)$. In other words, the Berkowitz LR test has a blind spot when it comes to skewness or tail heaviness.[9] Fortunately, this problem is easily dealt with by supplementing the LR test with an additional test for the normality of ε_t (e.g., such as a Jarque–Bera test): the latter test ought then to detect any non-normality in ε_t that the LR test might overlook.

There are also other ways in which we might test our z_t series for standard normality. For example, we might separate out our key predictions and test each in turn:

- *Mean prediction*: We can test the prediction that the z_t have a zero mean using a textbook *t*-test.
- *Variance prediction*: We can test the prediction that the z_t have a unit variance using a variance-ratio test. If s^2 is a sample estimate of the variance of a normally distributed random variable, and σ^2 is a hypothetical 'true' value (which is here equal to 1), then $(n-1)s^2/\sigma^2$ is distributed under the null as a chi-squared with $n-1$ degrees of freedom.
- *Skewness prediction*: We can test the skewness prediction by using an asymptotic normality test: given that the underlying distribution is normal, then as n gets large, the sample skewness is distributed as approximately $N(0,6n^{-1})$.
- *Kurtosis prediction*: We can also test the kurtosis prediction in a similar way. With an underlying normal distribution, then as n gets large, the sample excess kurtosis is distributed as approximately $N(0,24n^{-1})$.
- *Skewness and kurtosis predictions*: We can also test these predictions simultaneously, using a JB test, as described in the appendix to this chapter. This combined test can also be regarded as an omnibus test of normality itself.

These moments-based tests are particularly useful, because if a model fails one of them, the test result indicates the nature of the model inadequacy, and this is very useful for diagnostic purposes. If the model fails the mean test, then we conclude that the model's forecasts are biased in a particular direction; if the model fails the variance test, we conclude that the model's predicted dispersion is too high (if the sample variance is less than 1) or too low (if the sample variance exceeds 1). If the model fails the skewness text, we conclude its forecasts are skewed; and if the model fails the kurtosis tests, we conclude that its forecasts have excess or insufficient kurtosis, depending on whether the sample kurtosis is bigger or less than 3.

Of course, we also test any independence predictions, where these arise, and can do so using any of the methods mentioned in Box 15.3.

None of these tests requires any 'difficult' calculations, or even regressions, which makes them very easy to implement (e.g., on a spreadsheet).

[9] A plausible example and a longer discussion of this problem with the Berkowitz backtest is given in Dowd (2004a), p. 86.

Example 15.5 (Variance-ratio test)

Suppose we have a sample of 500 observations drawn from an assumed normal distribution. The sample variance is s^2, and under the null the true variance σ^2 equals 1. We know that the test statistic $(n-1)s^2/\sigma^2 = (n-1)s^2$ is distributed under the null as a χ^2 with $n-1$ degrees of freedom. Given a two-sided alternative hypothesis and a 5% significance level, we therefore accept the null if $499s^2$ lies between the 2.5% and 97.5% quantiles of a χ^2 with 499 degrees of freedom, and these are 438.9980 and 562.7895. It follows, then, that we accept the null if s^2 lies between 438.9980/499 = 0.8798 and 562.7895/499 = 1.1278.

Example 15.6 (JB test)

Suppose we have a sample of 500 observations drawn from an assumed normal distribution. The sample skewness and kurtosis are −0.0758 and 2.8888. Using the formula from Equation (A15.4) in the appendix, the JB test statistic is

$$JB = \frac{500}{6}\left(0.0758^2 + \frac{(3-2.888)^2}{4}\right) = 0.7364$$

This test value corresponds to a cdf-value of 0.3080 for a χ^2 with 2 degrees of freedom. This test result confirms that we can accept that the underlying distribution is indeed normal.

15.3.3 Overlapping Forecast Periods

We have mentioned several times that some risk models do, and others do not, predict that outcomes – whether exceedances, Rosenblatt-transformed data p_t, or Berkowitz-transformed data z_t – should be iid.

To explain further, outcomes are predicted to be independent when forecast periods do not overlap. The archetypal case is where a risk manager forecasts a daily VaR before trading starts; the realised P/L is taken from the end-of-day value of the portfolio, and can be compared to that day's VaR forecast. The cycle then repeats itself independently the next trading day. Using our earlier notation, p_t and z_t should be iid. In general, if we have some arbitrary forecast horizon – a day, a week, whatever – and if the forecasted variable is realised before the next set of forecasts is produced, then each forecast-realisation cycle finishes before the next one starts, and the model predicts iid.

Now consider a case where a risk model predicts P/L over a two-day horizon, say, *and* it makes such forecasts at the beginning of each trading day. Let the realised random P/Ls for each trading day be X_1, X_2, X_3, etc. It follows that p_1 and z_1 will depend on X_1 and X_2, p_2 and z_2 will depend on X_2 and X_3, and so on. Both p_1 and p_2 (or z_1 and z_2) are affected by X_2, so they cannot be expected to be independent. If we had a three-day horizon, each consecutive pair of outcomes would be affected by two realised X values, if we had a four-day horizon, each consecutive outcome pair would be affected by three realised X values, and so on. There is consequently no iid prediction when horizon periods (or forecast-realisation cycles) overlap.

The presence of overlapping forecast horizons affects backtesting in various ways. As just mentioned, it means that there is no iid prediction to test, but the other predictions – that p_t is U(0,1), or z_t is N(0,1) – remain, so we are testing for standard uniformity or standard normality in the absence of iid. Implementing such tests is more subtle than in the iid case. Perhaps the best way to carry out the tests is to postulate that the dependence structure takes a particular form – although what would be considered plausible would depend on the problem at hand. An example might be an AR(1) process such as Equation (15.7), although in some cases we might prefer others such as MA processes. We could then choose our test statistic (or statistics), estimate their sample values, and compare these against their critical values as determined by a purpose-built Monte Carlo simulation routine. However, unless we can predict the dependence structure – which is sometimes possible, but usually not – it would be wise to postulate a variety of alternative dependence structures, and check that our test results are robust across these alternatives. [10]

15.4 COMPARING ALTERNATIVE MODELS

It is often the case that we are not just interested in how individual models perform, but also in how different models compare to each other.[11] We can do so using forecast evaluation methods that give each model a score in terms of some loss function; we then use the loss scores to rank the models – the lower the loss, the better the model. These approaches are not statistical tests of model adequacy. Instead, their purpose is to rank models.[12] Because they are not statistical tests, forecast evaluation approaches do not suffer from the low power of standard tests such as basic frequency tests: this makes them attractive for backtesting with the small data sets typically available in real-world applications. In addition, they also allow us to tailor the loss function to take account of particular concerns: for example, we might be more concerned about higher losses than lower losses, and might therefore wish to give higher losses a greater weight in our loss function.

The ranking process has four key ingredients, and a single output, a final score for each model.

The first ingredient is a set of n paired observations – paired observations of losses (or P/L) each period and their associated VaR forecasts.

The second ingredient is a loss function that gives each observation a score depending on how the observed loss (or profit) compares to the VaR forecasted for that period. Thus, if L_t is

[10] An alternative approach is to put our data through yet another transformation. If we assume that the zt are AR(1) N(0, 1) this implies the ϵ_t have a variance of $(1 - p^2)$. We then use an estimate of p, $\hat{p}$, to recover an estimated residual series $\hat{\epsilon}_t = \hat{z}_t - \hat{p}z_{t-1}$, and test the prediction that ϵ_t are iid N(0, $1 - \hat{p}^2$).

[11] We can also rank models by means of their performance measure against selected standard statistics. We can either take a group of models and compare their risk measures to the average results for the group as a whole (which would give us a feel for which models produce higher or more volatile risk measures), or else we can compare them to those predicted by each individual model (which would give us a feel for the models' comparative accuracy). Such procedures are discussed at length by Hendricks (1996), who lists nine possible comparator statistics, including mean relative bias, root mean squared relative bias, empirical tail-loss frequency, ratio of average loss to VaR, and ratio of maximum loss to VaR.

[12] This raises a natural question. Why not rank models by means of prob-values obtained from statistical tests? The answer is that it would be nice if we could, but we can't (or at least we shouldn't). Prob-values are designed to evaluate hypotheses, and different prob-values are in general not directly comparable to each other, because the alternative hypotheses are usually non-nested. Adapting statistical tests to enable models to be compared is therefore a subtle business. For more on how this might be done, see Christoffersen *et al*. (2001) and Giacomini and Komunjer (2002). The latter paper also addresses how forecasts might be combined to improve their forecasting performance. However, even these approaches only allow us to make pairwise comparisons of models, rather than a full ranking, and there is no guarantee that they will produce mutually consistent sets of pairwise comparisons: for example, after going through all these tests we might conclude that model *A* is better than model *B*, model *B* is better than model *C*, and model *C* is better than *A*, in which case we get a set of pairwise comparisons but no overall ranking

the loss made over period t, and VaR_t is the forecasted VaR for that period, our loss function assigns the following value to the period-t observation:

$$C_t = \begin{cases} f(L_t, VaR_t) & \text{if} \quad L_t > VaR_t \\ g(L_t, VaR_t) & \quad\quad\; L_t \leq VaR_t \end{cases} \tag{15.10}$$

where $f(L_t, VaR_t) \geq g(L_t, VaR_t)$ to ensure that tail losses do not receive a lower value than other observations.[13]

The third ingredient is a benchmark, which gives us an idea of the score we could expect from a 'good' model.

The fourth ingredient is a score function, which takes as its inputs our loss function and benchmark values. For example, if we take our benchmark to be the expected value of C_t under the null hypothesis that the model is 'good', then Lopez (1998) suggests that we might use a quadratic probability score (QPS) function, given by:

$$QPS = \frac{2}{n}\sum_{t=1}^{n}(C_t - p)^2 \tag{15.11}$$

The QPS takes a value in the range [0,2], and the closer the QPS-value to zero, the better the model. We can therefore use the QPS (or some similar score function) to rank our models, with the better models having the lower scores.[14]

To implement this type of procedure, we need to specify the loss function, and a number of different loss functions have been proposed. Perhaps the most straightforward is the binary loss function proposed by Lopez (1998, p. 121), which gives an observation a value of 1 if it involves a tail loss, and a value of 0 otherwise. Equation (15.10) then takes the form:

$$C_t = \begin{cases} 1 & \text{if} \quad L_t > VaR_t \\ 0 & \quad\quad\; L_t \leq VaR_t \end{cases} \tag{15.12}$$

This 'Lopez I' loss function is intended for the user who is (exclusively) concerned with the frequency of tail losses. The benchmark for this loss function is p, the expected value of $E(C_t)$.

Example 15.7 (Lopez I backtest)

Suppose we have the same parameters as in Example 15.1: $n = 500$, $p = 0.05$ and $x = 30$. Applying the Lopez I backtest, we get a QPS value of $2/500 \times [30(1 - 0.05)^2 + 470(0.05)^2] = 0.1130$. Had we had an x-value equal to our expected value of $x = np = 25$, on the other hand, our QPS value would have been $2/500 \times [25(1 - 0.05)^2 + 475(0.05)^2] = 0.0950$. The Lopez I backtest therefore tells us that our model performs worse than a model that generated the expected number of exceedances. However, it cannot tell us whether our model's performance is *significantly* worse or not.

---

[13] If we wish it to, the loss function can incorporate asymmetries in the backtester's concerns about loss outcomes (e.g., it might treat losses in excess of VaR differently from losses below VaR). This is very useful because it allows us to take account of the economic as well as statistical consequences of high losses.

[14] The QPS criterion also has the attractive property that it (usually) encourages truth-telling by VaR modellers: if VaR modellers wish to minimise their QPS score, they will (usually) report their VaRs 'truthfully'. This is a useful property in situations where the backtester and the VaR modeller are different, and where the backtester might be concerned about the VaR modeller reporting false VaR forecasts to alter the results of the backtest.

A drawback of this Lopez loss function is that it ignores the magnitude of tail losses. If we wish to remedy this defect, Lopez himself suggests a second, size-adjusted, loss function:

$$C_t = \begin{cases} 1 + (L_t - VaR_t)^2 & \text{if} \quad L_t > VaR \\ 0 & \quad\quad\;\; L_t \leq VaR_t \end{cases} \tag{15.13}$$

This loss function allows for the sizes of tail losses in a way that Equation (15.12) does not: a model that generates higher tail losses would generate higher values of Equation (15.13) than one that generates lower tail losses, other things being equal. However, with this loss function, there is no longer a straightforward condition for the benchmark, so we need to estimate the benchmark by some other means (e.g., Monte Carlo simulation).[15] In addition, the size-adjusted loss function (Equation (15.13)) also has the drawback that it loses some of its intuition, because squared monetary returns have no ready monetary interpretation.

A way round this problem is suggested by Blanco and Ihle (1998), who suggest the following loss function:

$$C_t = \begin{cases} (L_t - VaR_t)/VaR_t & \text{if} \quad L_t > VaR_t \\ 0 & \quad\quad\;\; L_t \leq VaR_t \end{cases} \tag{15.14}$$

This loss function gives each tail-loss observation a weight equal to the tail loss divided by the VaR. This has a nice intuition, and ensures that higher tail losses get awarded higher C_t values without the impaired intuition introduced by squaring the tail loss. The benchmark for this forecast evaluation procedure is also easy to derive: the benchmark is equal to the difference between the ES and the VaR, divided by the VaR. Yet the Blanco–Ihle loss function also has a problem: because Equation (15.14) has the VaR as its denominator, it is not defined if the VaR is zero, and can give mischievous answers if VaR gets 'close' to zero or becomes negative. It can therefore be unreliable unless we can be confident of the VaR being sufficiently large and positive.[16]

We therefore seek a size-based loss function that avoids the squared term in the Lopez-II loss function, but also avoids denominators that might be zero valued. A promising candidate is the tail loss itself:

$$C_t = \begin{cases} L_t & \text{if} \quad L_t > VaR_t \\ 0 & \quad\quad\;\; L_t \leq VaR_t \end{cases} \tag{15.15}$$

The expected value of the tail loss is of course the ES, so we can choose the ES as our benchmark and use a quadratic score function such as:

$$QS = \frac{2}{n} \sum_{t=1}^{n} (C_t - ES_t)^2 \tag{15.16}$$

This approach penalises deviations of tail losses from their expected value, which makes intuitive sense. Moreover, because it is quadratic, it gives very high tail losses much greater weight than more 'normal' tail losses, and therefore comes down hard on large losses.

[15] One way to do so is suggested by Lopez (1998, pp. 123–124). He suggests that we assume that observed returns are independent and identically distributed (iid); we can then use this assumption to derive an empirical loss function and a value of the final score; if we repeat the operation a large number of times, we can use the average final score as our estimate of the benchmark.

[16] Blanco and Ihle also suggest a second approach that incorporates concerns about both the frequency and the size of tail losses. If we let $C_t^{frequency}$ be the Lopez-I frequency-loss-function (Equation (15.12)), and C_t^{size} be the Blanco–Ihle size-loss-function (Equation (15.14)), they suggest an alternative loss function that is a weighted average of both, with the weighing factor reflecting our relative concern about the two sources of loss. This is an appealing idea, but this suggestion does not produce reliable rankings: $C_t^{frequency}$ and C_t^{size} are not defined in terms of the same underlying metric, so irrelevant changes (e.g., like redefining our monetary units: say, going from dollars to cents) can alter our scores, and so change our rankings. The idea of taking a weighted average is a good one, but we need a more reliable way of implementing it.

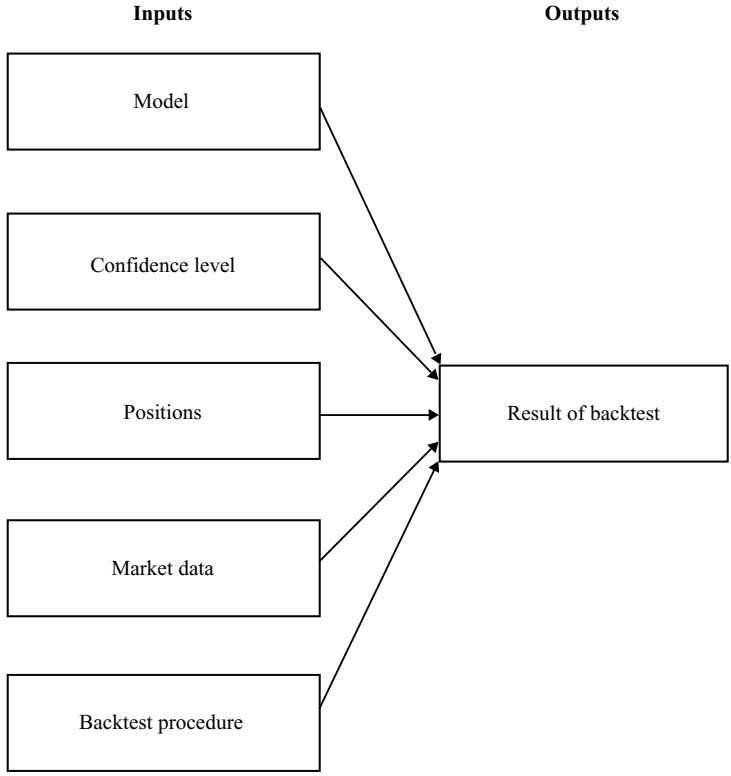

Figure 15.4 A general backtesting framework

15.5 BACKTESTING WITH ALTERNATIVE POSITIONS AND DATA

We can also get more information about the adequacy or otherwise of our risk model by backtesting with alternative inputs. Recall that the objective of any backtest procedure is to evaluate a given market risk model. However, as illustrated in the general backtesting framework outlined in Figure 15.4, in order to be able to do so, we also have to specify a particular VaR confidence level, particular positions in market instruments, and a particular set of market price/return data. Yet it is the model that we are really interested in and the confidence level, positions and data are only supplementary, albeit necessary, inputs. Once we have the model, the confidence level, the positions and the market data, the only other input is the particular backtest procedure used – the Kupiec test, Kolmogorov-Smirnov, or whatever. These then produce the output, which is a quantitative indication of the adequacy of the model.

So far, we have examined alternative backtests and taken the other inputs as given. However, we can also evaluate a model by varying the other inputs too: in particular, we can evaluate the model using different positions and different data.[17] This opens up new backtest possibilities.[18]

[17] We could also backtest using alternative confidence levels as well. However, we need to be careful, as a good performance at one confidence level does not necessarily mean that a model is good at other confidence levels (e.g., an EV model should be good at very high confidence levels, but there is no reason to expect it to be any good at lower confidence levels).

[18] We can also open up a vast array of additional backtesting possibilities by tuning into the large applied statistical literature on the properties of measures of distributional conformity as functions of risk factors. This literature uses methods based on a great variety

15.5.1 Backtesting with Alternative Positions

One route is to backtest our model using alternative positions. These might include:

- The actual portfolio we are currently holding.
- Each of the different portfolios we held over some past time period.
- Position at business-unit level, as well as at the aggregate, institutional, level: we can backtest our equity positions separately from our fixed-interest positions, and so on. We can also backtest at any business-unit level we want, right down to the level of the individual trader or asset manager, or the individual instrument held.
- We can carry out benchmark backtesting: we select a set of important or representative positions, which we regard as benchmarks, and backtest our model on these benchmarks. These benchmarks might be a typical stock market position, a typical FX options position, and so on. Such benchmark backtesting can be very useful for comparison purposes.
- We can carry out simulation exercises in which we randomly select a large number of portfolios, backtest against each, and then focus on average results (as in Hendricks (1996, pp. 44–45)). This approach has the advantage of giving us some protection against results that might be dependent on a particular portfolio.

15.5.2 Backtesting with Alternative Data

We can also carry out backtests using alternative sets of market-price data. In particular:

- We can vary the historical time period used in our backtesting procedures, and we would typically want to do so to check that our results were not specific to a particular choice of time period.
- For any given time period, we can bootstrap the data and work with bootstrapped data instead: the bootstrap then gives us the benefit of a potentially much larger data set.
- We can produce simulated data using Monte Carlo or other simulation methods, and parameterise such data on existing market-price data or our beliefs about the values of the parameters concerned.

Thus, there are many ways in which we can multiply our data, and so increase our scope for backtesting.

15.6 ASSESSING THE PRECISION OF BACKTEST RESULTS

One problem with any of these backtest procedures is that they place a great deal of weight on estimates, whether of rankings or of prob-values. The 'true' values will always remain unknown. This raises an obvious problem: we might get a poor estimate that leads us to make an incorrect judgement of the model and incorrectly reject a 'true' model or incorrectly accept a 'false' one, or incorrectly rank two different models. So how can we assess the precision of estimates and get a clearer idea of what the true prob-value or true ranking might be?

If we are dealing with statistical backtests, one solution is to estimate a confidence interval for the 'true' tail-loss probability using a bootstrap procedure. To illustrate with a plausible

of different assumptions, data structures, and technical tools. The models involved include, among others, time series/cross-section models, generalised mixed models, vector autoregressive moving average exogenous variable (VARMAX) models, Kalman filter models, and neural networks. For more on these, see Tilman and Brusilovskiy (2001).

example based on simulated numbers, suppose we have a portfolio that generates a daily P/L and we have a VaR model that assumes that P/L is normally distributed. Taking the confidence level to be 99%, the probability p of an excess loss is therefore 0.01 under the null hypothesis that the model is 'good'. Suppose too that we have a hypothetical sample of $n = 500$ daily P/L observations, and this 'historical' sample is drawn from a distribution with heavier than normal tails. Hence, our model is 'bad' because it incorrectly assumes normality. Our model tells us that we should expect $np = 5$ excess losses in our sample, but we actually get 8, and the basic frequency test then tells us that the probability of 8 or more excess losses is 6.71%. If we take this to be an estimate of the null probability and apply a standard significance level such as 5%, we would be inclined to 'pass' the model as acceptable, even if marginally so – and this conclusion would be incorrect.

Now suppose that we avoid the temptation to 'pass' the model on the basis of this result and bootstrap prob-values instead. We take 1000 sets of bootstrapped P/L samples from our original 'historical' P/L data. For each such sample, we estimate the VaR under the maintained hypothesis of normality and calculate the number of excess losses x in the sample. Our sample of 1000 x-values then generates a corresponding sample of null-probability estimates, and this latter sample gives us an estimate of the null-probability confidence interval. Such an exercise produces bootstrapped x-values with a mean of 7.93 and a 95% confidence interval of [3,14]. The corresponding bootstrapped null-probability estimates have a sample mean of 13.7% and an estimated 95% confidence interval of [0.0%, 44.0%]. This indicates that our earlier prob-value estimate is very imprecise.

This confidence interval also tells us that we cannot be confident at the 95% level that the 'true' null probability is either *at least* 5% (in which case we would confidently accept the model) or *less than* 5% (in which case we would confidently reject it). This confidence-interval information should then lead us to modify our assessment of the model: instead of (incorrectly) concluding that the model is 'true', as we would if we used the 'raw' frequency test result, we should only conclude that we can be 95% confident that the 'true' null probability lies between 0% and 44%. Since this range includes our significance level of 5%, this means that we *cannot* come to a clear (i.e., pass/fail) conclusion about the validity of our model. However, this more modest conclusion does justice to the uncertainty in our data, and helps prevent a false assessment of our model.

This illustrates that we cannot always arrive at binary conclusions. Instead of concluding that a model either 'passes' a backtest or 'fails' it, the correct conclusion is sometimes that we cannot be confident whether the model passes or not. However, if that is what the data tell us, then so be it: it is better to conclude that the evidence does not allow us to come to a clear-cut judgement on the model one way or the other, and be right, than to come to clear conclusions and be wrong.

We can also use simulation methods to evaluate our backtests in other ways. For example, we can 'backtest our backtests' to see how reliable our backtests seem to be. To do so, we establish a range of plausible alternative hypotheses: if we were concerned about tails being heavier than assumed in our model, we might assume a range of alternative heavier tailed possibilities, and so on. The null hypothesis and our alternatives are chosen taking account of the context. We then carry out Monte Carlo simulations to evaluate the reliability of the various backtests we are using: we estimate their type I and type II error rates, and so on. The results of these exercises would give us an indication of the reliability of the different backtests in the context we are working in. We might then find that some respectable backtests seem to perform badly in our particular situation, whereas others perform much better. This would

give us some idea of how well the different backtests seem to be working, and we can then choose the particular backtests that work best for us.[19]

15.7 SUMMARY AND CONCLUSIONS

The main points covered in this chapter can be summarised as follows:

- The first requirement in backtesting is to clean our data and carry out a preliminary data analysis. This analysis should include the use of a backtesting chart and some summary statistical analysis.
- The most straightforward tests are those based on a comparison of the actual and predicted frequencies of exceedances – the basic binomial (or Kupiec) test, Christoffersen tests, etc. However, these often lack power with the sample sizes commonly available to us, and they throw away potentially valuable information about tail losses.
- It is therefore better to use tests that make use of the sizes as well as frequencies of exceedances, or to use outcomes across the whole P/L spectrum or at least along the spectrum of tail losses. To do so, we would use the Rosenblatt and Berkowitz transformations to make our data comparable and malleable, and then apply distribution-equality tests to evaluate the predictions of the underlying null hypothesis. Such tests have much more power than frequency-based tests, and there are a variety of such tests available.
- We can also compare alternative models by ranking their performance in terms of loss-score forecast evaluation methods (as in Lopez, etc.).
- We can also carry out additional backtests by changing other inputs besides the actual backtest procedure itself: in particular, we can change the positions and market data used. Additional backtests can tell us a lot about model adequacy that would not otherwise be apparent.
- Sometimes the data do not admit to clear results about model adequacy (i.e., sometimes the correct result is not 'guilty' or 'innocent', but 'not proven'). We should therefore bootstrap our backtests to get a better idea of the distribution of prob-value estimates. We might also use simulation methods to 'backtest our backtests' and get an idea of how well particular tests work in the particular contexts in which we wish to use them.

All this said, we should not rely on any one backtest, and if a particular procedure gives a strong result, either positive or negative, we should seek to confirm or disconfirm that result using other backtests: strong conclusions require strong evidence. In any case, it is good practice to use a battery of whatever backtests are feasible, and to use their results to get a feel for the strengths and weaknesses of our risk model. In the final analysis, the purpose of backtesting is not just to provide an overall assessment of a risk model, but to develop a sense of how the model works and to learn its strengths and limitations.

[19] Finally, we can also gauge the precision of our backtest results by stress testing them. We might change the input data in plausible ways, or change the calibration, and see what effect these changes have on our results. If the results are robust to these changes, then we can have more confidence in them; if they are not, we need to be more careful with them.

Appendix
Testing Whether two Distributions are Different

This appendix addresses the principles involved when we seek to test whether one distribution is different from another. This problem is the central issue in backtesting, and arises whenever we wish to test whether an empirical distribution is different from a predicted distribution. There are two broad approaches we can take to this problem:

- We can focus on a particular parameter (such as the mean, median, variance, etc.) and then test the hypothesis that this parameter has the same value in both distributions.
- We can test whether one distribution considered as a whole is indistinguishable from the other distribution as a whole.

To distinguish between them, we shall refer to the former for convenience as 'parameter-equality' tests and the latter as 'distribution-equality' tests.

A15.1 PARAMETER-EQUALITY TESTS

In the risk management area, parameter-equality tests are often applied to the frequency of exceedances, and we discussed frequency-based tests at some length in Chapter 15, section 15.2. However, we can also apply parameter-equality tests to other parameters, including the following:

- *Mean*: Assuming other factors are suitably 'well behaved', we can test for the equality of two means using textbook tests such as t-tests and F-tests. For example, if we use a t-test to test the hypothesis that the 'true' mean is μ, then under the null our test statistic $(m - \mu)/(s\sqrt{n})$ is distributed as a Student-t with $n - 1$ degrees of freedom, where m and s are the sample mean and standard deviation.
- *Median*: We can test for the equality of two medians using tests such as a binomial sign test (based on the idea that the sample proportion above and below the median should be about a half), a Wilcoxon signed ranks test (based on the idea that the sum of ranks above and below the median should be similar), a Van der Waerden test (based on ranks that are transformed to normal variates), or a Kruskal–Wallis test.
- *Variance*: If the distribution underlying the null hypothesis is a normal one, we can test whether the empirical and predicted variances are equal using a variance-ratio test. This test is based on the prediction that $(n - 1)s^2/\sigma^2$ is distributed as a χ^2 with $n - 1$ degrees of freedom under the null hypothesis. Alternatively, depending on what other assumptions we might make, we can apply an F-test, a Siegel–Tukey test, a Bartlett test, a Levene test, a Brown–Forsythe test, and so on.
- *Skewness and kurtosis*: If the underlying null distribution is normal, we can also test whether empirical skewnesses and kurtoses are consistent with the null using χ^2 tests.

Further explanations of these tests can be found in textbooks on non-parametric statistics.

Some of these tests were discussed in Chapter 15, where we discussed how they might be used to test the prediction that the z_t series (i.e., the Rosenblatt/Berkowitz-transformed data) are distributed as $N(0,1)$. However, all parameter-equality tests have the limitation that they ignore potentially useful information contained in other parameters. Some of these tests are also dependent on ancillary decisions or assumptions that may not always be appropriate (e.g., frequency tests are conditional on an arbitrary 'target' frequency, the variance-ratio test assumes normality, most textbook tests assume independence, etc.).

There are also runs tests, in which we test whether the number of runs in a time series is consistent with what we might expect under a null hypothesis of independence. These are very easy to apply. Suppose our data are time ordered and expressed in binary form (e.g., we might give exceedance observations a value of 1 and non-exceedances a value of 0). A run is then a sequence of consecutive identical numbers, and the number of runs R is equal to the number of sign changes plus 1. If m is the number of observations taking one value and n the number taking the other value, then under the independence null the mean and variance of the number of runs are respectively:

$$\mu_R = 1 + \frac{2mn}{m+n} \tag{A15.1a}$$

$$\sigma_R^2 = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)} \tag{A15.1b}$$

If the total number of observations is large, then R is approximately normal, and we can test the null using a normal approximation. For example, if we have a two-sided alternative hypothesis, we might carry out the test using the 90% confidence interval for R under the null, which is:

$$[\hat{\mu}_R - 1.645\hat{\sigma}_R, \hat{\mu}_R + 1.645\hat{\sigma}_R] \tag{A15.2}$$

where $\hat{\mu}_R$ and $\hat{\sigma}_R$ are sample estimators of μ_R and σ_R.

A15.2 DISTRIBUTION-EQUALITY TESTS

'Distribution-equality' (or 'whole distribution') approaches are often much better, because they make use of more information than the parameter-equality approaches. There are a number of such approaches to choose from.[20]

A15.2.1 The Chi-squared Test

One of the classic distribution-equality tests is the χ^2 (or chi-squared) test. This can be used to test the significance of the difference between any empirical and predicted distribution functions. The chi-squared test is applied to binned (or classified) data. Hence, if the data come from a continuous distribution, we would need to bin the data, which involves dividing the data into k classes. We then compute the test statistic:

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{A15.3}$$

[20] This discussion of distribution-equality tests is illustrative rather than exhaustive, and also omits certain tests because they are fairly close relatives of the some of the tests considered. The tests omitted include, for example, the Cramér–von Mises test, which is a relative of the Kolmogorov–Smirnov and Anderson–Darling tests, and the Shapiro–Francia test, which is a close relative of the Shapiro–Wilks test.

where O_i is the observed frequency of data in bin i, and E_i is the expected frequency of data in bin i.[21] Under the null hypothesis, this test statistic is distributed as χ^2_{k-c}, where c is the number of estimated parameters (location, scale, etc.) for the assumed distribution. We therefore reject the null hypothesis if the test statistic exceeds the critical value associated with our chosen significance level.

The main disadvantage of the χ^2 test is that results are dependent on the way in which the data are binned, and binning is (largely) arbitrary. (Any optimal bin size, if there is one, would depend on the distribution.) In using it, we should therefore be careful to check the sensitivity of results to alternative ways of binning the data. A second drawback is that the test is only approximate, and the approximation only works for reasonably large sample sizes.[22]

A15.2.2 The Jarque–Bera Test

Where our assumed density is normal, we can also use the Jarque–Bera (JB) test, which is widely used in econometrics. The JB test statistic is:

$$JB = \frac{n}{6}\left(s^2 + \frac{(\kappa - 3)^2}{4}\right) \tag{A15.4}$$

where n is the sample size and s and κ are the coefficients of sample skewness and kurtosis. Under the null hypothesis, the JB test statistic is asymptotically distributed as χ^2 with 2 degrees of freedom. The JB test statistic is therefore easy to calculate and its critical values are easy to obtain. The JB test is very useful in risk management, because we can use it to test for heavy tails (e.g., in return distributions). In addition, as explained in Chapter 15, it is useful for backtesting because we can use it to test that Rosenblatt–Berkowitz transformed series, z_t, are normal.

A15.2.3 The Kolmogorov–Smirnov Test

The Kolmogorov–Smirnov (KS) test is one of the oldest and best-known distribution tests. It focuses on the difference between predicted and empirical cumulative density functions, and is applied to data coming from a continuous density function (i.e., so does not involve binning). To apply this test, we first construct the predicted cdf $F(x)$ and empirical cdf $\hat{F}(x)$.[23] The KS test statistic is then the maximum distance between the two cdfs evaluated over each data point X_i:

$$D = \max_i \left|F(X_i) - \hat{F}(X_i)\right| \tag{A15.5}$$

The test value of the KS statistic is then compared to the relevant critical value, and the null is accepted or rejected accordingly.

To illustrate, Figure A15.1 plots cumulative and empirical cdfs drawn from a normal distribution. In this case, the null (that the data are drawn from an assumed normal) is true and

[21] The latter can be calculated as the number of observations in our data set times the difference between the value of the distribution function at the right hand of the bin and the value of the distribution function at the left hand of the bin.

[22] Chi-squared tests are also capable of considerable refinement. For example, Anderson (1994) shows how they can be directed at a variety of alternative hypotheses and concerns (e.g., about skewness, kurtosis, etc.) and still retain considerable power.

[23] The former comes from the assumed distribution; to construct the latter, we order the data from smallest to largest, and the value of the CDF associated with the ith observation is then the number of observations less than or equal to this observation divided by the total sample size n.

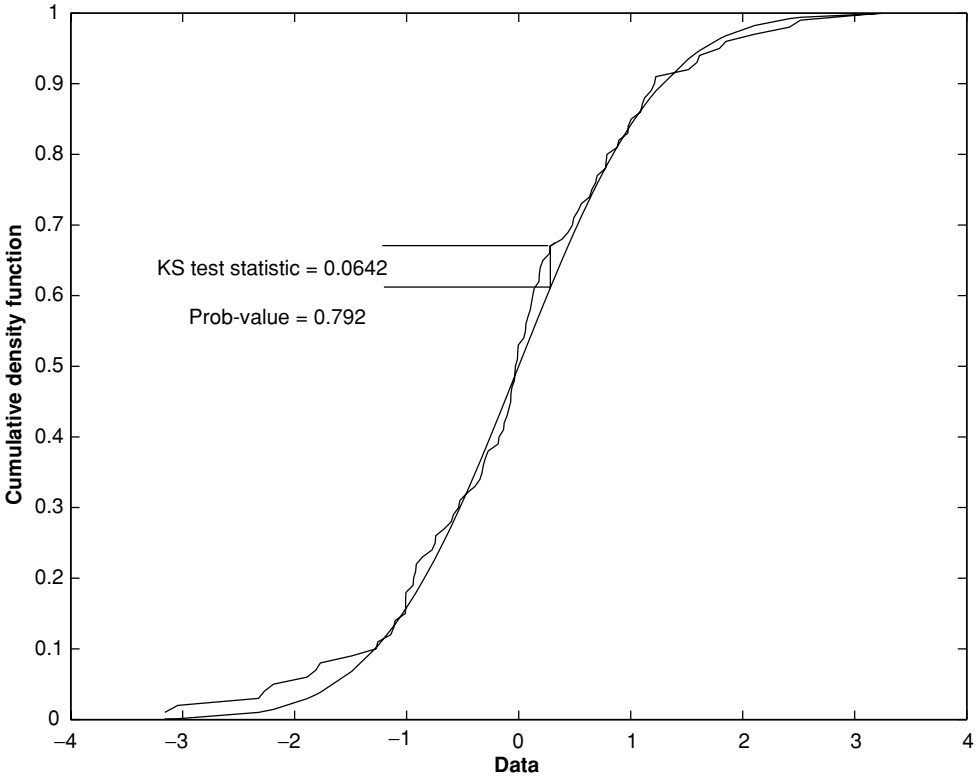

Figure A15.1 The Kolmogorov–Smirnov test
Note: Empirical cdf (the rough curve) is calculated from 100 random drawings from a standard normal, and the predicted cdf is that of a normal distribution.

the two curves are close together. The KS test yields a prob-value of 0.792, and indicates (correctly) that the null can be accepted.

This test is easy to implement, because the test statistic is straightforward to calculate and its critical values are easily obtained using Monte Carlo simulation. A coding strategy to obtain a confidence interval for the KS statistic is illustrated in the MMR Toolbox function 'ks_test_stat':

- Assume a theoretical distribution, which can be the standard normal.
- Take a large number m of random samples of size n from this distribution.
- For each sample we obtain empirical and predicted cdf series, and take the sample KS test statistic as the largest absolute value of the difference between empirical and predicted cdf series.
- Sort the sample KS test statistic values smallest to highest, and obtain the cut-off points of the lower and upper tails of the distribution of KS test values. For example, if we want a 90% confidence interval, we select the cut-offs of the lowest and highest 5% of the distribution. These cut-off points give us the confidence interval for the KS test statistic under the null hypothesis.

As an example, Figure A15.2 plots the cdf for the KS test statistic for an assumed sample size of 100, and the cut-offs of the 5% tails tell us that the 90% confidence interval for the KS test is [0.0455,0.1305].

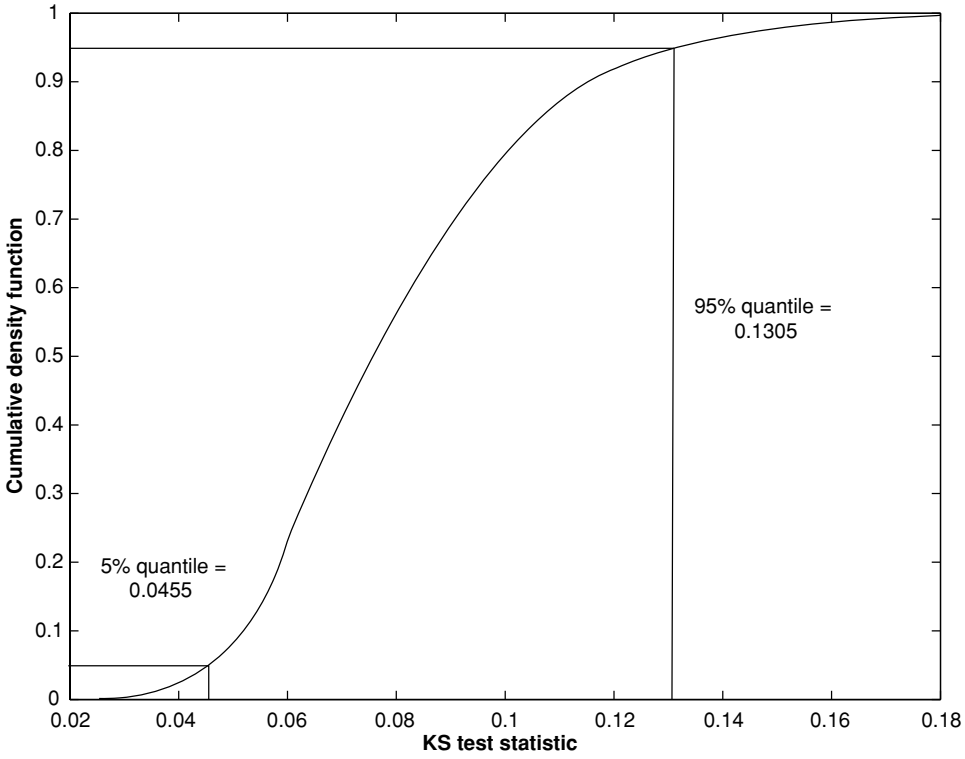

Figure A15.2 Cumulative density function for the KS test
Note: Based on 10 000 simulation trials using the MMR Toolbox function 'KS_test_stat'. The sample size $n = 100$.

Having obtained a function such as the one used here, all we need do in any practical situation is to plug the sample size into it and the function will give us the confidence interval for the specified sample size.

The KS test has the advantages that it does not depend on binning, is exact, and is distribution-free in the sense that its critical values do not depend on the assumed distribution $F(x)$. However, its disadvantages are:

- It only applies to continuous distributions.
- It tends to be more sensitive to the distributional differences near the centre of the distribution, and is less sensitive at the tails. This is obviously a drawback when using the KS test to backtest risk models, where we are usually much more interested in the tail than in the central mass of a distribution.
- The KS test is only strictly valid if the parameters of the distribution are known, and is not valid if the parameters are estimated. This is a drawback if we are using risk models with estimated rather than known parameters.

A15.2.4 The Lilliefors Test

As just noted, one of the limitations of the KS test is that it is not valid when distribution function parameters need to be estimated. However, if the predicted distribution function is

normal, an alternative that avoids this problem is the Lilliefors test. This applies the KS test not to the raw data, as such, but to their normalised equivalents, normalised by sample mean and standard deviation. In other words, we transform our data X_i by means of the normalisation:

$$Z_i = \frac{X_i - m}{s} \tag{A15.6}$$

where m and s are the sample mean and standard deviation of our data. We then calculate the KS statistic applied to the Z_i. We can obtain critical values (or a confidence interval) for the Lilliefors test statistic using a Monte Carlo routine that is almost identical to the one used for the KS test.

In taking account of the fact that parameters are estimated, the Lilliefors test avoids one of the main limitations of the KS test, but it is only reliable if the predicted distribution is normal and, like the KS test, it is more sensitive to central discrepancies than to tail ones.

A15.2.5 The Kuiper Test

A second close relative of the KS test is the Kuiper test, whose test statistic is the sum of the maximum amount by which each distribution exceeds the other:

$$D^* = \max_i\{F(X_i) - \hat{F}(x_i)\} + \max_i\{\hat{F}(X_i) - F(X_i)\} \tag{A15.7}$$

The Kuiper test can be implemented in much the same way as the KS test: its test statistic is straightforward to calculate and its critical values can easily be obtained by an almost identical Monte Carlo simulation approach to that used for the KS test, the only difference being that the test statistic is Equation (A15.7) instead of Equation (A15.5).

To illustrate the Kuiper test, Figure A15.3 gives the Kuiper equivalent of our earlier KS plot in Figure A15.2. This shows that the cdf for the Kuiper test statistic is similar to that of the KS test statistic, although the test statistic itself is usually a little bigger than the KS test (as we would expect). The 90% confidence interval for the Kuiper test statistic is [0.0731, 0.1627].

The Kuiper test has the advantage over the KS test that it is more sensitive to deviations in the tail regions. It is also believed to be more robust to transformations in the data, and to be good at detecting cyclical and other features in the data. However, there is also evidence that it is data intensive, and needs large data sets to get reliable results.[24]

A15.2.6 The Anderson–Darling Test

Another useful test is the Anderson–Darling test,[25] whose test statistic is defined as

$$\begin{aligned} A^2 &= n\int_{-\infty}^{\infty} \frac{[\hat{F}(x) - F(x)]^2}{F(x)[1 - F(x)]} dF(x) \\ &= -n - \frac{1}{n}\sum_{i=1}^{n}(2i - 1)[\ln F(X_i) + \ln(1 - F(X_{n+1-i})] \end{aligned} \tag{A15.8}$$

The AD test is a modification of the KS test, and can be implemented in much the same way.

[24] See, e.g., Crnkovic and Drachman (1996, p. 140). They also suggest that observations can be weighted by means of a 'worry function' that reflects users' risk-aversion, and that the Kuiper test (and by implication, other tests too) can be applied to the weighted observations to take account of their risk aversion.

[25] For more on this test, see Stevens (1974).

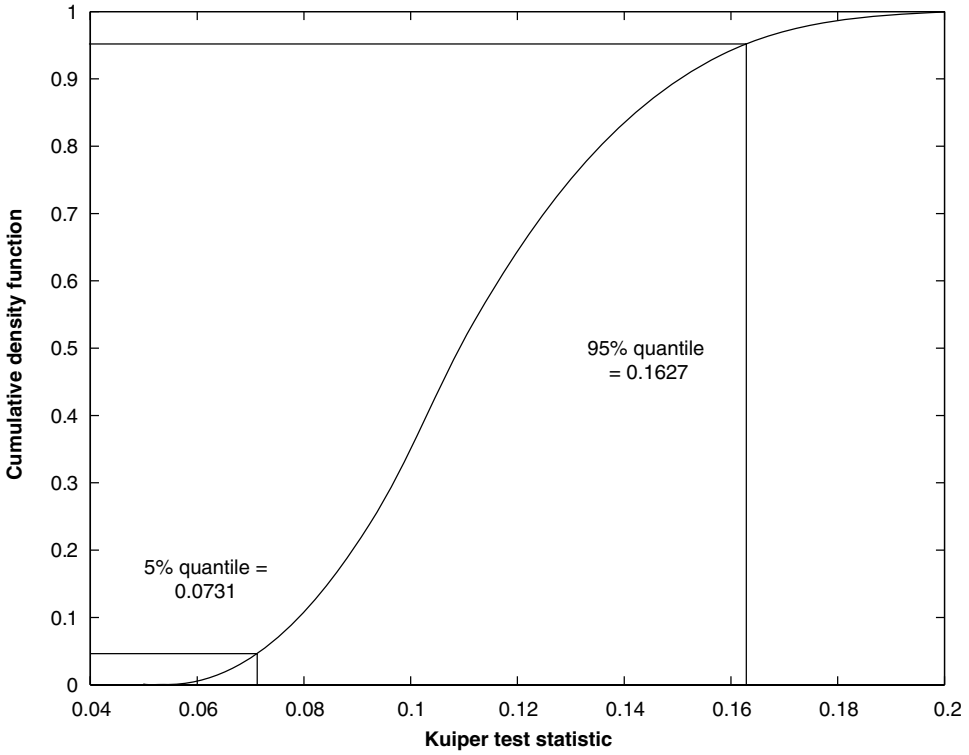

Figure A15.3 Cumulative density function for the Kuiper test
Note: Based on 10 000 simulation trials using the MMR Toolbox function 'Kuiper_test_stat'. The sample size $n = 100$.

The AD works better on the tails than the KS test and accommodates estimated parameters. It is also more sensitive than KS, and is good for any data set and any assumed distribution. However, unlike the KS test, it is not distribution-free, so its critical values depend on the assumed distribution $F(x)$ and therefore have to be obtained for the particular distribution at hand. Adjustments also have to be made when n is small.

A15.2.7 The Shapiro–Wilks Test

A final test is the Shapiro–Wilks test, which is another useful omnibus test of normality. Assuming our data are in ordered form (i.e., are order statistics), the Shapiro–Wilks test statistic W is given by

$$W_i = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \tag{A15.9}$$

where the a_i are constants that depend on the problem at hand. The Shapiro–Wilks test is highly regarded, but one needs to refer to published literature or specialist software to obtain the a_i values and critical values to implement the test.

16
Model Risk

This chapter considers our last subject of model risk. Loosely speaking, model risk is the risk of error in our estimated risk measures due to inadequacies in our risk models. Model risk is a very subtle subject, and the risk of error is an inescapable consequence of model use.[1] Model risk is often hidden or glossed over, and is often overlooked. However, it is also a very important subject, because inadequate attention to model risk can lead to firms being much more vulnerable to high losses than they think they are. Thus, a failure to consider model risk can lead a firm to disaster, and sometimes has.

We begin by reviewing model methodology; we then consider what model risk entails, where it comes from, how to estimate it and, finally, how to deal with it.

16.1 MODELS AND MODEL RISK

Models are formal frameworks that enable us to determine the values of outputs (e.g., such as asset prices, hedge ratios, VaR, etc.) based on postulates about the factors that determine those outputs. There are three main types of model, and the most important are 'fundamental' models, which are formal systems tying outputs to inputs based on assumptions about dynamic processes, interrelationships between variables, and so on. Some examples are the Black–Scholes option pricing model, which links the option price to the underlying price, the strike price, etc., based on assumptions such as a lognormally distributed asset price; and parametric VaR models, which link VaR to assumptions about the distribution of P/L or returns. We also get 'descriptive' models, which can be regarded as short-cuts to fundamental models. These are more superficial, but often more intuitive and easier to work with. An example is a bond price model based on assumptions about yield movements – a model that sidesteps the complexities of the term structure by focusing instead on simplified 'stories' about yields. Both fundamental and descriptive models attempt to explain cause and effect, although to differing levels of depth. Finally, we also get statistical models that attempt to capture the relationship between variables using some sort of statistical best fit, with the emphasis usually on the correlation between them rather than any attempt at a causal explanation as such.

One of the first points to appreciate about a model is that it is only a representation of something, and should never be mistaken for what it represents. In the eloquent words of Emanuel Derman:

> even the finest model is only a model of the phenomena, and not the real thing. A model is just a toy, though occasionally a very good one, in which case people call it a theory. A good scientific toy can't do everything, and shouldn't even try to be totally realistic. It should represent as naturally as possible the most essential variables of the system, and the relationships between them, and allow the investigation of cause and effect. A good toy doesn't reproduce every feature of the real object; instead, it illustrates for its intended audience the qualities of the original object most

[1] For some different perspectives on this subject, see, e.g., Derman (1997), Cairns (2000), Lhabitant (2000) and Kato and Yoshiba (2000).

> important to them. A child's toy train makes noises and flashes lights; an adult's might contain a working miniature steam engine. Similarly, good models should aim to do only a few important things well.[2]

A model is thus, by its very nature, a highly simplified structure, and we should not expect it to give a perfect answer. Some degree of error is to be expected, and we can think of this risk of error as a form of model risk. However, the term 'model risk' is more subtle than it looks: not all output errors are due to model inadequacy (e.g., simulation methods generally produce errors due to sampling variation) and models that are theoretically inappropriate can sometimes achieve good results (e.g., simple options-pricing models often work well even when some of the assumptions on which they are based are known to be invalid).

We are particularly concerned in this chapter with how models can go wrong, and to appreciate these sorts of problems it clearly helps to understand how our models are constructed in the first place. To get to know our financial models Derman suggests that we should:

- Understand the securities involved, and the markets in which they are traded.
- Isolate the most important variables, and separate out the causal variables (or exogenous variables) from the caused (or endogenous) variables.
- Decide which exogenous variables are deterministic and which are stochastic or random, decide how the exogenous variables are to be modelled, and decide how the exogenous variables affect the endogenous ones.
- Decide which variables are measurable, and which are not; decide how the former are measured, and consider whether and how the non-measurable variables can be proxied or implicitly solved from other variables.
- Consider how the model can be solved, and look for the simplest possible solutions. We should also consider the possible benefits and drawbacks of using approximations instead of exact solutions.
- Program the model, taking account of programming considerations, computational time, and so on.
- Test and backtest the model.
- Implement the model, and evaluate its performance.

In financial models, model risk often boils down to some form of pricing or valuation risk: the risk that we have mispriced something somewhere in the model. It follows, then, that model risk is less of an issue when we have simple instruments: for example, we can easily price a bond using present-value methods. However, model risk can be a much greater problem for exotic positions because unobserved variables (e.g., such as volatilities), interactions between risk factors, calibration issues, numerical approximations, lack of transparency, and so on, all make pricing more difficult. Calculating the price of a bond is one thing; calculating the price of a complicated exotic option is quite another. In addition, when dealing with models of financial risk, we face the additional task of trying to integrate risks across different positions or business units, and this raises a host of issues (e.g., potential inconsistencies, aggregation problems, etc.) that do not (typically) arise in 'stand-alone' pricing models. So risk models involve all the model risk problems, and more, of pricing models.

[2] Derman (1997), p. 85.

16.2 SOURCES OF MODEL RISK

16.2.1 Incorrect Model Specification

Model risk can arise from many different sources, and one of the most important is incorrect model specification. This can manifest itself in many ways:

- *Stochastic processes might be misspecified*: We might assume that a stochastic process follows a geometric Brownian motion when it is in fact heavy tailed, we might mistake a lognormal P/L process for a normal one, and so on, and it is rarely easy to identify the 'correct' process.
- *Missing risk factors*: We might ignore factors such as stochastic volatility or fail to consider enough points across the term structure of interest rates.
- *Misspecified relationships*: We might misspecify relationships between variables (e.g., we might ignore correlations or get correlations wrong).
- *Ignoring of transactions costs, crisis and liquidity factors*: Many models ignore transactions costs and assume that markets are perfectly liquid. Such assumptions are very convenient for modelling purposes, but can lead to major errors where transactions costs are significant or market liquidity is limited. Such problems were highlighted by the difficulties experienced by portfolio insurance strategies in the October 1987 crash – where strategies predicated on dynamic hedging were unhinged by the inability to unwind positions as the market fell. The failure to allow for illiquidity led to much larger losses than the models anticipated – a classic manifestation of model risk.

There is evidence that model misspecification risk can be substantial. Hendricks (1996) investigated differences between alternative VaR estimation procedures applied to 1000 randomly selected simple FX portfolios, and found that these differences were sometimes substantial. More alarmingly, a study by Beder (1995) examined eight common VaR methodologies used by a sample of commercial institutions, applied to three hypothetical portfolios, and found that alternative VaR estimates for the same portfolio could differ by a factor of up to 14 – which is a worrying magnitude by anyone's standards. VaR estimates are thus in practice (and not just in theory!) very dependent upon the methodology and assumptions underlying their estimation, and this point is confirmed by the Berkowitz–O'Brien (2002) study discussed in Box 16.1.

Box 16.1 How Good are Banks' Risk Measurement Models?

Given the amounts invested in risk measurement models, one would hope that the models banks actually use would be fairly good ones – but, in fact, the evidence on this issue is not especially reassuring. A recent study by Berkowitz and O'Brien (2002) examines the VaR models used by six leading US financial institutions. Their results indicate that these models tend to be too conservative and in some cases highly inaccurate: banks sometimes experienced high losses very much larger than their models predicted, which suggests that these models are poor at dealing with heavy tails or extreme risks. Their results also indicate that banks' models have difficulty dealing with changes in volatility.

In addition, a comparison of banks' models with a simple univariate GARCH model indicates that the latter gives roughly comparable coverage of high losses, but also tends

to produce lower VaRs and is much better at dealing with volatility changes (e.g., such as those of August–September 1998). These results suggest that the banks' structural models embody so many approximations and other implementation compromises that they lose any edge over much simpler models such as GARCH ones. They could also be interpreted as suggesting that banks would be better off ditching their structural risk models in favour of much simpler GARCH models.[3]

These results suggest that the Basel regulations on the internal models approach to capital adequacy regulation might be counterproductive. Banks' models might be too conservative in part because regulations require subportfolio VaRs to be added for capital adequacy purposes: this would ignore any diversification benefits that arise when subportfolios are combined together, and so produce conservative (i.e., excessively high) VaR estimates. Banks' models might also have difficulty tracking changes in volatility because Basel regulations require VaR estimates to reflect market volatility over a period of at least a year, and this makes it difficult to respond to large sudden changes in market volatility.

16.2.2 Incorrect Model Application

Model risk can also arise because a good model is incorrectly applied. To quote Derman again,

> There are always implicit assumptions behind a model and its solution method. But human beings have limited foresight and great imagination, so that, inevitably, a model will be used in ways its creator never intended. This is especially true in trading environments, where not enough time can be spent on making interfaces fail-safe, but it's also a matter of principle: you just cannot foresee everything. So, even a 'correct' model, 'correctly' solved, can lead to problems. The more complex the model, the greater this possibility.[4]

One can give many instances of this type of problem: we might use the wrong model (e.g., we might use a Black–Scholes model for pricing interest-rate options when we should have used a Heath–Jarrow–Morton model, etc.); we might have initially had the right model, but have fallen behind best market practice and not kept the model up to date, or not replaced it when a superior model became available; we might run Monte Carlo simulations with a poor random number generator or an insufficient number of trials, and so on. 'The only practical defence,' as Derman continued, 'is to have informed and patient users who clearly comprehend both the model and its method of solution, and, even more important, understand what can go wrong'.[5]

16.2.3 Implementation Risk

Model risk also arises from the ways in which models are implemented. A formal model does not and cannot provide a complete specification of model implementation in every conceivable circumstance, because of the very large number of possible instruments and markets, and their varying institutional and statistical properties. However complete the model, implementation decisions still need to be made: about valuation (e.g., mark to market vs mark to model, whether to use the mean bid–ask spread, etc.), whether and how to clean the P/L series, how

[3] Similar findings are also reported by Lucas (2000) and Lopez and Walter (2001, p. 25), who find that sophisticated risk models based on estimates of complete variance–covariance matrices fail to perform much better than simpler univariate VaR models that require only volatility estimates.

[4] Derman (1997), p. 86.

[5] See note 4.

to map instruments, and so on. The extent of implementation risk can be appreciated from a study by Marshall and Siegel (1997). They sought to quantify implementation risk by looking at differences between how various commercial systems applied the RiskMetrics variance–covariance approach to specified positions based on a common set of assumptions (i.e., a 1-day holding period, a 95% VaR confidence level, delta-valuation of derivatives, RiskMetrics mapping systems, etc.). They found that any two sets of VaR estimates were always different, and that VaR estimates could vary by up to nearly 30% depending on the instrument class; they also found these variations were in general positively related to complexity: the more complex the instrument or portfolio, the greater the range of variation of reported VaRs. These results suggested that

> a naive view of risk assessment systems as straightforward implementations of models is incorrect. Although software is deterministic (i.e., given a complete description of all the inputs to the system, it has well-defined outputs), as software and the embedded model become more complex, from the perspective of the only partially knowledgeable user, they behave stochastically. . . . Perhaps the most critical insight of our work is that as models and their implementations become more complex, treating them as entirely deterministic black boxes is unwise, and leads to real implementation and model risks.[6]

16.2.4 Other Sources of Model Risk

16.2.4.1 Incorrect calibration

Incorrect calibration of an otherwise good model can also produce model risk. Parameters might be estimated with error, not kept up to date, estimated over inappropriate sample periods, and so on. Incorrect calibration can then lead to major losses if the models are then used to price traded instruments. A good example is the £90m loss made by the NatWest Bank over 1995–97. Over this period, a trader had fed his own estimates of volatility into a model used to price long-dated OTC interest-rate options. These estimates were too high and led to fictitious profits and real trading losses. Shortly after, BZW sustained a £15m loss on mispriced currency options, and the Bank of Tokyo-Mitsubishi announced a loss of £83m from faulty use of a one-factor Black–Derman–Toy model to trade swaptions. In the latter case, this model had been initially calibrated to the market prices of at-the-money swaptions, but was subsequently used to price out-of-the-money and Bermudan swaptions. Unfortunately, it wasn't designed for these options, and the mispricing didn't come to light for several years.

Risk models are also prone to calibration problems, particularly with the estimation of volatility and correlation. When volatility rises unexpectedly, firms tend to experience higher losses than suggested by their risk models, because 'true' volatility is higher than previously estimated volatility, and a highly publicised example was the experience of LTCM in the summer of 1998. Similar problems can arise when correlations unexpectedly polarise: in such cases, the portfolio loses much of its effective diversification, and 'true' risk can be much greater than estimates based on earlier correlations would suggest.

16.2.4.2 Programming problems

In addition, there is always the potential for error arising from poor programming. Programs might have errors or bugs in them, simulation methods might use poor random number

[6] Marshall and Siegel (1997), pp. 105–106.

generators or suffer from discretisation errors, approximation routines might be inaccurate or fail to converge to sensible solutions, rounding errors might add up, and so on. We can also get problems when programs are revised by people who did not originally write them, when programs are not compatible with user interfaces or other systems (e.g., datafeeds), when programs become complex or hard to read (e.g., when programs are rewritten to make them computationally more efficient but then become less easy to follow). We can also get simple counting problems, and Derman (1997, p. 87) reports the example of a convertible bond model that was good at pricing many of the options features embedded in convertible bonds, but sometimes miscounted the number of coupon payments left to maturity.

16.2.4.3 Data problems

Finally, of course, models can give incorrect answers because poor data are fed into them. The outputs of models are only as good as the inputs fed into them – 'garbage in, garbage out', as the old saying goes. Data problems can arise from many sources: from the way data are constructed (e.g., whether we mark to market or mark to model, whether we use actual trading data or end-of-day data, how we deal with bid–ask spreads, etc.), from the way time is handled (e.g., whether we use calendar time, trading time, how we deal with holidays, etc.), from data being non-synchronous, and from other sources.

Box 16.2 Endogenous Model Risk

A particularly subtle and difficult form of model risk arises from the ways in which traders or asset managers respond to the models themselves (e.g., how they respond to VaR limits or VaR incentives). Traders are likely to have a reasonable idea of the errors in the parameters – particularly volatility or correlation parameters – used to estimate VaR, and such knowledge will give the traders an idea of which positions have under- and overestimated risks. If traders face VaR limits, or face risk-adjusted remuneration with risks specified in VaR terms, they will have an incentive to seek out such positions and trade them. To the extent they do, they will take on more risk than suggested by VaR estimates, and our VaR estimates will be biased downwards. VaR estimates are also likely to be biased even if traders do not have superior knowledge of underlying parameter values. The reason for this is that if a trader uses an estimated variance–covariance matrix to select trading positions, then he or she will tend to select positions with low estimated risks, and the resulting changes in position sizes mean that the initial variance–covariance matrix will tend to underestimate the resulting portfolio risk.

Some plausible estimates of the sizes of this bias are reported by Ju and Pearson (1999). For instance, if a trader maximises expected returns subject to a risk constraint specified in terms of estimated VaR, their results suggest that this bias is large when K, the dimension of the variance–covariance matrix, is high (e.g., 50 or more) and the sample size is small or moderate. The bias is also high when an exponentially weighted moving average estimator is used, regardless of the size of K. These results suggest that firms should be very careful about using VaR estimates to control or remunerate trading, as they are subject to considerable endogenous model risk.

16.3 QUANTIFYING MODEL RISK

It often helps to try to quantify our model risk, and we can quantify model risk if we are prepared to make suitable assumptions about what it is we are uncertain about. However, before getting into this issue in any depth, we should first recognise that any quantification of model risk is inevitably limited. We might specify the nature of our uncertainty in some precise way (along lines to be discussed), but any such specification will inevitably involve some assumptions that we take for granted, and these assumptions may or may not be valid. This means that any estimates of model risk are conditional on assumptions, whose ultimate veracity is unknown. Any estimate of model risk is therefore an estimate of *some particular form* of model risk, which inevitably ignores *other* aspects of model risk. Hence, any estimates of model risk are likely to be underestimates of overall model risk. This in turn leads to the somewhat depressing conclusion that if we produce a low estimate of model risk, then we have little real idea of whether overall model risk is high or low, but if we produce a high estimate of model risk, then we can be fairly confident that the model risk is high – and also probably higher than we think. The consequence is a pessimist's paradise: if it looks good, don't count on it, and if it looks bad, it's probably worse.

With this important caveat out of the way, we now consider some ways in which model risk can be estimated conditional on our being able to express our uncertainty in a suitably precise form.[7] To demonstrate the principles involved, we will consider five distinct cases, in ascending order of sophistication.

Case one: single unknown parameter

In the first case, we have a single unknown parameter. In the case of a risk model, this parameter would typically be a volatility. To illustrate this, suppose we are confident that P/L is normal with a mean 0, say, but an unknown standard deviation σ. However, we have an estimate s of the standard deviation based on a sample of size n. Statistical theory tells us that, if we draw a random sample of size n from a normal distribution, then $(n-1)s^2/\sigma^2$ will be distributed as a χ^2 with $n-1$ degrees of freedom. It follows after a little rearrangement that the 90% confidence interval for σ^2 must be:

$$\frac{(n-1)s^2}{\chi^2_{0.95}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.05}} \tag{16.1}$$

Given a sample value of s, the confidence interval for the true standard deviation σ is then:

$$\sqrt{\frac{n-1}{\chi^2_{0.95}}} \times s < \sigma < \sqrt{\frac{n-1}{\chi^2_{0.05}}} \times s \tag{16.2}$$

and this yields the following confidence interval for the VaR:

$$-\sqrt{\frac{n-1}{\chi^2_{0.95}}} \times sz_\alpha < VaR = -\sigma z_\alpha < -\sqrt{\frac{n-1}{\chi^2_{0.05}}} \times sz_\alpha \tag{16.3}$$

This confidence interval gives us an estimate of our model risk, which in this case takes the form of uncertainty about a parameter (i.e., parameter risk).

[7] For more on these issues in the VaR context, see, e.g., Dowd (2000a) or Siu *et al.* (2001) .

Case two: two unknown parameters

This earlier example is extremely simplistic, and in more realistic cases we will usually be uncertain about the values of more than one parameter. To extend our earlier example and make it more plausible, we might be uncertain about the mean μ as well as the standard deviation σ. In this case, a closed-form solution such as Equation (16.3) does not appear to be possible, so we would resort to simulation instead. Given the information at hand, we might reasonably assume that we have an estimated mean m drawn from a normal distribution with unknown mean μ and unknown standard deviation $\sigma/\sqrt{n}$. For any given values of μ and σ, we can estimate the VaR as $VaR = -\mu + \sigma z_\alpha$, so we use a simulation routine to provide us with values of μ and σ from our sample estimates m and s. More particularly, we carry out a large number M of simulation trials, each of which has two stages. In the first stage, we use the fact that $(n-1)s^2/\sigma^2$ is distributed as a χ^2 with $n-1$ degrees of freedom to simulate a value of σ from the known value s and the simulated value of a χ^2 variate. Using this value of σ, we then draw a corresponding value of μ from a normal distribution with mean m and standard deviation $\sigma/\sqrt{n}$. We then use each trial's values of μ and σ to provide us with that trial's estimate of the VaR. Carrying out a large number of trials gives us a large number of VaR estimates, and we can infer the confidence interval for our VaR from the distribution of these simulated VaR estimates. This type of estimation routine is straightforward to program and can be easily applied to larger numbers of unknown parameters: the key is to simulate the values of unknown parameters from what we think we know about the density functions that describe our subjective (or Bayesian, if you like) uncertainty about them.

Case three: unknown correlation(s)

This approach extends easily to multiple factor problems where there are one or more unknown correlations. We would draw up a more elaborate multiple position version of the Monte Carlo routine just considered: means would be drawn from normal distributions, and variances and covariances from χ^2 distributions, and the simulated correlations would be inferred from the simulated variances and covariances. In principle, this type of approach could be extended to n assets, and allows us to estimate the model risk associated with a variance–covariance risk model.

Case four: mixing parameter and distribution risk

More generally, we might be uncertain about both parameter values and the distribution(s) we are facing. In these cases, we would again make assumptions about the density functions of the parameters involved, but in this case we have to specify the P/L distributions as well, and choose a 'mixing' variable that determines which P/L distribution is selected in any given trial. So, for example, we might assume a probability p that P/L is drawn from a normal and a probability (1 − p) that it is drawn from, say, a t with 5 degrees of freedom. In each trial, we first draw from the mixing distribution – which in this case would be a binomial – to determine whether our P/L for that trial is to be drawn from a normal or a t. Given the result of that drawing, we select the appropriate distribution, draw random values for the parameters involved in much the same way as before, and then, finally, draw a set of simulated P/L values and obtain a trial VaR estimate from these values. We carry out this exercise a large number of times, obtain a corresponding set of VaR estimates, and use these to estimate a confidence interval for our VaR, which then gives us an estimate of our model risk.

As we extend the generality of our approach, we eventually run into the problem that we don't in general know the true distributions from which our parameters are to be drawn; nor do we generally know the parameters of those distributions. Outside the special case (and others related to it) where P/L is normal – which leads means to be normal and variances and covariances to be χ^2 – it is difficult to specify what those distributions should be.

However, even if we don't believe that we are actually dealing with a normal distribution, it may still be the case that we can be confident that a normal distribution *under*estimates model risk. In such circumstances, we can still use estimates of model risk based on a normal distribution to give us lower-bound estimates of the model risk we think we face.

In general, we can only make what we think are reasonable assumptions about the parameter distributions and hope for the best. In the final analysis, there is no avoiding the fact that we have to make *some* assumptions somewhere down the line, and our results will depend on what we assume.

16.4 MANAGING MODEL RISK

Given that we can never eliminate model risk, the only option left is to learn to live with (i.e., manage) it. There are many ways it can be managed, and these fall under three main headings: those applicable by the individual risk measurement/management practitioners who build models and use them; those applicable by the managers the risk practitioners report to; and organisational methods, which involve the establishment of suitable institutional or procedural structures to detect and counteract model risk.

16.4.1 Managing Model Risk: Some Guidelines for Risk Practitioners

There are a number of ways practitioners can protect themselves against model risk:

- *Be aware of model risk*: First and foremost, practitioners should simply be aware of it, and aware of the limitations of the models they use. They should also be aware of the comparative strengths and weaknesses of different models, be knowledgeable of which models suit which problems, and be on the lookout for models that are applied inappropriately.
- *Identify, evaluate and check key assumptions*: Users should explicitly set out the key assumptions on which a model is based, evaluate the extent to which the model's results depend on these assumptions, and check them as much as possible (e.g., using statistical tests).
- *Test models against known problems*: It is always a good idea to check a model on simple problems to which one already knows the answer, and many problems can be distilled to simple special cases that have known answers. If the model fails to give the correct answer to a problem whose solution is already known, then we immediately know that there is something wrong with it.
- *Choose the simplest reasonable model*: Exposure to model risk is reduced if practitioners always choose the simplest reasonable model for the task at hand. Occam's razor applies just as much in model selection as in anything else: unnecessary complexity is never a virtue. Whenever we choose a more complex model over a simpler one, we must have a clear reason for doing so.
- *Backtest and stress test the model*: Practitioners should evaluate model adequacy using stress tests (explained in Chapter 13) and backtests (explained in Chapter 15).
- *Estimate model risk quantitatively*: Where feasible, practitioners should estimate model risk quantitatively (e.g., using simulation methods), keeping in mind the point made earlier that

any quantitative estimate of model risk is almost certainly an underestimate, because not all model risk is quantifiable.

- *Don't ignore small problems*: Practitioners should resist the temptation to explain away small discrepancies in results and sweep them under the rug. Small discrepancies are often good warning signals of larger problems that will manifest themselves later if they are not sorted out.
- *Plot results and use non-parametric statistics:* Graphical outputs can be extremely revealing, and simple histograms or plots often show up errors that might otherwise be very hard to detect. For example, a plot might have the wrong slope or shape or have odd features such as kinks that flag up an underlying problem. Summary statistics and simple non-parametric tests can also be useful in giving a feel for data and results.
- *Re-evaluate models periodically*: Models should be recalibrated and re-estimated on a regular basis, and the methods used should be kept up to date.

16.4.2 Managing Model Risk: Some Guidelines for Senior Managers

Senior managers can respond to model risk by ensuring that they are properly informed themselves: without being expert risk modellers, they should have some basic appreciation of the issues involved so they can understand what their risk managers are talking about. They should try to avoid thinking of risk models as black boxes, and they should learn what questions to ask and how to judge the answers. Most of all, they should learn from the mistakes of others: derivatives mistakes are well publicised and many of these stem from model risk problems and the failure of managers to pay attention to warning signals or ensure that their risk control systems were working.

Senior managers should also listen to their risk people and take their concerns seriously. They should be on their guard against the temptation to put too much trust in traders and disregard those who question what they are up to. All too often, 'star' traders have turned out to be making large losses rather than profits, and the managers they reported to have turned a blind eye because they were dazzled by the profits they appeared to be making. The tendency to believe what one wants to believe is a natural human failing, and it is also a key factor in many major derivatives disasters: in Orange County in 1994 (where the board of supervisors ignored warnings about the exposure of the County's investment portfolio to a rise in interest rates), in Barings in 1994–95 (where Barings' senior management ignored repeated warnings about the activities of Nick Leeson), and in many other cases. Managers need to inculcate a culture that takes risk management seriously – and they need to resist the temptation to regard risk management as an obstacle to their next bonus.

To combat model risk, senior managers should also be on their guard for specific types of situation where model risk can create serious problems. One such situation is 'model creep'. This occurs where a model is initially designed for one type of problem, and performs well on that problem, but is gradually applied to more diverse situations to which it is less suited or even not suited at all. A perfectly good model can then end up as a major liability not because there is anything wrong with it, but because users don't appreciate its limitations.

Similarly, managers need to be aware of product cycles and the constraints of product development. For example, when a new model is initially developed, its superior pricing properties will tend to make large profits for those who first trade it; these profits will encourage others to enter the market, and profits will rapidly fall. However, initially high profits will tempt other firms to get into the market prematurely, before their own models are fully functional: firms

that enter the market too quickly will be unable to compete effectively with more established operators, and will sometimes make large losses. Similarly, managers should resist the temptation to push for risk models to be made operational before they are ready: good risk models can take a long time to develop, especially if the data and organisational infrastructure that they require have to be developed first.

Senior managers should also have a good appreciation of how risk estimates are affected by trading strategies (i.e., they should be aware of the endogenous model risk issue; see Box 16.2). As Shaw points out,

> many factor models fail to pick up the risks of *typical* trading strategies which can be the greatest risks run by an investment bank. According to naïve yield factor models, huge spread positions between on-the-run bonds and off-the-run bonds are riskless! According to naïve volatility factor models, hedging one year (or longer dated) implied volatility with three month implied volatility is riskless, provided it is done in the 'right' proportions – i.e., the proportions built into the factor model! It is the *rule*, not the exception, for traders to put on spread trades which defeat factor models *since they use factor type models to identify richness and cheapness*! (his emphasis)[8]

In other words, managers need to appreciate how the choice of model affects trading strategies, and how the latter can then distort the model outputs.

More fundamentally, senior managers need to be on their guard against the tricks that people play – how traders can hide losses and 'game' VaR models to their advantage, and so on. It is also important that managers do not underrate the abilities of those below them to play the system and get away with it: a 1997 survey by Cap Gemini found that although three-quarters of risk managers believed that their organisation was immune to a Barings-style scandal, almost the same proportion of traders believed the opposite, and 85% of traders believed they could hide trades from their managers. These findings suggest that many firms are a lot *less* secure than their managers think. In combating these sorts of problems, managers also need to recognise that if they are to be effective, their risk managers must have the knowledge and skills to match traders, and this means that they must be suitably remunerated – otherwise, there is little incentive for those with the necessary skills to want to do anything but trade.

Senior managers can also reduce their vulnerability to model risk problems by encouraging a multidisciplinary team approach to model building. They should not see models as incomprehensible formulas that quants or risk modellers hand over to programmers to make even more incomprehensible. Instead, they should see models as the product of an interdisciplinary team, involving inputs from mathematicians, statisticians, computer scientists, finance experts, accountants, traders, model users, and others. They should encourage people from these disparate groups to understand each other, and foster a climate of constructive criticism.

16.4.3 Institutional Methods to Manage Model Risk

16.4.3.1 Procedures to vet, check and review models

Firms themselves can also combat the dangers of model risk through appropriate institutional devices. One basic defence is a sound system to vet models before they are approved for use and then check and periodically review them. A good model-vetting procedure is proposed by Crouhy *et al.* and involves the following four steps:[9]

[8] Shaw (1997), p. 215.
[9] Crouhy *et al.* (2001), pp. 607–608.

- *Documentation*: The risk manager should ask for a complete specification of the model, including its mathematics, components, computer code, and implementation features (e.g., numerical methods and pricing algorithms used). The information should be in sufficient detail to enable the risk manager to reproduce the model from the information provided.
- *Soundness*: The risk manager should check that the model is a reasonable one for the instrument(s) or portfolio concerned.
- *Benchmark modelling*: The risk manager should develop a benchmark model and test it against well-understood approximation or simulation methods.
- *Check results and test the proposed model*: The final stage involves the risk manager using the benchmark model to check the performance of the proposed model. The model should also be checked for zero-arbitrage properties such as put–call parity, and should then be stress tested to help determine the range of parameter values for which it will give reasonable estimates.

All these stages should be carried out free of undue pressures from the front office, and traders should not be allowed to vet their own pricing models. However, there is also a need to vet risk models as well, and this raises a difficult problem: if we use risk managers to check the traders' models, then who checks the risk models? A partial answer, inevitably, is that much of the responsibility for vetting risk management models must lie with the risk management function itself. In larger firms, one could envisage this task as carried out in a separate unit within the risk management function, so one unit of risk management builds risk models, and the other checks them. In addition, it is also good practice for firms to institute periodic risk audits – audits of all aspects of the firm's risk management, including audits of the models used, carried out by external parties with the skills to do so.[10] Such audits would also provide useful feedback, especially on areas where the firm was lagging behind best market practice.

It is also important to keep good records, so each model should be fully documented in the middle (or risk) office. Risk managers should have full access to the model at all times, as well as access to real trading and other data that might be necessary to check models and validate results. The ideal should be to give the middle office enough information to be able to check any model or model results at any time, and do so using appropriate (i.e., up-to-date) data sets. This information set should include a log of model performance with particular attention to any problems encountered and what (if anything) has been done about them. Finally, there should be a periodic review (as well as an occasional spot check) of the models in use, to ensure that model calibration is up to date and that models are upgraded in line with market best practice, and to ensure that obsolete models are identified as such and taken out of use.

16.4.3.2 Independent risk oversight

At a more fundamental level – and this is absolutely critical to sound risk management – the firm should also set up a suitable independent risk oversight (IRO) or middle office unit. This unit should encompass risk measurement as well as risk management, should be independent of line-execution areas (such as treasury, trading, portfolio management, asset-liability management, etc.), and its head, the chief risk officer (CRO), should report to the CEO and, ideally, sit on the board or other governing body. The middle office should have a clear mandate from

[10] Such audits are not only good practice in themselves, but also help senior managers establish that they have carried out due diligence – which is a particularly important consideration for firms operating in the US with the advent of Sarbanes-Oxley and other corporate governance reforms.

senior management, and its policies should reflect the corporate policies towards risk – the corporate risk appetite, and so on. It should also have its own designated budget and, to help mitigate any temptation for the middle office to go along with excessive risk-taking elsewhere in the organisation, the remuneration of the CRO and his or her staff should not be tied to the performance of other units (such as trading profits).

This unit should have authority to block any trading or asset management activity within the organisation, and have authority over the use of all pricing or risk models. It should have responsibility for monitoring risk independently of other business units, and particularly the front office. In carrying out these tasks, it should seek to ensure a balance between an excessively prohibitionist stance on risk (i.e., everything not expressly allowed is forbidden) and an excessively lenient stance (i.e., everything not expressly forbidden is allowed), and also aim to ensure that all interested parties are fully involved in the firm's internal risk measurement/management dialogue. The middle office should also be responsible for all aspects of risk estimation, including stress testing, backtesting and (at least some) contingency planning, for ensuring that all models are adequate for the tasks to which they are being used (i.e., take responsibility for vetting, checking and monitoring the models used), for reporting and disseminating risk information throughout the organization, and for protecting and maintaining the integrity of the firm's risk measurement and risk management systems.[11]

16.5 CONCLUSIONS

Model risk is one of the most important and least appreciated areas of market risk measurement. We go about our work in risk management as if we know a lot that we actually don't: we often treat our models as if they are correct, we might treat parameters as if they are known, and so on. And yet in the strict sense of the word we actually *know* very little at all. Instead, we only ever work with assumptions and have no choice but to do so. However, it is then all too easy to fall into the trap of starting to think of our assumptions as if they were true knowledge. We are particularly vulnerable to this trap because it is a basic human characteristic to seek confirmation of our beliefs: we all want the world to confirm our views of it, and we tend to brush aside inconvenient evidence that we might have got it wrong. These mental habits then make us vulnerable in those situations where reality rudely asserts itself and reminds us that our beliefs can sometimes be disastrously off the mark. Model risk casts its shadow over everything we do in risk management, and prudence suggests that we should always ask ourselves what would happen if our assumptions fail to hold. Of course, this is easier to say than to do, but the financial markets are littered with the remains of those who have ignored model risk and thought they could get away with it, and these include some of the biggest names of their day. Ultimately, model risk is like the proverbial ghost at the banquet – an unwelcome guest, but one that we would be very unwise to ignore.

[11] Besides having good model vetting procedures and independent risk oversight, there are also other ways that firms can deal with model risk. For example, one sound practice is for firms to keep reserves against possible losses from model risk. The reserves attributed to a position should reflect some measure of the model risk involved, so that positions with higher model risk get higher reserve charges than positions with lower model risk. Such charges not only provide the firm with a cushion to absorb possible losses from model risk, but also help to ensure that the cost of model risk is accounted for and attributed to the positions concerned. Firms can also counteract model risk by taking account of it in setting position limits. If a position is known to have considerable model risk, a firm can limit its exposure to this source of model risk by imposing a tighter position limit.

Bibliography

Abken, P. A. (2000) 'An empirical evaluation of value at risk by scenario simulation.' *Journal of Derivatives* Summer issue: 12–29.

Acerbi, C. (2002) 'Spectral measures of risk: a coherent representation of subjective risk aversion.' *Journal of Banking and Finance* 26: 1505–1518.

Acerbi, C. (2004) 'Coherent representations of subjective risk aversion.' Pp. 147–207 in G. Szegö (ed.) *Risk Measures for the 21st Century*. Chichester: John Wiley and Sons, Ltd.

Acerbi, C. and D. Tasche (2002) 'On the coherence of expected shortfall.' *Journal of Banking and Finance* 26: 1487–1503.

Adelman, I. (1990) 'Factor analysis.' Pp. 90–94 in J. Eatwell, M. Milgate and P. Newman (eds) *The New Palgrave Time Series and Statistics*. London and Basingstoke: Macmillan.

Albanese, C., K. Jackson and P. Wiberg (2001) 'Hedging with value-at-risk.' Mimeo. University of Toronto.

Alexander, C. (1998) 'Volatility and correlation: measurement, models and applications.' Pp. 125–171 in C. Alexander (ed.) *Risk Measurement and Analysis, Vol. One: Measuring and Modelling Financial Risk*. Chichester: John Wiley and Sons, Ltd.

Alexander, C. (2000) 'Orthogonal methods for generating large positive semi-definite variance–covariance matrices.' University of Reading ISMA Centre *Discussion Paper in Finance* 2000–06.

Alexander, C. and A. Chibumba (1997) 'Orthogonal GARCH: an empirical validation in equities, foreign exchange and interest rates.' Mimeo. University of Sussex.

Almgren, R. and N. Chriss (1999) 'Value under liquidation.' *Risk* 12 (December): 61–63.

Anderson, G. (1994) 'Simple tests of distributional form.' *Journal of Econometrics* 62: 265–276.

Andersen, L. (2000) 'A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR model.' *Journal of Computational Finance* 3 (Winter): 1–32.

Andersen, T. G. and T. Bollerslev (1997) 'Intraday periodicity and volatility persistence in financial markets.' *Journal of Empirical Finance* 4: 115–158.

Andersen, T. G. and T. Bollerslev (1998a) 'Deutsche Mark–Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies.' *Journal of Finance* 53: 219–265.

Andersen, T. G. and T. Bollerslev (1998b) 'Answering the sceptics: yes, standard volatility models do provide accurate forecasts.' *International Economic Review* 39: 885–905.

Andersen, T. G., T. Bollerslev, F. Diebold and P. Labys (2001) 'The distribution of realized exchange rate volatility.' *Journal of the American Statistical Association* 96: 42–55.

Andrews, D. K. and M. Buchinsky (1997) 'On the number of bootstrap repetitions for bootstrap standard errors, confidence intervals, and tests.' Cowles Foundation Paper 1141R, July 1997.

Aragonés, J. R., C. Blanco and J. Mascarenas (2001) 'Active management of equity investment portfolios.' *Journal of Portfolio Management* 27 (Spring): 39–46.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1997) 'Thinking coherently.' *Risk* 10 (November): 68–71.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999) 'Coherent measures of risk.' *Mathematical Finance* 9 (November): 203–228.

Ashton, D. and M. Tippett (2004) 'The distribution of "short period" equity returns.' Mimeo. University of Bristol and Loughborough University.

Bahar, R., M. Gold, T. Kitto and C. Polizu (1997) 'Making the best of the worst.' *Risk* 10 (August): 100–103.

Bangia, A., F. Diebold, T. Schuermann and J. Stroughair (1999) 'Liquidity on the outside.' *Risk* 12 (June): 68–73.

Bank for International Settlements (2000) *Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues*. BIS Committee on the Global Financial System. Basel: Bank for International Settlements, April 2000.

Bank for International Settlements (2004a) *Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in April 2004: Preliminary Global Results*. Basel: Bank for International Settlements. September.

Bank for International Settlements (2004b) *Triennial and Semiannual Surveys on Positions in Global Over-the-Counter (OTC) Derivatives Markets at End-June 2004. Concentration Measures for OTC Derivatives Markets from December 1998 to June 2004.* Basel: Bank for International Settlements.

Barone-Adesi, G., F. Bourgoin and K. Giannopoulos (1998) 'Don't look back.' *Risk* 11 (August) 1998: 100–103.

Barone-Adesi, G., and K. Giannopoulos (2000) 'Non-parametric VaR techniques. Myths and realities.' Mimeo. Universita della Svizzera Italiana and City University Business School; and Westminster Business School. November 2000.

Barone-Adesi, G., K. Giannopoulos and L. Vosper (1999) 'VaR without correlations for portfolios of derivatives securities.' *Journal of Futures Markets* 19: 583–602.

Barraquand, J. (1995) 'Numerical valuation of high dimensional multivariate European securities.' *Management Science* 41: 1882–1891.

Bartlmae, K. and F. A. Rauscher (2000) 'Measuring DAX market risk: a neural network volatility mixture approach.' Mimeo. DaimlerChrysler AG, Research and Technology, Ulm.

Basak, S. and A. Shapiro (2001) 'Value-at-risk-based risk management: optimal policies and asset prices.' *Review of Financial Studies* 14: 371–405.

Basel Committee on Banking Supervision (1996) *Amendment to the Capital Accord to Incorporate Market Risks*. Basel: Bank for International Settlements.

Basel Committee on Banking Supervision (2004) *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Basel: Bank for International Settlements.

Bassi, F., P. Embrechts and M. Kafetzaki (1998) 'Risk management and quantile estimation.' Pp. 111–130 in R. J. Adler, R. E. Feldman and M. S. Taqqu (eds) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston: Birkhaüser.

Bauer, C. (2000) 'Value at risk using hyperbolic distributions.' *Journal of Economics and Business* 52: 455–467.

Baumol, W. J. (1963) 'An expected gain confidence limit criterion for portfolio selection.' *Management Science* 10: 174–182.

Bauwens, L., S. Laurent and J. Rombouts (2004) **'**Multivariate GARCH models: a survey.' *Journal of Applied Econometrics*. Forthcoming.

Beder, T. (1995) 'VaR: seductive but dangerous.' *Financial Analysts Journal* 51 (September/October): 12–24.

Beirlant, J., Y. Goegebeur, J. Teugels, J. Seger, D. De Waal and C. Ferro (2004) *Statistics of Extremes: Theory and Applications*. Chichester: John Wiley and Sons, Ltd.

Berkowitz, J. (2000a) 'A coherent framework for stress-testing.' *Journal of Risk* 2: 1–11.

Berkowitz, J. (2000b) 'Breaking the silence.' *Risk* 13 (October): 105–108.

Berkowitz, J. (2001) 'Testing density forecasts, with applications to risk management.' *Journal of Business and Economic Statistics* 19: 465–474.

Berkowitz, J. and J. O'Brien (2002) 'How accurate are value-at-risk models at commercial banks?' *Journal of Finance* 57: 1093–1112.

Bertsimas, D. and A. Lo (1998) 'Optimal control of execution costs.' *Journal of Financial Markets* 1: 1–50.

Bickel, P. J. and K. A. Docksum (1981) 'An analysis of transformations revisited.' *Journal of the American Statistical Association* 76: 296–311.

Billio, M. and L. Pelizzon (1997) 'A switching volatility approach to improve the estimation of value-at-risk.' Mimeo. University of Venice and London Business School.

Black, F. (1976) 'The pricing of commodity contracts.' *Journal of Financial Economics* 2: 167–179.
Black, F. and M. Scholes (1973) 'The pricing of options and corporate liabilities'. *Journal of Political Economy* 81: 637–654.
Blake, D., A. J. G. Cairns and K. Dowd (2001) 'PensionMetrics: stochastic pension plan design and value-at-risk during the accumulation phase.' *Insurance: Mathematics and Economics* 29 (October): 187–215.
Blake, D., A. J. G. Cairns and K. Dowd (2003) 'PensionMetrics 2: stochastic pension plan design during the distribution phase.' *Insurance: Mathematics and Economics*, 33: 29–47.
Blanco, C. and G. Ihle (1998) 'How good is your VaR? Using backtesting to assess system performance.' *Financial Engineering News* (August): 1–2.
Blanco, C. and D. Soronow (2001) 'Jump diffusion processes – energy price processes used for derivatives pricing and risk management.' *Commodities Now* (June): 83–87.
Bolancé, C., M. Guillen and J. P. Nielsen (2003) 'Kernel density estimation of actuarial loss functions.' *Insurance: Mathematics and Economics* 32: 19–36.
Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity.' *Journal of Econometrics* 31: 307–327.
Borse, G. J. (1997) *Numerical Methods with MATLAB: A Resource for Scientists and Engineers.* Boston: PWS Publishing Company.
Bouchaud, J.P. and M. Potters (2000a) *Theory of Financial Risks: From Statistical Physics to Risk Management.* Cambridge: Cambridge University Press.
Bouchaud, J. P. and M. Potters (2000b) 'Worse fluctuation method for fast value-at-risk estimates.' Mimeo. Science and Finance, 109–111 rue Victor Hugo, 92532 Levallois CEDEX, France.
Boudoukh, J., M. Richardson and R. Whitelaw (1995) 'Expect the worst.' *Risk* 8 (September): 100–101.
Boudoukh, J., M. Richardson and R. Whitelaw (1998) 'The best of both worlds: a hybrid approach to calculating value at risk.' *Risk* 11 (May): 64–67.
Bouyé, E., A. Nikeghbali, G. Riboulet and T. Roncalli (2000) 'Copulas for finance: a reading guide and some applications.' Mimeo. Financial Econometric Research Centre, City University Business School, London; and Groupe de Recherche Opérationnelle, Crédit Lyonnais, Paris.
Box, G. E. P. and D. R. Cox (1964) 'An analysis of transformations.' *Journal of the Royal Statistical Society, Series B*, 26: 211–252.
Boyle, P., M. Broadie and P. Glasserman (1997) 'Monte Carlo methods for security pricing.' *Journal of Economic Dynamics and Control* 21: 1267–1321.
Brace, A., D. Gatarek and M. Musiela (1997) 'The market model of interest rate dynamics.' *Mathematical Finance* 7: 127–155.
Breckling, J., E. Eberlein and P. Kovic (2000) 'A tailored suit for risk management: hyperbolic model.' Pp. 189–202 in J. Franke, W. Härdle and G. Stahl (eds) *Measuring Risk in Complex Stochastic Systems*. New York: Springer Verlag. Lecture Notes in Statistics Series No. 147.
Breuer, T. and G. Krenn (1999) *Guidelines on Market Risk. Volume 5. Stress Testing*. Vienna: Oesterreichische Nationalbank.
Breuer, T. and G. Krenn (2000) 'Identifying stress test scenarios.' Mimeo. Fachhochschule Voralberg, Dornbirn, Austria, and Oesterreichische Nationalbank, Vienna.
Britten-Jones, M. and S. M. Schaefer (1999) 'Non-linear value-at-risk.' *European Finance Review* 2: 161–187.
Broadie, M. and P. Glasserman (1998) 'Simulation for option pricing and risk management.' Pp. 173–207 in C. Alexander (ed.) *Risk Management and Analysis. Vol. 1: Measuring and Modelling Financial Risk*. Chichester: John Wiley and Sons, Ltd.
Brock, W. A., W. D. Dechert and J. Scheinkman (1987) 'A test for independence based on the correlation dimension.' University of Wisconsin-Madison *SSRI Workshop Paper 8702*.
Butler, J. S. and B. Schachter (1998) 'Improving value at risk with a precision measure by combining kernel estimation with historical simulation.' *Review of Derivatives Research* 1: 371–390.
Byström, H. N. E. (2000) 'Orthogonal GARCH and covariance matrix forecasting in a stress scenario: the Nordic stock markets during the Asian financial crisis 1997–1998.' Mimeo. Lund University.
Cairns, A. J. G. (2000) 'A discussion of parameter and model uncertainty in insurance.' *Insurance: Mathematics and Economics* 27: 313–330.
Cárdenas, J., E. Fruchard, E. Koehler, C. Michel and I. Thomazeau (1997) 'VaR: one step beyond.' *Risk* 10 (October): 72–75.

Cárdenas, J., E. Fruchard, J.-F. Picron, C. Reyes, K. Walters and W. Yang (1999) 'Monte Carlo within a day.' *Risk* 12 (February): 55–59.

Cerny, A. and S. Hodges (1999) 'The theory of no-good-deal pricing in financial markets.' Working paper, Imperial College.

Chappell, D. and K. Dowd (1999) 'Confidence intervals for VaR.' *Financial Engineering News*, March: 1–2.

Chernozhukov, V. and L. Umantsev (2001) 'Conditional value-at-risk: aspects of modelling and estimation.' *Empirical Economics* 26: 271–292.

Cherubini, U. and G. Della Lunga (2001) 'Fuzzy value-at-risk: accounting for market liquidity.' *Economic Notes by Banca Monte dei Paschi di Siena* 30: 293–312.

Cherubini, U. and E. Luciano (2000) 'Value at risk trade-off and capital allocation with copulas.' Mimeo. Universities of Bologna and Turin.

Cherubini, U., E. Luciano and W. Vecchiato (2004) *Copula Methods in Finance.* Chichester: John Wiley and Sons, Ltd.

Christoffersen, P. F. (1998) 'Evaluating interval forecasts.' *International Economic Review* 39: 841–862.

Christoffersen, P. F. and F. X. Diebold (2000) 'How relevant is volatility forecasting for financial risk management?' *Review of Economics and Statistics* 82: 12–22.

Christoffersen, P. F., J. Hahn and A. Inoue (2001) 'Testing and comparing value at risk measures.' *Journal of Empirical Finance* 8: 325–342.

Clewlow, L. and C. Strickland (1998) *Implementing Derivatives Models*. Chichester: John Wiley and Sons, Ltd.

Clewlow, L. and C. Strickland (2000) *Energy Derivatives: Pricing and Risk Management*. London: Lacima Publications.

Conover, W. J. (1980) *Practical Nonparametric Statistics*. 2nd edition. New York: John Wiley and Sons, Ltd.

Cont, R., M. Potters and J.-P. Bouchaud (1997) 'Scaling in stock market data: stable laws and beyond.' Mimeo. Centre d'Etudes de Saclay, and Université de Nice.

Corsi, F., G. Zumbach, U. A. Müller and M. M. Dacorogna (2001) 'Consistent high-precision volatility from high-frequency data.' *Economic Notes by Banca Monte dei Paschi di Siena*, 30: 183–204.

Cosandey, D. (2001) 'Adjusting value-at-risk for market liquidity.' *Risk* 14 (October): 115–118.

Cotter, J. (2001) 'Margin exceedances for European stock index futures using extreme value theory.' *Journal of Banking and Finance* 25: 1475–1502.

Cotter, J. (2004) 'Tail behaviour of the Euro.' *Applied Economics*. Forthcoming.

Crnkovic, C. and J. Drachman (1995) 'A universal tool to discriminate among risk measurement techniques.' Mimeo. Corporate Risk Management Group, JP Morgan.

Crnkovic, C. and J. Drachman (1996) 'Quality control.' *Risk* 9 (September): 139–143.

Crouhy, M., D. Galai and R. Mark (1998) 'The new 1998 regulatory framework for capital adequacy: 'standardized approach' versus 'internal models'.' Pp. 1–37 in C. Alexander (ed.) *Risk Management and Analysis. Vol. 1: Measuring and Modelling Financial Risk*. Chichester: John Wiley and Sons, Ltd.

Crouhy, M., D. Galai and R. Mark (2001) *Risk Management*. New York: McGraw Hill.

Cvitanic, J. and I. Karatzas (1999) 'On dynamic measures of risk.' *Finance and Stochastics* 3: 451–482.

Daniel, W. W. (1978) *Applied Nonparametric Statistics*. Boston: Houghton Mifflin Company.

Danielsson, J. (2002) 'The emperor has no clothes: limits to risk modelling.' *Journal of Banking and Finance* 26: 1273–1296.

Danielsson, J. and C. G. de Vries (1997a) 'Beyond the sample: extreme quantile and probability estimation.' Mimeo. Tinbergen Institute, Rotterdam.

Danielsson, J. and C. G. de Vries (1997b) 'Tail index estimation with very high frequency data.' *Journal of Empirical Finance* 4: 241–257.

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault and H. S. Shin (2001) 'An academic response to Basel II.' *Special Paper No. 130*, LSE Financial Markets Group.

Danielsson, J. and J.-P. Zigrand (2001) 'What happens when you regulate risk? Evidence from a simple general equilibrium model.' Mimeo. London School of Economics Financial Markets Group.

Danneberg, D. (1990) 'Distorted probabilities and insurance premiums.' *Methods of Operations Research* 63: 3–5.

Davé, R. D. and G. Stahl (1997) 'On the accuracy of VaR estimates based on the variance–covariance approach.' Mimeo. Olsen and Associates, Zürich; and Federal Banking Supervision Office, Berlin.

Davison, A. C. and D. V. Hinkley (1997) *Bootstrap Methods and their Applications*. Cambridge: Cambridge University Press.
Deans, J. (2000) 'Backtesting.' Pp. 261–289 in M. Lore and L. Borodovsky (eds) *The Professional's Handbook of Financial Risk Management*. Oxford: Butterworth-Heinemann.
Dembo, R. (1997) 'Value-at-risk and return,' *NetExposure* No. 1 (October).
Denecker, K., S. Van Assche, J. Cromberz, R. Vander Vennet and I. Lemahieu (2001) 'Value-at-risk prediction using context modelling.' *European Physical Journal B* 20: 481–492.
Derivatives Policy Group (1995) *A Framework for Voluntary Oversight*. New York: Derivatives Policy Group.
Derman, E. (1997) 'Model risk.' Pp. 83–88 in S. Grayling (ed.) *VaR – Understanding and Applying Value-at-Risk*. London: Risk Publications.
Dias, M. A. G. (2001) 'Monte Carlo simulation of stochastic processes.' Mimeo. PUC-Rio, Rio de Janeiro.
Diebold, F. X., T. A. Gunther and A. S. Tay (1998a) 'Evaluating density forecasts with applications to financial risk management.' *International Economic Review* 39: 863–883.
Diebold, F. X., A. Hickman, A. Inoue and T. Schuermann (1998b) 'Scale models.' *Risk* 11 (January): 104–107.
Diebold, F. X., T. Schuermann and J. D. Stroughair (2000) 'Pitfalls and opportunities in the use of extreme value theory in risk management.' *Journal of Risk Finance* 1: 30–35.
Dowd, K. (1999a) 'A VaR approach to risk-return analysis.' *Journal of Portfolio Management* 25 (Summer): 60–67.
Dowd, K. (1999b) 'VaR by increments.' *Risk* (*Enterprise-Wide Risk Management Special Report*) 11 (November): 31–33.
Dowd, K. (1999c) 'Financial risk management.' *Financial Analysts Journal* 55, No. 4: 65–71.
Dowd, K. (2000a) 'Estimating value-at-risk: a subjective approach.' *Journal of Risk Finance* 1 (Summer): 43–46.
Dowd, K. (2000b) 'Accounting for value at risk.' *Journal of Risk Finance* 2 (Fall): 51–58.
Dowd, K. (2002) 'A bootstrap backtest.' *Risk* 15 (October): 93–94.
Dowd, K. (2003) 'Temporal aggregation of volatility processes: a simulation study.' Mimeo. Nottingham University Business School.
Dowd, K. (2004a) 'A modified Berkowitz back-test.' *Risk* 17 (April): 86.
Dowd, K. (2004b) 'The inflation "fan charts": an evaluation.' *Greek Economic Review*, 23 (Spring): 99–111.
Dowd, K., D. Blake and A. J. G. Cairns (2004) 'Long-term value at risk.' *Journal of Risk Finance*, 5, No. 2 (Winter/Spring 2004): 52–57.
Dowd, K. and P. Fackler (2004) 'Estimating VaR with copulas.' *Financial Engineering News* No. 39, September–October 2004: 9, 21.
Dowd, K. and D. Rowe (2004) 'Introduction to VaR Models.' Chapter A2 in C. O. Alexander and E. Sheedy (eds) *The Professional Risk Managers' Handbook: A Comprehensive Guide to Current Theory and Best Practices*. PRMIA Professional Risk Manager.
Drost, F. and T. Nijman (1993) 'Temporal aggregation of Garch processes.' *Econometrica* 61: 909–927.
Drzik, J. (1996) 'Putting risk in its place.' *Firmwide Risk Management: A Risk Special Supplement* July: 14–16.
Duffie, D. and J. Pan (1997) 'An overview of value at risk.' *Journal of Derivatives* 4 (Spring): 7–49.
Duffie, D. and J. Pan (1999) 'Analytical value-at-risk with jumps and credit risk.' Mimeo. Graduate School of Business, Stanford University.
Duffie, D. and A. Ziegler (2001) 'Liquidation risk.' Mimeo, Graduate School of Business, Stanford University.
D'Vari, R. and J. C. Sosa (2000) 'Value at risk estimates for Brady bond portfolios.' *Journal of Fixed Income* 10 (December): 7–23.
Eberlein, E. (1999) 'Recent advances in more realistic risk management: the hyperbolic model.' Mimeo. University of Freiburg.
Eberlein, E., J. Kallsen and J. Kristen (2001) 'Risk management based on stochastic volatility.' Mimeo. University of Freiburg.
Eberlein, E., Keller, U. and K. Prause (1998) 'New insights into smile, mispricing and value at risk: the hyperbolic model.' *Journal of Business* 71: 371–406.

Eberlein, E. and K. Prause (2000) 'The generalized hyperbolic model: financial derivatives and risk measures.' *FDM Preprint 56*, University of Freiburg.

Efron, B. (1979) 'Bootstrap methods: another look at the jackknife.' *Annals of Statistics* 7: 1–26.

Efron, B. and R. J. Tibshirani (1993) *An Introduction to the Bootstrap*. New York and London: Chapman and Hall.

Elton, E. J. and M. J. Gruber (1995) *Modern Portfolio Theory and Investment Analysis*. 5th edition. New York: John Wiley and Sons, Ltd.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997) *Modelling Extreme Events for Insurance and Finance*. Berlin: Springer Verlag.

Embrechts, P., A. McNeil and D. Straumann (1999a) 'Correlation pitfalls and alternatives.' *Risk* 12, No. 5, May: 69–71.

Embrechts, P., A. McNeil and D. Straumann (1999b) 'Correlation and dependence in risk management: properties and pitfalls.' Mimeo. ETHZ Zentrum, Zürich.

Engle, R. F. (1982) 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.' *Econometrica* 50: 987–1007.

Engle, R. F. and S. Manganelli (1999) 'CAViaR: conditional autoregressive value at risk by regression quantiles.' Mimeo. University of California, San Diego.

Engle, R. F. and J. Mezrich (1996) 'GARCH for groups.' *Risk* 9 (June): 36–40.

Engle, R. F. and V. Ng (1993) 'Measuring the impact of news on volatility.' *Journal of Finance* 48: 1749–1778.

Estrella, A. (1996) 'Taylor, Black and Scholes: series approximations and risk management pitfalls.' Pp. 359–379 in *Risk Measurement and Systemic Risk. Proceedings of a Joint Central Bank Research Conference*. Washington, DC: Board of Governors of the Federal Reserve System.

Estrella, A., D. Hendricks, J. Kambhu, S. Shin and S. Walter (1994) 'The price risk of options positions: measurement and capital requirements.' Federal Reserve Bank of New York *Economic Policy Review* (Fall): 27–43.

Evans, M., N. Hastings and B. Peacock (2000) *Statistical Distributions*. 3rd edition. New York: John Wiley and Sons, Ltd.

Fallon, W. (1996) 'Calculating value-at-risk.' Wharton School Financial Institutions Center *Working Paper 96–49*.

Feuerverger, A. and A. C. M. Wong (2000) 'Computation of value-at-risk for nonlinear portfolios.' *Journal of Risk* 3 (Fall): 37–55.

Figlewski, S. (1994) 'Forecasting volatility using historical data.' New York University *Working Paper S-94-13*.

Figlewski, S. and B. Gao (1999) 'The adaptive mesh model: a new approach to efficient option pricing.' *Journal of Financial Economics* 53: 313–351.

Finger, C. C. (1996) 'Accounting for "pull to par" and "roll down" for RiskMetrics™ cashflows.' *RiskMetrics™ Monitor*, Third Quarter: 4–11.

Fischer, T. (2003) 'Risk capital allocation by coherent risk measures based on one-sided moments.' *Insurance: Mathematics and Economics* 32: 135–146.

Fishburn, P. C. (1977) 'Mean-risk analysis with risk associated with below-target returns.' *American Economic Review* 67: 116–126.

Focardi, S. M. and F. J. Fabozzi (2003) 'Fat tails, scaling, and stable laws: a critical look at modeling extremal events in financial phenomena.' *Journal of Risk Finance* (Fall): 5–26.

Fong, H. G. and K.-C. Lin (1999) 'A new analytical approach to value at risk.' *Journal of Portfolio Management* 25 (May): 88–98.

Frey, R. (2000) 'Market illiquidity as a source of model risk in dynamic hedging.' Mimeo. Swiss Banking Institute, University of Zürich.

Froot, K., P. O'Connell and M. Seasholes (2001) 'The portfolio flows of international investors.' *Journal of Financial Economics* 59: 151–193.

Frye, J. (1996) 'Principals of risk: finding value-at-risk through factor-based interest rate scenarios.' Mimeo. NationsBank-CRT, Chicago.

Fusai, G. and E. Luciano (1998) 'Measuring VaR under optimal and suboptimal portfolio polifices.' Mimeo. Universities of Florence and Turin.

Garman, M. B. (1996a) 'Making VaR more flexible.' *Derivatives Strategy* (April): 52–53.

Garman, M. B. (1996b) 'Improving on VaR.' *Risk* 9 (May): 61–63.

Garman, M. B. (1996c) 'Making VaR proactive.' Berkeley, CA: Financial Engineering Associates. FEA.
General Accounting Office (1994) *Financial Derivatives: Actions Needed to Protect the Financial System*. GAO, Washington, DC.
General Accounting Office (1996) *Financial Derivatives: Actions Taken or Proposed Since May 1994*. GAO, Washington, DC.
Giacomini, E. and I. Komunjer (2002) 'Evaluation and combination of conditional quantile forecasts.' Mimeo. Department of Economics, UCSD.
Giannopoloulos, K. (2000) 'Measuring volatility.' Pp. 42–74 in M. Lore and L. Borodovsky (eds) *The Professional's Handbook of Financial Risk Management*. Oxford: Butterworth-Heinemann.
Giannopoloulos, K. and R. Tunaru (2004) 'Coherent risk measures under the FHS.' *Journal of Banking and Finance.* Forthcoming.
Gibson, M. S. (2001) 'Incorporating event risk into value-at-risk.' Mimeo. Federal Reserve Board.
Giot, P. and J. Grammig (2003) 'How large is liquidity risk in an automated auction market?' Mimeo. University of Namur and University of Tübingen.
Giot, P. and S. Laurent (2003) 'Value-at-risk for long and short trading positions.' *Journal of Applied Econometrics* 18: 641–664.
Glasserman, P., P. Heidelberger and P. Shahabuddin (1999a) 'Importance sampling and stratification for value-at-risk.' Pp. 7–24 in Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo and A. S. Weigend (eds) *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance)*. Cambridge, MA: MIT Press.
Glasserman, P., P. Heidelberger and P. Shahabuddin (1999b) 'Stratification issues in estimating value-at-risk.' Pp. 351–358 in P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W. Evans (eds) *Proceedings of the 1999 Winter Simulation Conference*. Piscataway, NJ: IEEE Computer Society Press.
Glasserman, P., P. Heidelberger and P. Shahabuddin (2000a) 'Portfolio value-at-risk with heavy-tailed risk factors.' Columbia Business School *Paine Webber Working Papers in Money, Economics and Finance PW-00-06*.
Glasserman, P., P. Heidelberger and P. Shahabuddin (2000b) 'Variance reduction techniques for estimating value-at-risk.' *Management Science* 46: 1349–1364.
Glosten, L., R. Jagannathan and D. Runkle (1993) 'On the relationship between the expected value and the volatility of the nominal excess return on stocks.' *Journal of Finance* 48: 1779–1801.
Golub, B. W. and L. M. Tilman (1997) 'Measuring yield curve risk using principal components analysis, value at risk, and key rate durations.' *Journal of Portfolio Management* 23 (Summer): 72–84.
Gottschling, A., C. Haefke and H. White (1999) 'Closed form integration of artificial neural networks with some applications to finance.' Mimeo. Deutsche Bank, Frankfurt, and UC-San Diego.
Granger, C. W. J. and C.-Y. Sin (2002) 'Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk.' *Journal of Forecasting* 19: 277–298.
Greenspan, A. (2000) 'Banking evolution.' Remarks to the 36th annual conference on bank structure and competition of the Federal Reserve Bank of Chicago, May 4, 2000.
Grootveld, H. and W. G. Hallerbach (2004) 'Upgrading value-at-risk from diagnostic metric to decision variable: a wise thing to do?' Pp. 33–50 in G. Szegö (ed.) *Risk Measures for the 21st Century*. Chichester: John Wiley and Sons, Ltd.
Group of Thirty (1993) *Derivatives: Practices and Principles*. New York: Group of Thirty.
Guldimann, T. (2000) 'The story of RiskMetrics.' *Risk* 13 (January): 56–58.
Gumbel, E. J. (1958) *Statistics of Extremes*. New York: Columbia University Press.
Hallerbach, W. G. (1999) 'Decomposing portfolio value-at-risk: a general analysis.' Mimeo. Erasmus University.
Hauksson, H. A., M. Dacorogna, T. Domenig, U. Müller and G. Samorodnitsky (2001) 'Multivariate extremes, aggregation and risk estimation.' *Quantitative Finance* 1: 79–95.
Heath, D., R. A. Jarrow and A. Morton (1990a) 'Bond pricing and the term structure of interest rates: a discrete time approximation.' *Journal of Financial and Quantitative Analysis* 25: 419–440.
Heath, D., R. A. Jarrow and A. Morton (1990b) 'Contingent claim valuation with a random evolution of interest rates.' *Review of Futures Markets* 9: 54–76.
Heath, D., R. A. Jarrow and A. Morton (1992) 'Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation.' *Econometrica* 60: 77–105.
Hendricks, D. (1996) 'Evaluation of value-at-risk models using historical data.' Federal Reserve Bank of New York *Economic Policy Review* 2 (April): 39–70.

Henrard, M. (2000) 'Comparison of cashflow maps for value-at-risk.' *Journal of Risk* 3 (Fall): 57–71.
Hereford, N. and G. Shuetrim (2000) 'Using stratified sampling methods to improve percentile estimates in the context of risk measurement.' Australian Prudential Regulation Authority *Working Paper 5*, June.
Hill, B. M. (1975) 'A simple general approach to inference about the tail of a distribution.' *Annals of Statistics* 35: 1163–1173.
Hisata, Y. and Y. Yamai (2000) 'Research toward the practical application of liquidity risk evaluation models.' Bank of Japan *Monetary and Economic Studies* (December): 83–127.
Ho, L.-C., P. Burridge, J. Cadle and M. Theobald (2000) 'Value-at-risk: applying the extreme value approach to Asian markets in the recent financial turmoil.' *Pacific-Basin Finance Journal* 8: 249–275.
Ho, T. (1992) 'Key rate durations: measures of interest rate risks.' *Journal of Fixed Income* 2: 29–44.
Ho, T., M. Chen and F. Eng (1996) 'VaR analytics: portfolio structure, key rate convexities and VaR betas.' *Journal of Portfolio Management* 23 (Fall): 89–98.
Holton, G. A. (2003) *Value-at-Risk: Theory and Practice*. San Diego: Academic Press.
Hoppe, R. (1998) 'VaR and the unreal world.' *Risk* 11 (July): 45–50.
Hosking, J. R. M., J. R. Wallis and E. F. Wood (1985) 'Estimation of the generalized extreme-value distribution by the method of probability-weighted moments.' *Technometrics* 27: 251–261.
Hua, P. and P. Wilmott (1997) 'Crash courses.' *Risk* 10 (June): 64–67.
Hull, J. (2003) *Options, Futures and Other Derivatives*. 5th edition. Upper Saddle River, NJ: Prentice Hall. Chapter 18.
Hull, J. and A. White (1987) 'The pricing of options on assets with stochastic volatilities.' *Journal of Finance* 42: 281–300.
Hull, J. and A. White (1998a) 'Value at risk when daily changes in market variables are not normally distributed.' *Journal of Derivatives* 5 (Spring): 9–19.
Hull, J. and A. White (1998b) 'Incorporating volatility updating into the historical simulation method for value-at-risk.' *Journal of Risk* 1 (Fall): 5–19.
Huschens, S. (1997) 'Confidence intervals for the value-at-risk.' Technische Universität Dresden Fakultät für Wirtschaftswissenschaften Dresdner Beiträge zu Quantitativen Verfahren Nr. 9/97.
Illinski, K. (2001) *The Physics of Finance: Gauge Modeling in Non-equilibrium Pricing*. Chichester: John Wiley and Sons, Ltd.
Ingersoll, J. E., Jr (1987) *Theory of Financial Decision Making*. Totowa, NJ: Rowman and Littlefield.
Inui, K. and M. Kijima (2003) 'On the significance of expected shortfall as a coherent risk measure.' Mimeo. Graduate School of Economics, Kyoto University.
Jäckel, P. (2002) *Monte Carlo Methods in Finance*. Chichester: John Wiley and Sons, Ltd.
Jahel, L. El, W. Perraudin and P. Sellin (1999) 'Value at risk for derivatives.' *Journal of Derivatives* 6: 7–26.
Jakobsen, S. (1996) 'Measuring value-at-risk for mortgage-backed securities.' Pp. 185–208 in F. Bruni, D. E. Fair and R. O'Brien (eds) *Risk Management in Volatile Financial Markets*. Dordrecht: Kluwer Academic Publishers.
James, J. and N. Webber (2000) *Interest Rate Modelling*. Chichester: John Wiley and Sons, Ltd.
Jamshidian, F. and Y. Zhu (1996) 'Scenario simulation model for risk management.' *Capital Market Strategies* 12 (December): 26–30.
Jamshidian, F. and Y. Zhu (1997) 'Scenario simulation: theory and methodology.' *Finance and Stochastics* 1: 43–67.
Jarque, C. M. and A. K. Bera (1987) 'A test for normality of observations and regression residuals.' *International Statistical Review* 55: 163–172.
Jarrow, R. A. and A. Subramanian (1997) 'Mopping up liquidity.' *Risk* 10 (December): 170–173.
Jarrow, R. A. and S. M. Turnbull (1995) 'Pricing options on derivative securities subject to credit risk.' *Journal of Finance* 50: 53–85.
Jarrow, R. A. and S. M. Turnbull (1998) 'The intersection of market and credit risks.' Working paper. Toronto: CIBC.
Jaschke, S. and U. Küchler (2000) 'Coherent risk measures, valuation bounds, and (μ, ρ)-portfolio optimization.' Mimeo. Weierstrass-Institut für Angewandte Analysis und Stochastik, Berlin and Humboldt University of Berlin.
Johnson, N.L. (1949) 'Systems of frequency curves generated by methods of translation.' *Biometrika* 36: 149–176.
Jones, B. L. and R. Zitikis (2003) 'Empirical estimation of risk measures and related quantities.' *North American Actuarial Journal* 7 (October): 44–54.

Jones, M. C. and M. J. Faddy (2003) 'A skew extension of the *t*-distribution, with applications.' *Journal of the Royal Statistical Society* B 65, Part 1: 159–174.
Jorion, P. (1997) 'In defense of VaR.' *Derivatives Strategy* 2 (April): 20–23.
Jorion, P. (2002) 'How informative are value-at-risk disclosures?' *Accounting Review* 77: 911–931.
Ju, X. and N. D. Pearson (1999) 'Using value-at-risk to control risk taking: how wrong can you be?' *Journal of Risk* 1 (2): 5–36.
Kato, T. and T. Yoshiba (2000) 'Model risk and its control.' Bank of Japan Institute for Monetary and Economic Studies *Discussion Paper No. 2000-E-15*.
Kearns, P. and A. R. Pagan (1997) 'Estimating the density tail index for financial time series.' *Review of Economics and Statistics* 79: 171–175.
Kendall, M. and A. Stuart (1972) *The Advanced Theory of Statistics, Vol. 1: Distribution Theory*. 4th edition. London: Charles Griffin and Co. Ltd.
Kendall, M. G. and A. Stuart (1973) *The Advanced Theory of Statistics. Volume 2: Inference and Relationship.* 3th edition. Griffin: London.
Kennedy, W. J., Jr. and J. E. Gentle (1980) *Statistical Computing*. New York and Basel: Marcel Dekker, Inc.
Kloek, T. (1990) 'Principal components.' Pp. 204–207 in J. Eatwell, M. Milgate and P. Newman (eds) *The New Palgrave Time Series and Statistics*. London and Basingstoke: Macmillan.
Koenker, R. and G. Bassett (1978) 'Quantile regression.' *Econometrica* 46: 33–50.
Kolmogorov, A. N. (1933) 'Sulla determinazione empirica di una legge di distribuzione.' *Giornale dell' Istituto Italiano degli Attuari* 4: 83–91; N. V. Smirnov (1939) 'Estimate of deviation between empirical distribution functions in two independent samples.' (in Russian) *Moscow University Bulletin* 2(2): 3–16.
Krakovsky, A. (1999) 'Pricing liquidity into derivatives.' *Risk* 12 (No. 12): 65–67.
Kreinin, A. and A. Levin (2000) 'Robust Monte Carlo simulation for approximate covariance matrices and VaR analyses.' Pp. 160–172 in S. Uryasev (ed.) *Probabilistic Constrained Optimization: Methodology and Applications*. Dordrecht: Kluwer Academic Publishers.
Kreyszig, E. (1999) *Advanced Engineering Mathematics*. 8th edition. New York: John Wiley and Sons, Ltd.
Kuiper, N. H. (1960) 'Tests concerning random points on a circle.' *Proc. Koninkl. Neder. Akad. Van Wetenschappen A* 63: 38–47.
Kupiec, P. (1995) 'Techniques for verifying the accuracy of risk management models.' *Journal of Derivatives* 3: 73–84.
Kupiec, P. H. (1999) 'Stress testing in a value at risk framework.' *Journal of Derivatives* 6: 7–24.
Kuruc, A. and B. Lee (1998) 'How to trim your hedge.' *Risk* 11 (December): 46–49.
Lagnado, R., G. Delianedis and S. Tikhonov (2000) 'Monte Carlo simulation of non-normal processes.' MKIRisk Discussion paper. MKIRisk, London.
Lakhany, A. and H. Mausser (2000) 'Estimating the parameters of the generalized lambda distribution.' *Algo Research Quarterly* 3 (December): 47–58.
Landsman, Z. and E. A. Valdez (2003) 'Tail conditional expectations for elliptical distributions.' *North American Actuarial Journal*, 7: 55–71.
Laubsch, A. (1996) 'Estimating index tracking error for equity portfolios.' *RiskMetrics™ Monitor*, Second Quarter 1996: 34–41.
Lawrence, C. and G. Robinson (1995a) 'How safe is RiskMetrics?' *Risk* 8 (January): 26–29.
Lawrence, C. and G. Robinson (1995b) 'Liquid measures.' *Risk* 8 (July): 52–55.
Lawrence, C. and G. Robinson (1995c) 'Value at risk: addressing liquidity and volatility issues.' *Capital Market Strategies* 9 (November): 24–28.
Le Saout, E. (2002) 'Incorporating liquidity risk in VaR models.' Mimeo. Paris I University – CREFIB & CREREG.
Lee, Y.-S. and T.-K. Lin (1992) 'Higher order Cornish–Fisher expansion.' *Applied Statistics* 41: 233–240.
Lhabitant, F.-S. (2000) 'Coping with model risk.' Pp. 415–440 in M. Lore and L. Borodovsky (eds) *The Professional's Handbook of Financial Risk Management*. Oxford: Butterworth-Heinemann.
Lilliefors, H. W. (1967) 'On the Kolmogorov–Smirnov test for normality with mean and variance unknown.' *Journal of the American Statistical Association* 66: 534–544.
Lin, C. H. and C.-C. C. Chien (2002) 'Incorporating a GED model into the historical simulation method for value-at-risk.' National Kaohsiung First University of Science and Technology.

Linsmeier, T. J. and N. D. Pearson (1996) 'Risk measurement: an introduction to value at risk.' Mimeo. University of Illinois at Urbana-Champaign.
Litterman, R. (1996) 'Hot spots™ and hedges.' *Journal of Portfolio Management Special Issue*: 52–75.
Longerstaey, J. and P. Zangari (1995) 'A transparent tool.' *Risk* 8 (January): 30–32.
Longin, F. (1996) 'The asymptotic distribution of extreme stock market returns.' *Journal of Business* 63: 383–408.
Longin, F. (1999) 'From value at risk to stress testing: the extreme value approach.' *Journal of Banking and Finance* 24: 1097–1130.
Longstaff, F. A. and E. S. Schwartz (2001) 'Valuing American options by simulation: a simple least-squares approach.' *Review of Financial Studies* 14: 113–147.
Lopez, J. A. (1998) 'Regulatory evaluation of value-at-risk models.' Federal Reserve Bank of New York *Economic Policy Review* (October): 119–124.
Lopez, J. A. (1999) 'Regulatory evaluation of value-at-risk models.' *Journal of Risk* 1: 37–64.
Lopez, J. A. and C. A. Walter (2001) 'Evaluating covariance matrix forecasts in a value-at-risk framework.' Mimeo. Federal Reserve Bank of San Franscisco and Credit Suisse Group, Zürich.
Lucas, A. (2000) 'A note on optimal estimation from a risk management perspective under possibly mis-specified tail behavior.' *Journal of Business and Economic Statistics* 18: 31–39.
Lucas, A. and P. Klaassen (1998) 'Extreme returns, downside risk, and optimal asset allocation.' *Journal of Portfolio Management* 24 (Fall): 71–79.
Malevergne, Y. and D. Sornette (2001) 'General framework for a portfolio theory with non-Gaussian risks and non-linear correlations.' Mimeo. Laboratoire de Physique de la Matière Condensée, Université de Nice-Sophia Antipolis; and Institute of Geophysics and Planetary Physics and Department of Earth and Space Science, UCLA. March.
Mandelbrot, B. (1963) 'The variation of certain speculative prices.' *Journal of Business* 36: 394–419.
Mantegna, R. N. and H. E. Stanley (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press.
Markowitz, H. M. (1952) 'Portfolio selection.' *Journal of Finance* 7: 77–91.
Markowitz, H. M. (1959) *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley and Sons, Ltd.
Markowitz, H. M. (1992) 'Mean-variance analysis.' Pp. 683–685 in P. Newman, M. Milgate and J. Eatwell (eds) *The New Palgrave Dictionary of Money and Finance. Volume 2*. New York: Stockton Press.
Marshall, C. and M. Siegel (1997) 'Value at risk: implementing a risk measurement standard.' *Journal of Derivatives* 4: 91–110.
Mausser, H. (2001) 'Calculating quantile-based risk analytics with L-estimators.' *Algo Research Quarterly* 4 (December): 33–47.
McNeil, A. J. (1998) 'Calculating quantile risk measures for financial return series using extreme value theory.' Mimeo. ETHZ Zentrum, Zürich.
McNeil, A. J. and R. Frey (2000) 'Estimation of tail-related risk for heteroscedastic financial time series: an extreme value approach.' *Journal of Empirical Finance* 7: 271–300.
McNeil, A. J. and T. Saladin (1997) 'The peaks over thresholds model for estimating high quantiles of loss data.' Mimeo. ETHZ Zentrum, Zürich. April.
Milevsky, M. (1998) 'Optimal asset allocation towards the end of the life cycle: to annuitize or not to annuitize?' *Journal of Risk and Insurance* 65: 401–426.
Mina, J. (2001) 'Calculating VaR through quadratic approximations.' *Journal of Risk Finance* 2 (Winter): 49–55.
Miranda, M. S. and P. L. Fackler (2002) *Applied Computational Economics and Finance*. Cambridge, MA: MIT Press.
Mittnik, S., M. S. Paolella and S. T. Rachev (2000) 'Diagnosing and treating the fat tails in financial returns data.' *Journal of Empirical Finance* 7: 389–416.
Mittnik, S., S. T. Rachev and M. S. Paolella (1998) 'Stable Paretian modelling in finance: some empirical and theoretical aspects.' Pp. 79–110 in R. J. Adler, R. E. Feldman and M. S. Taqqu (eds) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston: Birkhaüser.
Modigliani, F. and M. H. Miller (1958) 'The cost of capital, corporation finance, and the theory of investment.' *American Economic Review 48*, June: 261–297.
Moosa, I. A. and B. Bollen (2001) 'A benchmark for measuring bias in estimated daily value at risk.' Forthcoming, *International Review of Financial Analysis*.

Moosa, I. A. and J. J. Knight (2001) 'Firm characteristics and value at risk analysis: a survey of Australia's public shareholding companies.' Mimeo. La Trobe University.
Morgan Guaranty Trust Company (1996) *RiskMetrics*™*– Technical Document*. 4th edition. New York: Morgan Guaranty Trust Company.
Mori, A., M. Ohsawa and T. Shimizu (1996) 'A framework for more effective stress testing.' Bank of Japan Institute for Monetary and Economic Studies *Discussion Paper 96-E-2*.
Müller, U. A., M. M. Dacorogna and O. V. Pictet (1998) 'Heavy tails in high-frequency financial data.' Pp. 55–77 in R. J. Adler, R. E. Feldman and M. S. Taqqu (eds) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston: Birkhaüser.
Muranaga, J. and M. Ohsawa (1996) 'Measurement of liquidity risk in the context of market risk calculation.' Research paper, Bank of Japan: Institute for Monetary and Economic Studies.
Nelson, D. B. (1990) 'Conditional heteroscedasticity in asset returns: a new approach.' *Econometrica* 59: 347–370.
Nelson, R. B. (1999) *An Introduction to Copulas*. Berlin: Springer Verlag.
Niederreiter, H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM.
Niffikeer, C. I., R. D. Hewins and R. B. Flavell (2000) 'A synthetic factor approach to the estimation of value-at-risk of a portfolio of interest rate swaps.' *Journal of Banking and Finance* 24: 1903–1932.
Panning, W. H. (1999) 'The strategic uses of value at risk: long-term capital management for property/casualty insurers.' *North American Actuarial Journal* 3(2): 84–105.
Pant, V. and W. Chang (2001) 'An empirical comparison of methods for incorporating fat tails into value-at-risk models.' *Journal of Risk* 3: 99–119.
Paul-Choudhury, S. (1997) 'This year's model.' *Risk* 10 (May): 18–23.
Persaud, A. (2000) 'The liquidity puzzle.' *Risk* 13 (June): 64–66.
Phelan, M. J. (1997) 'Probability and statistics applied to the practice of financial risk management: the case of J.P. Morgan's RiskMetrics™.' *Journal of Financial Services Research* 12(2–3): 175–200.
Phoa, W. (2000) 'Yield curve risk factors: domestic and global contexts.' Pp. 155–184 in M. Lore and L. Borodovsky (eds) *The Professional's Handbook of Financial Risk Management*. Oxford: Butterworth-Heinemann.
Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1992) *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edition. Cambridge: Cambridge University Press.
Prinzler, R. (1999) 'Reliability of neural network based value-at-risk estimates.' Mimeo. Technical University of Dresden.
Pritsker, M. (1997) 'Evaluating value at risk methodologies: accuracy versus computational time.' *Journal of Financial Services Research* 12: 201–242.
Pritsker, M. (2001) 'The hidden dangers of historical simulation.' Mimeo. Board of Governors of the Federal Reserve System. April.
Ratchev, S. and S. Mittnik (2000) *Stable Paretian Models in Finance*. New York: John Wiley and Sons, Ltd.
Rebonato, R. (1998) *Interest-Rate Option Models*. 2nd edition. Chichester: John Wiley and Sons, Ltd.
Reiss, R.-D. (1989) *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*. Berlin: Springer Verlag.
Reiss, R.-D. and M. Thomas (1997) *Statistical Analysis of Extreme Values from Insurance, Finance, Hydrology and Other Fields*. Basel: Birkhäuser Verlag.
Rockafellar, R. T. and S. Uryasev (2002) 'Conditional value-at-risk for general loss distributions.' *Journal of Banking and Finance* 26: 1443–1471.
Ross, S. A. (1976) 'The arbitrage theory of asset pricing.' *Journal of Economic Theory* 13: 341–360.
Rouvinez, C. (1997) 'Going Greek with VaR.' *Risk* 10 (February): 57–65.
Roy, A. D. (1952) 'Safety first and the holding of assets.' *Econometrica* 20: 431–449.
Scaillet, O. (2000a) 'Nonparametric estimation of copulas for time series.' Mimeo. IRES, Catholic University of Louvain.
Scaillet, O. (2000b) 'Nonparametric estimation and sensitivity analysis of expected shortfall.' Mimeo. IRES, Catholic University of Louvain.
Scaillet, O. (2000c) 'Nonparametric estimation of conditional expected shortfall.' Mimeo. IRES, Catholic University of Louvain.
Schachter, B. (1995) 'Comments on "Taylor, Black and Scholes: Series approximations and risk management pitfalls" by Arturo Estrella.' Mimeo. Office of the Comptroller of the Currency.

Schachter, B. (1998) 'The value of stress testing in market risk management.' Pp. 5F-0–5F-11 in T. Haight (ed.) *Derivatives Risk Management Service*. Arlington, VA: A. S. Pratt & Sons.

Schachter, B. (2000) 'Stringent stress tests.' *Risk* 13 (No. 12, *Special Report on Enterprise-Wide Risk Management*): S22–S24.

Sentana, E. (2001) 'Mean-variance allocation with a value at risk constraint.' Mimeo. CEMFI.

Shapiro, S. S. and M. B. Wilk (1965) 'An analysis of variance test for normality (complete samples).' *Biometrika*, 52 (3 and 4): 591–611.

Shaw, J. (1997) 'Beyond VaR and stress testing.' Pp. 221–224 in *VAR – Understanding and Applying Value at Risk*. London: KPMG/Risk Publications.

Shimko, D. (1996) 'VaR for corporates.' *Risk* 9 (June): 28–29.

Shimko, D., B. Humpheys and V. Pant (1998) 'Hysterical simulation.' *Risk* 11 (June): 47.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London and New York: Chapman and Hall.

Singer, R. (1997) 'To VaR, a sister.' *Risk* 10 (August): 86–87.

Singh, M. K. (1997) 'Value at risk using principal components analysis.' *Journal of Portfolio Analysis*, Fall: 101–112.

Siu, T. K., H. Tong and H. Yang (2001) 'On Bayesian value at risk: from linear to non-linear portfolios.' Mimeo. National University of Singapore, University of Hong Kong, and London School of Economics.

Sklar, A. (1959) 'Fonctions de repartition à n dimensions et leur merges.' *Publ. Inst. Stat. Univ. Paris* 8: 229–231.

Smirnov, N. V. (1948) 'Table for estimating goodness of fit of empirical distributions.' *The Annals of Mathematical Statistics* 19: 279–281.

Stevens, M. A. (1974) 'EDF statistics for goodness of fit and some comparisons.' *Journal of the American Statistical Association* 69: 730–737.

Studer, G. (1999) 'Market risk computation for nonlinear portfolios.' *Journal of Risk* 1 (4): 33–53.

Studer, G. and H.-J. Lüthi (1997) 'Analyzing nonlinear portfolios with quadratic maximum loss.' *Net Exposure* 1 (October). Available at http://www.netexposure.co.uk/regd/issues/1/studer/index.html.

Taleb, N. (1997a) 'The world according to Nassim Taleb.' *Derivatives Strategy* 2 (December/January): 37–40.

Taleb, N. (1997b) 'Against VaR.' *Derivatives Strategy* 2 (April): 21–26.

Tasche, D. (2000) 'Risk contributions and performance measurement.' Working paper. Munich University of Technology.

Tasche, T. and L. Tibiletti (2003) 'A shortcut to sign incremental value-at-risk for risk allocation.' *Journal of Risk Finance* 4 (Winter): 43–46.

Tay, A. S. and K. F. Wallis (2000) 'Density forecasting: a survey.' *Journal of Forecasting* 19: 275–254.

Taylor, J. W. (2000) 'A quantile regression neural network approach to estimating the conditional density of multiperiod returns.' *Journal of Forecasting* 19: 299–311.

Thisted, R. A. (1988) *Elements of Statistical Computing: Numerical Computations*. New York and London: Chapman and Hall.

Tilley, J. A. (1993) 'Valuing American options in a path simulation model.' *Transactions of the Society of Actuaries* 45: 83–104.

Tilman, L. M. and P. Brusilovskiy (2001) 'Measuring predictive accuracy of value-at-risk models: issues, paradigms, and directions.' *Journal of Risk Finance* 2, No. 3 (Spring 2001): 83–91.

Tsay, R. S. (2002) *Analysis of Financial Time Series.* (New York: John Wiley and Sons, Ltd, Chapter 8.7).

Uryasev, S. (2000) 'Conditional value-at-risk: optimization algorithms and applications.' *Financial Engineering News* 14 (February): 1–5.

Venkataraman, S. (1997) 'Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques.' Federal Reserve Bank of Chicago *Economic Perspectives* (March/April): 3–13.

Vlaar, P. J. G. (2000) 'Value at risk models for Dutch bond portfolios.' *Journal of Banking and Finance* 24: 1131–1154.

Wakeman, L. (1998) 'Credit enhancement.' Pp. 255–275 in C. Alexander (ed.) *Risk Analysis and Management. Volume 1: Measuring and Modelling Financial Risk*. Chichester: John Wiley and Sons, Ltd.

Wang, C., G. Li and H. Wan (2001) 'Improving the estimation of value at risk using Markov chain Monte Carlo simulation.' Mimeo. Management School, Tianjin University.

Wang, S. S. (1996) 'Premium calculation by transforming the layer premium density.' *ASTIN Bulletin* 26: 71–92.
Wang, S. S. (2000) 'A class of distortion operators for pricing financial and insurance risks.' *Journal of Risk and Insurance* 67: 15–36.
Wang, S. S. (2002) 'A risk measure that goes beyond coherence.' Mimeo. SCOR Reinsurance Co., Itasca IL.
Wang, S. S., V. R. Young and H. H. Panjer (1997) 'Axiomatic characterization of insurance prices.' *Insurance: Mathematics and Economics* 21: 173–183.
Wang, T. (1999) 'A class of dynamic risk measures.' Faculty of Commerce and Business Administration, University of British Columbia.
Wee, L.-S. and J. Lee (1999) 'Integrating stress testing with risk management.' *Bank Accounting and Finance* (Spring): 7–19.
Wiener, Z. (1999) 'Comment on "Non-linear value at risk".' *European Finance Review* 2: 189–193.
Wilmott, P. (2000) *Paul Wilmott on Quantitative Finance. Volumes 1 and 2*. Chichester: John Wiley and Sons, Ltd.
Wilson, T. C. (1994a) 'Debunking the myths.' *Risk* 7 (April): 67–72.
Wilson, T. C. (1994b) 'Plugging the gap.' *Risk* 7 (October): 74–80.
Wilson, T. C. (1996) 'Calculating risk capital.' Pp. 193–232 in C. Alexander (ed.) *The Handbook of Risk Analysis and Management*. Chichester: John Wiley and Sons, Ltd.
Woods, M., K. Dowd and C. Humphrey (2004) 'The value of value-at-risk reporting: a critical evaluation of value-at-risk disclosures in the banking sector.' Mimeo. Nottingham University Business School and University of Manchester.
Yamai, Y. and T. Yoshiba (2001a) 'On the validity of value-at-risk: comparative analyses with expected shortfall.' Bank of Japan Institute of Monetary and Economic Studies *Discussion Paper No. 2001-E-4*.
Yamai, Y. and T. Yoshiba (2001b) 'Comparative analyses of expected shortfall and VaR: their estimation error, decomposition, and optimisation.' Bank of Japan Institute of Monetary and Economic Studies *Discussion Paper No. 2001-E-12*.
Yoshiba, T. and Y. Yamai (2001) 'Comparative analyses of expected shortfall and VaR (2): expected utility maximisation and tail risk.' Bank of Japan Institute of Monetary and Economic Studies *Discussion Paper No. 2001-E-14*.
Zangari, P. (1996a) 'A VaR methodology for portfolios that include options.' *RiskMetrics*™ *Monitor*, First Quarter: 4–12.
Zangari, P. (1996b) 'An improved methodology for measuring VaR.' *RiskMetrics*™ *Monitor*, Second Quarter: 7–25.
Zangari, P. (1996c) 'How accurate is the delta-gamma methodology?' *RiskMetrics*™ *Monitor*, Third Quarter: 12–29.
Zangari, P. (1996d) 'When is non-normality a problem? The case of 15 time series from emerging markets.' *RiskMetrics*™ *Monitor*, Fourth Quarter: 20–32.
Zangari, P. (1997) 'What risk managers should know about mean reversion and jumps in prices.' *RiskMetrics*™ *Monitor*, Fourth Quarter: 12–41.
Zangari, P. (1998) 'Exploratory stress-scenario analysis with applications to EMU.' *RiskMetrics*™ *Monitor Special Edition*: 30–53.

Index