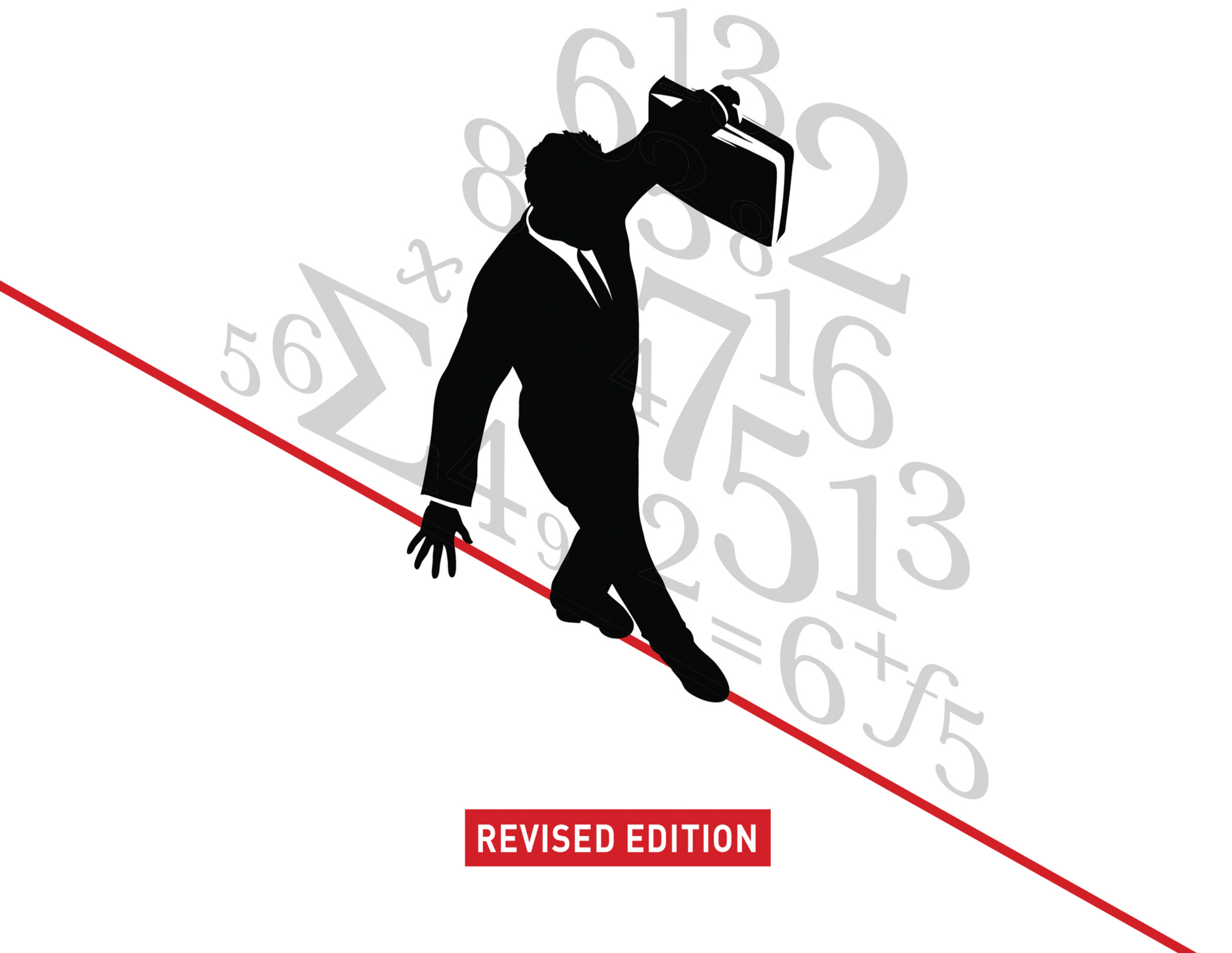
PRINCETON SERIES IN FINANCE

QUANTITATIVE RISK MANAGEMENT

CONCEPTS, TECHNIQUES AND TOOLS

Alexander J. McNeil, Rüdiger Frey and Paul Embrechts



REVISED EDITION

Quantitative Risk Management

Quantitative Risk Management: Concepts, Techniques and Tools
IS A PART OF THE
PRINCETON SERIES IN FINANCE

Finance as a discipline has been growing rapidly. The numbers of researchers in academy and industry, of students, of methods and models have all proliferated in the past decade or so. This growth and diversity manifests itself in the emerging cross-disciplinary as well as cross-national mix of scholarship now driving the field of finance forward. The intellectual roots of modern finance, as well as the branches, will be represented in the Princeton Series in Finance.

Titles in this series will be scholarly and professional books, intended to be read by a mixed audience of economists, mathematicians, operations research scientists, financial engineers, and other investment professionals. The goal is to provide the finest cross-disciplinary work in all areas of finance by widely recognized researchers in the prime of their creative careers.

OTHER BOOKS IN THIS SERIES

Financial Econometrics: Problems, Models, and Methods by Christian Gourieroux and Joann Jasiak

Credit Risk: Pricing, Measurement, and Management by Darrell Duffie and Kenneth J. Singleton

Microfoundations of Financial Economics: An Introduction to General Equilibrium Asset Pricing by Yvan Lengwiler

Credit Risk Modeling: Theory and Applications by David Lando

Quantitative Risk Management: Concepts, Techniques and Tools, revised edition, by Alexander J. McNeil, Rüdiger Frey and Paul Embrechts

Quantitative Risk Management

Concepts, Techniques and Tools

Revised Edition

Alexander J. McNeil

Rüdiger Frey

Paul Embrechts

Princeton University Press
Princeton and Oxford

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

press.princeton.edu

ISBN 978-0-691-16627-8

Library of Congress Control Number: 2015934589

British Library Cataloguing-in-Publication Data is available

This book has been composed in Times and typeset
by T&T Productions Ltd, London

Printed on acid-free paper ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Janine, Alexander and Calliope
Alexander

Für Catharina, Sebastian und Michaela
Rüdiger

Voor Levi, Mila, Ben en Marlon
Paul

Contents

Preface

Why have we written this book? In recent decades the field of financial risk management has developed rapidly in response to both the increasing complexity of financial instruments and markets and the increasing regulation of the financial services industry. This book is devoted specifically to *quantitative* modelling issues arising in this field. As a result of our own discussions and joint projects with industry professionals and regulators over a number of years, we felt there was a need for a textbook treatment of quantitative risk management (QRM) at a technical yet accessible level, aimed at both industry participants and students seeking an entrance to the area.

We have tried to bring together a body of methodology that we consider to be core material for any course on the subject. This material and its mode of presentation represent the blending of our own views, which come from the perspectives of financial mathematics, insurance mathematics and statistics. We feel that a book combining these viewpoints fills a gap in the existing literature and emphasises the fact that there is a need for quantitative risk managers in banks, insurance companies and beyond to have broad, interdisciplinary skills.

What is new in this second edition? The second edition of this book has been extensively revised and expanded to reflect the continuing development of QRM methodology since the 2005 first edition. This period included the 2007–9 financial crisis, during which much of the methodology was severely tested. While we have added to the detail, we are encouraged that we have not had to revise the main messages of the first edition in the light of the crisis. In fact, many of those messages—the importance of extremes and extremal dependence, systematic risk and the model risk inherent in portfolio credit models—proved to be central issues in the crisis.

Whereas the first edition had a Basel and banking emphasis, we have added more material relevant to Solvency II and insurance in the second edition. Moreover, the methodological chapters now start at the natural starting point: namely, a discussion of the balance sheets and business models of a bank and an insurer.

This edition contains an extended treatment of credit risk in four chapters, including new material on portfolio credit derivatives and counterparty credit risk. There is a new market-risk chapter, bringing together more detail on mapping portfolios to market-risk factors and applying and backtesting statistical methods. We have also extended the treatment of the fundamental topics of risk measures and risk aggregation.

We have revised the structure of the book to facilitate teaching. The chapters are a little shorter than in the first edition, with more advanced or specialized material now placed in a series of "Special Topics" chapters at the end. The book is split into four parts: (I) An Introduction to Quantitative Risk Management, (II) Methodology, (III) Applications, (IV) Special Topics.

Who was this book written for? This book is primarily a textbook for courses on QRM aimed at advanced undergraduate or graduate students and professionals from the financial industry. *A knowledge of probability and statistics at least at the level of a first university course in a quantitative discipline and familiarity with undergraduate calculus and linear algebra are fundamental prerequisites*. Though not absolutely necessary, some prior exposure to finance, economics or insurance will be beneficial for a better understanding of some sections.

The book has a secondary function as a reference text for risk professionals interested in a clear and concise treatment of concepts and techniques that are used in practice. As such, we hope it will facilitate communication between regulators, end-users and academics.

A third audience for the book is the community of researchers that work in the area. Most chapters take the reader to the frontier of current, practically relevant research and contain extensive, annotated references that guide the reader through the vast literature.

Ways to use this book. The material in this book has been tested on many different audiences, including undergraduate and postgraduate students at ETH Zurich, the Universities of Zurich and Leipzig, Heriot-Watt University, the London School of Economics and the Vienna University of Economics and Business. It has also been used for professional training courses aimed at risk managers, actuaries, consultants and regulators. Based on this experience we can suggest a number of ways of using the book.

A taught course would generally combine material from Parts I, II and III, although the exact choice of material from Parts II and III would depend on the emphasis of the course. Chapters 2 and 3 from Part I would generally be core taught modules, whereas Chapter 1 might be prescribed as background reading material.

A general course on QRM could be based on a complete treatment of Parts I–III. This would require a minimum of two semesters, with 3–4 hours of taught courses per week for an introductory course and longer for a detailed treatment. A quantitative course on enterprise risk management for actuaries would follow a very similar selection, probably omitting material from Chapters 11 and 12, which contain Basel-specific details of portfolio credit risk modelling and an introduction to portfolio credit derivatives.

For a course on credit risk modelling, there is a lot of material to choose from. A comprehensive course spanning two semesters would include Part I (probably omitting Chapter 3), Chapters 6 and 7 from Part II, and Chapters 10–12 from Part III. Material on counterparty credit risk (Chapter 17) might also be included from Part IV.

A one-semester, specialized course on market risk could be based on Part I, Chapters 4–6 from Part II, and Chapter 9 from Part III. An introduction to risk management for financial econometricians could follow a similar selection but might cover all the chapters in Part II.

It is also possible to devise more specialized courses, such as a course on risk-measurement and aggregation concepts based on Chapters 2, 7 and 8. Moreover, material from various chapters could be used as interesting examples to enliven statistics courses on subjects like multivariate analysis, time-series analysis and generalized linear modelling. In Part IV there are a number of potential topics for seminars at postgraduate and PhD level.

What we have not covered. We have not been able to address all the topics that a reader might expect to find under the heading of QRM. Perhaps the most obvious omission is the lack of a section on the risk management of derivatives by hedging. Here we felt that the relevant techniques, and the financial mathematics required to understand them, are already well covered in a number of excellent textbooks. Other omissions include modelling techniques for price liquidity risk and models for systemic risk in national and global networks of financial firms, both of which have been areas of research since the 2007–9 crisis. Besides these larger areas, many smaller issues have been neglected for reasons of space but are mentioned with suggestions for further reading in the "Notes and Comments" sections, which should be considered as integral parts of the text.

Acknowledgements. The origins of this book date back to 1996, when A.M. and R.F. began postdoctoral studies in the group of P.E. at the Federal Institute of Technology (ETH) in Zurich. All three authors are grateful to ETH for providing the environment in which the project initially flourished. A.M. and R.F. thank Swiss Re and UBS, respectively, for providing the financial support for their postdoctoral positions. P.E. thanks the Swiss Finance Institute, which continues to provide support through a Senior SFI Professorship.

A.M. acknowledges the support of Heriot-Watt University, where he now works, and thanks the Isaac Newton Institute for Mathematical Sciences for a Visiting Fellowship during which he was able to put the finishing touches to the manuscript. R.F. has subsequently held positions at the Swiss Banking Institute of the University of Zurich, the University of Leipzig and the Vienna University of Economics and Business, and he is grateful to all these institutions for their support. P.E. thanks the London School of Economics, where he enjoyed numerous fruitful discussions with colleagues during his time as Centennial Professor of Finance, and the Oxford-Man Institute at the University of Oxford for hospitality during his visit as OMI Visiting Man Chair.

The Forschungsinstitut für Mathematik of ETH Zurich provided generous financial support throughout the project and facilitated meetings of all three authors in Zurich on many occasions. At a crucial juncture in early 2004 the Mathematisches Forschungsinstitut Oberwolfach was the venue for a memorable week of progress. We also acknowledge the invaluable contribution of RiskLab Zurich to the enterprise: the agenda for the book was strongly influenced by joint projects and discussions with the RiskLab sponsors UBS, Credit Suisse and Swiss Re. We have also benefited greatly from the NCCR FINRISK research programme in Switzerland, which funded doctoral and postdoctoral research on topics in the book.

We are indebted to numerous readers who have commented on various parts of the manuscript, and to colleagues in Zurich, Leipzig, Edinburgh, Vienna and beyond who have helped us in our understanding of QRM and the mathematics underlying it. These include Stefan Altner, Philippe Artzner, Jochen Backhaus, Guus Balkema, Michał Barski, Uta Beckmann, Reto Baumgartner, Wolfgang Breymann, Reto Bucher, Hans Bühlmann, Peter Bühlmann, Valérie Chavez-Demoulin, Dominik Colangelo, Marius Costeniuc, Enrico De Giorgi, Freddy Delbaen, Rosario Dell'Aquila, Stefano Demarta, Stefan Denzler, Alexandra Dias, Catherine Donnelly, Douglas Dwyer, Damir Filipovic, Tom Fischer, Gabriel Frahm, Hansjörg Furrer, Rajna Gibson, Kay Giesecke, Michael Gordy, Bernhard Hodler, Marius Hofert, Andrea Höing, Kurt Hornik, Friedrich Hubalek, Christoph Hummel, Edgars Jakobsons, Alessandro Juri, Roger Kaufmann, Philipp Keller, Erwan Koch, Hans Rudolf Künsch, Filip Lindskog, Hans-Jakob Lüthi, Natalia Markovich, Benoît Metayer, Andres Mora, Alfred Müller, Johanna Nešlehová, Monika Popp, Giovanni Puccetti, Lars Rösler, Wolfgang Runggaldier, David Saunders, Hanspeter Schmidli, Sylvia Schmidt, Thorsten Schmidt, Uwe Schmock, Philipp Schönbucher, Martin Schweizer, Torsten Steiger, Daniel Straumann, Dirk Tasche, Hideatsu Tsukahara, Laura Vana, Eduardo Vilela, Marcel Visser, Ruodu Wang, Jonathan Wendin and Mario Wüthrich. For their help in preparing the manuscript we thank Gabriele Baltes and Edgars Jakobsons. We should also, of course, thank the scores of (unnamed) students who took QRM lectures based on this book and contributed through their critical questions and remarks.

We thank the team at Princeton University Press for all their help in the production of this book, particularly our editors Richard Baggaley (first edition) and Hannah Paul (second edition). We are also grateful to anonymous referees who provided us with exemplary feedback, which has shaped this book for the better. Special thanks go to Sam Clark at T&T Productions Ltd, who took our uneven LaTeX code and turned it into a more polished book with remarkable speed and efficiency.

We are delighted that the first edition appeared in a Japanese translation and thank the translation team (Hideatsu Tsukahara, Shun Kobayashi, Ryozo Miura, Yoshinori Kawasaki, Hiroaki Yamauchi and Hidetoshi Nakagawa) for their work on this project, their valuable feedback on the book and their hospitality to A.M. at the Japanese book launch.

To our wives, Janine, Catharina and Gerda, and our families our sincerest debt of gratitude is due. Though driven to distraction no doubt by our long contemplation of risk, without obvious reward, their support was constant.

Further resources. Readers are encouraged to visit the book's homepage at

www.pupress.princeton.edu/titles/8056.html

and the site

www.qrmtutorial.org

where they will find supplementary resources for this book. We are particularly grateful to Marius Hofert, not only for his proofreading, but also for his help in developing slides, exercises and R libraries and scripts to illustrate many of the topics in the book.

Special abbreviations. A number of abbreviations for common terms in probability are used throughout the book; these include “rv” for random variable, “df” for distribution function, “iid” for independent and identically distributed and “se” for standard error.

Part I

An Introduction to Quantitative Risk Management

1
Risk in Perspective

In this chapter we provide a non-mathematical discussion of various issues that form the background to the rest of the book. In Section 1.1 we begin with the nature of risk itself and discuss how risk relates to randomness; in the financial context (which includes insurance) we summarize the main kinds of risks encountered and explain what it means to measure and manage such risks.

A brief history of financial risk management and the development of financial regulation is given in Section 1.2, while Section 1.3 contains a summary of the regulatory framework in the financial and insurance industries.

In Section 1.4 we take a step back and attempt to address the fundamental question of why we might want to measure and manage risk at all. Finally, in Section 1.5 we turn to quantitative risk management (QRM) explicitly and set out our own views concerning the nature of this discipline and the challenge it poses. This section in particular should give more insight into our choice of methodological topics in the rest of the book.

1.1 Risk

The *Concise Oxford English Dictionary* defines risk as "hazard, a chance of bad consequences, loss or exposure to mischance". In a discussion with students taking a course on financial risk management, ingredients that are typically discussed are events, decisions, consequences and uncertainty. It is mostly only the downside of risk that is mentioned, rarely a possible upside, i.e. the potential for a gain. While for many people risk has largely negative connotations, it may also represent an opportunity. Much of the financial industry would not exist were it not for the presence of financial risk and the opportunities afforded to companies that are able to create products and services that offer more financial certainty to their clients.

For financial risks no single one-sentence definition of risk is entirely satisfactory. Depending on context, one might arrive at notions such as "any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies" or, alternatively, "the quantifiable likelihood of loss or less-than-expected returns".

1.1.1 Risk and Randomness

Regardless of context, risk strongly relates to uncertainty, and hence to the notion of randomness. Randomness has eluded a clear, workable definition for many centuries;

it was not until 1933 that the Russian mathematician A. N. Kolmogorov gave an axiomatic definition of randomness and probability (see Kolmogorov 1933). This definition and its accompanying theory provide the language for the majority of the literature on risk, including this book.

Our reliance on probability may seem unsatisfactorily narrow to some. It bypasses several of the current debates on risk and uncertainty (Frank Knight), the writings on probabilistic thinking within economics (John Maynard Keynes), the unpredictability of unprecedented financial shocks, often referred to as Black Swans (Nassim Taleb), or even the more political expression of the known, the unknown and the unknowable (Donald Rumsfeld); see the Notes and Comments section for more explanation. Although these debates are interesting and important, at some point clear definitions and arguments are called for and this is where mathematics as a language enters. The formalism of Kolmogorov, while not the only possible approach, is a tried-and-tested framework for mathematical reasoning about risk.

In Kolmogorov's language a probabilistic model is described by a triplet $(\Omega, \mathcal{F}, P)$. An element ω of Ω represents a realization of an experiment, in economics often referred to as a state of nature. The statement "the probability that an event A occurs" is denoted (and in Kolmogorov's axiomatic system defined) as $P(A)$, where A is an element of $\mathcal{F}$, the set of all events. P denotes the probability measure. For the less mathematically trained reader it suffices to accept that Kolmogorov's system translates our intuition about randomness into a concise, axiomatic language and clear rules.

Consider the following examples: an investor who holds stock in a particular company; an insurance company that has sold an insurance policy; an individual who decides to convert a fixed-rate mortgage into a variable one. All of these situations have something important in common: the investor holds today an asset with an uncertain future value. This is very clear in the case of the stock. For the insurance company, the policy sold may or may not be triggered by the underlying event covered. In the case of a mortgage, our decision today to enter into this refinancing agreement will change (for better or for worse) the future repayments. So randomness plays a crucial role in the valuation of current products held by the investor, the insurance company and the home owner.

To model these situations a mathematician would now define the value of a risky position X to be a function on the probability space $(\Omega, \mathcal{F}, P)$; this function is called a *random variable*. We leave for the moment the range of X (i.e. its possible values) unspecified. Most of the modelling of a risky position X concerns its *distribution function* $F_X(x) = P(X \leqslant x)$: the probability that by the end of the period under consideration the value of the risk X is less than or equal to a given number x. Several risky positions would then be denoted by a random vector $(X_1, \dots, X_d)$, also written in bold face as $\boldsymbol{X}$; time can be introduced, leading to the notion of random (or so-called stochastic) processes, usually written $(\boldsymbol{X}_t)$. Throughout this book we will encounter many such processes, which serve as essential building blocks in the mathematical description of risk.

We therefore expect the reader to be at ease with basic notation, terminology and results from elementary *probability and statistics*, the branch of mathematics dealing with *stochastic* models and their application to the real world. The word "stochastic" is derived from the Greek "stochazesthai", the art of guessing, or "stochastikos", meaning skilled at aiming ("stochos" being a target). In discussing stochastic methods for risk management we hope to emphasize the skill aspect rather than the guesswork.

1.1.2 Financial Risk

In this book we discuss risk in the context of finance and insurance (although many of the tools introduced are applicable well beyond this context). We start by giving a brief overview of the main risk types encountered in the financial industry.

The best-known type of risk is probably *market risk*: the risk of a change in the value of a financial position or portfolio due to changes in the value of the underlying components on which that portfolio depends, such as stock and bond prices, exchange rates, commodity prices, etc. The next important category is *credit risk*: the risk of not receiving promised repayments on outstanding investments such as loans and bonds, because of the "default" of the borrower. A further risk category is *operational risk*: the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events.

The three risk categories of market, credit and operational risk are the main ones we study in this book, but they do not form an exhaustive list of the full range of possible risks affecting a financial institution, nor are their boundaries always clearly defined. For example, when a corporate bond falls in value this is market risk, but the fall in value is often associated with a deterioration in the credit quality of the issuer, which is related to credit risk. The ideal way forward for a successful handling of financial risk is a *holistic* approach, i.e. an integrated approach taking all types of risk and their interactions into account.

Other important notions of risk are *model risk* and *liquidity risk*. The former is the risk associated with using a misspecified (inappropriate) model for measuring risk. Think, for instance, of using the Black–Scholes model for pricing an exotic option in circumstances where the basic Black–Scholes model assumptions on the underlying securities (such as the assumption of normally distributed returns) are violated. It may be argued that model risk is always present to some degree.

When we talk about liquidity risk we are generally referring to price or market liquidity risk, which can be broadly defined as the risk stemming from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. Liquidity can be thought of as "oxygen for a healthy market"; a market requires it to function properly but most of the time we are not aware of its presence. Its absence, however, is recognized immediately, with often disastrous consequences.

In banking, there is also the concept of *funding liquidity risk*, which refers to the ease with which institutions can raise funding to make payments and meet withdrawals as they arise. The management of funding liquidity risk tends to be a specialist activity of bank treasuries (see, for example, Choudhry 2012) rather than trading-desk risk managers and is not a subject of this book. However, funding liquidity and market liquidity can interact profoundly in periods of financial stress. Firms that have problems obtaining funding may sell assets in fire sales to raise cash, and this in turn can contribute to market illiquidity, depressing prices, distorting the valuation of assets on balance sheets and, in turn, making funding even more difficult to obtain; this phenomenon has been described as a liquidity spiral (Brunnermeier and Pedersen 2009).

In insurance, a further risk category is *underwriting risk*: the risk inherent in insurance policies sold. Examples of risk factors that play a role here are changing patterns of natural catastrophes, changes in demographic tables underlying (long-dated) life products, political or legal interventions, or customer behaviour (such as lapsation).

1.1.3 Measurement and Management

Much of this book is concerned with techniques for the statistical measurement of risk, an activity which is part of the process of managing risk, as we attempt to clarify in this section.

Risk measurement. Suppose we hold a portfolio consisting of d underlying investments with respective weights $w_1, \dots, w_d$, so that the change in value of the portfolio over a given holding period (the so-called profit and loss, or P&L) can be written as $X = \sum_{i=1}^{d} w_i X_i$, where X_i denotes the change in value of the ith investment. Measuring the risk of this portfolio essentially consists of determining its distribution function $F_X(x) = P(X \leqslant x)$, or functionals describing this distribution function such as its mean, variance or 99th percentile.

In order to achieve this, we need a properly calibrated *joint* model for the underlying random vector of investments $(X_1, \dots, X_d)$, so statistical methodology has an important role to play in risk measurement; based on historical observations and given a specific model, a statistical estimate of the distribution of the change in value of a position, or one of its functionals, is calculated. In Chapter 2 we develop a detailed framework framework for risk measurement. As we shall see—and this is indeed a main theme throughout the book—this is by no means an easy task with a unique solution.

It should be clear from the outset that good risk measurement is essential. Increasingly, the clients of financial institutions demand objective and detailed information on the products that they buy, and firms can face legal action when this information is found wanting. For any product sold, a proper quantification of the underlying risks needs to be explicitly made, allowing the client to decide whether or not the product on offer corresponds to his or her risk appetite; the 2007–9 crisis saw numerous violations of this basic principle. For more discussion of the importance of the quantitative approach to risk, see Section 1.5.

Risk management. In a very general answer to the question of what risk management is about, Kloman (1990) writes:

> To many analysts, politicians, and academics it is the management of environmental and nuclear risks, those technology-generated macro-risks that appear to threaten our existence. To bankers and financial officers it is the sophisticated use of such techniques as currency hedging and interest-rate swaps. To insurance buyers or sellers it is coordination of insurable risks and the reduction of insurance costs. To hospital administrators it may mean "quality assurance". To safety professionals it is reducing accidents and injuries. In summary, risk management is *a discipline for living with the possibility that future events may cause adverse effects*.

The last phrase in particular (the emphasis is ours) captures the general essence of risk management: it is about ensuring *resilience* to future events. For a financial institution one can perhaps go further. A financial firm's attitude to risk is not passive and defensive; a bank or insurer actively and willingly takes on risk, because it seeks a return and this does not come without risk. Indeed, risk management can be seen as the core competence of an insurance company or a bank. By using its expertise, market position and capital structure, a financial institution can manage risks by repackaging or bundling them and transferring them to markets in customized ways.

The management of risk at financial institutions involves a range of tasks. To begin with, an enterprise needs to determine the capital it should hold to absorb losses, both for regulatory and economic capital purposes. It also needs to manage the risk on its books. This involves ensuring that portfolios are well diversified and optimizing portfolios according to risk–return considerations. The risk profile of the portfolio can be altered by hedging exposures to certain risks, such as interest-rate or foreign-exchange risk, using derivatives. Alternatively, some risks can be repackaged and sold to investors in a process known as securitization; this has been applied to both insurance risks (weather derivatives and longevity derivatives) and credit risks (mortgage-backed securities, collateralized debt obligations). Firms that use derivatives need to manage their derivatives books, which involves the tasks of pricing, hedging and managing collateral for such trades. Finally, financial institutions need to manage their counterparty credit risk exposures to important trading partners; these arise from bilateral, over-the-counter derivatives trades, but they are also present, for example, in reinsurance treaties.

We also note that the discipline of risk management is very much the core competence of an actuary. Indeed, the Institute and Faculty of Actuaries has used the following definition of the actuarial profession:

> Actuaries are respected professionals whose innovative approach to making business successful is matched by a responsibility to the public interest. Actuaries identify solutions to financial problems. They manage assets and liabilities by analysing past events, assessing the present risk involved and modelling what could happen in the future.

Actuarial organizations around the world have collaborated to create the Chartered Enterprise Risk Actuary qualification to show their commitment to establishing best practice in risk management.

1.2 A Brief History of Risk Management

In this section we treat the historical development of risk management by sketching some of the innovations and some of the events that have shaped modern risk management for the financial industry. We also describe the more recent development of regulation in the industry, which has, to some extent, been a process of reaction to a series of incidents and crises.

1.2.1 From Babylon to Wall Street

Although risk management has been described as "one of the most important innovations of the 20th century" by Steinherr (1998), and most of the story we tell is relatively modern, some concepts that are used in modern risk management, and in derivatives in particular, have been around for longer. In our selective account we stress the example of financial derivatives as these have played a role in many of the events that have shaped modern regulation and increased the complexity of the risk-management challenge.

The ancient world to the twentieth century. A derivative is a financial instrument derived from an underlying asset, such as an option, future or swap. For example, a European call option with strike K and maturity T gives the holder the right, but not the obligation, to obtain from the seller at maturity the underlying security for a price K; a European put option gives the holder the right to dispose of the underlying at a price K.

Dunbar (2000) interprets a passage in the Code of Hammurabi from Babylon of 1800 BC as being early evidence of the use of the option concept to provide financial cover in the event of crop failure. A very explicit mention of options appears in Amsterdam towards the end of the seventeenth century and is beautifully narrated by Joseph de la Vega in his 1688 *Confusión de Confusiones*, a discussion between a lawyer, a trader and a philosopher observing the activity on the Beurs of Amsterdam. Their discussion contains what we now recognize as European call and put options and a description of their use for investment as well as for risk management—it even includes the notion of short selling. In an excellent recent translation (de la Vega 1996) we read:

> If I may explain "opsies" [further, I would say that] through the payment of the premiums, one hands over values in order to safeguard one's stock or to obtain a profit. One uses them as sails for a happy voyage during a beneficent conjuncture and as an anchor of security in a storm.

After this, de la Vega continues with some explicit examples that would not be out of place in any modern finance course on the topic.

Financial derivatives in general, and options in particular, are not so new. Moreover, they appear here as instruments to manage risk, "anchors of security in a storm", rather than as dangerous instruments of speculation, the "wild beasts of finance" (Steinherr 1998), that many believe them to be.

Academic innovation in the twentieth century. While the use of risk-management ideas such as derivatives can be traced further back, it was not until the late twentieth century that a theory of valuation for derivatives was developed. This can be seen as perhaps the most important milestone in an age of academic developments in the general area of quantifying and managing financial risk.

Before the 1950s, the desirability of an investment was mainly equated to its return. In his groundbreaking publication of 1952, Harry Markowitz laid the foundation of the theory of portfolio selection by mapping the desirability of an investment onto a risk–return diagram, where risk was measured using standard deviation (see Markowitz 1952, 1959). Through the notion of an *efficient frontier* the portfolio manager could optimize the return for a given risk level. The following decades saw explosive growth in risk-management methodology, including such ideas as the Sharpe ratio, the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT). Numerous extensions and refinements that are now taught in any MBA course on finance followed.

The famous Black–Scholes–Merton formula for the price of a European call option appeared in 1973 (see Black and Scholes 1973). The importance of this formula was underscored in 1997 when the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel was awarded to Robert Merton and Myron Scholes (Fischer Black had died some years earlier) "for a new method to determine the value of derivatives".

In the final two decades of the century the mathematical finance literature developed rapidly, and many ideas found their way into practice. Notable contributions include the pioneering papers by Harrison and Kreps (1979) and Harrison and Pliska (1981) clarifying the links between no-arbitrage pricing and martingale theory. A further example is the work on the term structure of interest rates by Heath, Jarrow and Morton (1992). These and other papers elaborated the mathematical foundations of financial mathematics. Textbooks on stochastic integration and Itô calculus became part of the so-called quant's essential reading and were, for a while, as likely to be seen in the hands of a young investment banker as the *Financial Times*.

Growth of markets in the twentieth century. The methodology developed for the rational pricing and hedging of financial derivatives changed finance. The "wizards of Wall Street" (i.e. the mathematical specialists conversant in the new methodology) have had a significant impact on the development of financial markets over the last few decades. Not only did the new option-pricing formula work, it transformed the market. When the Chicago Options Exchange first opened in 1973, fewer than a thousand options were traded on the first day. By 1995, over a million options were changing hands each day, with current nominal values outstanding in the derivatives markets in the tens of trillions. So great was the role played by the Black–Scholes–Merton formula in the growth of the new options market that, when the American stock market crashed in 1987, the influential business magazine *Forbes* attributed the blame squarely to that one formula. Scholes himself has said that it was not so much the formula that was to blame, but rather that market traders had not become sufficiently sophisticated in using it.

Along with academic innovation, developments in information technology (IT) also helped lay the foundations for an explosive growth in the volume of new risk-management and investment products. This development was further aided by worldwide deregulation in the 1980s. Important additional factors contributing to an increased demand for risk-management skills and products were the oil crises of the 1970s and the 1970 abolition of the Bretton Woods system of fixed exchange rates. Both energy prices and foreign exchange risk became highly volatile risk factors and customers required products to hedge them. The 1933 Glass–Steagall Act—passed in the US in the aftermath of the 1929 Depression to prohibit commercial banks from underwriting insurance and most kinds of securities—indirectly paved the way for the emergence of investment banks, hungry for new business. Glass–Steagall was replaced in 1999 by the Financial Services Act, which repealed many of the former's key provisions, although the 2010 Dodd–Frank Act, passed in the aftermath of the 2007–9 financial crisis, appears to mark an end to the trend of deregulation.

Disasters of the 1990s. In January 1992 the president of the New York Federal Reserve, E. Gerald Corrigan, speaking at the Annual Mid-Winter Meeting of the New York State Bankers Association, said:

> 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. Off-balance-sheet activities [i.e. derivatives] have a role, but they must be managed and controlled carefully and they must be understood by top management as well as by traders and rocket scientists.

Corrigan was referring to the growing volume of derivatives in banks' trading books and the fact that, in many cases, these did not appear as assets or liabilities on the balance sheet. His words proved prescient.

On 26 February 1995 Barings Bank was forced into administration. A loss of £700 million ruined the oldest merchant banking group in the UK (established in 1761). Besides numerous operational errors (violating every qualitative guideline in the risk-management handbook), the final straw leading to the downfall of Barings was a so-called straddle position on the Nikkei held by the bank's Singapore-based trader Nick Leeson. A straddle is a short position in a call and a put with the same strike—such a position allows for a gain if the underlying (in this case the Nikkei index) does not move too far up or down. There is, however, considerable loss potential if the index moves down (or up) by a large amount, and this is precisely what happened when the Kobe earthquake occurred.

Three years later, Long-Term Capital Management (LTCM) became another prominent casualty of losses due to derivatives trading when it required a \$3.5 billion payout to prevent collapse, a case made all the more piquant by the fact that

Myron Scholes and Robert Merton were principals at the hedge fund. Referring to the Black–Scholes formula, an article in the *Observer* newspaper asked: "Is this really the key to future wealth? Win big, lose bigger."

There were other important cases in this era, leading to a widespread discussion of the need for increased regulation, including Metallgesellschaft in 1993 (speculation on oil prices using derivatives) and Orange County in 1994 (speculation on interest rates using derivatives).

In the life insurance industry, Equitable Life, the world's oldest mutual insurer, provided a case study of what can happen when the liabilities arising from insurance products with embedded options are not properly hedged. Prior to 1988, Equitable Life had sold pension products that offered the option of a guaranteed annuity rate at maturity of the policy. The guarantee rate of 7% had been set in the 1970s when inflation and annuity rates were high, but in 1993 the current annuity rate fell below the guarantee rate and policyholders exercised their options. Equitable Life had not been hedging the option and it quickly became evident that they were faced with an enormous increase in their liabilities; the Penrose Report (finally published in March 2004) concluded that Equitable Life was underfunded by around £4.5 billion by 2001. It was the policyholders who suffered when the company reneged on their pension promises, although many of the company's actions were later ruled unlawful and some compensation from the public purse was agreed. However, this case provides a good illustration of the need to regulate the capital adequacy of insurers to protect policyholders.

The turn of the century. The end of the twentieth century proved to be a pivotal moment for the financial system worldwide. From a value of around 1000 in 1996, the Nasdaq index quintupled to a maximum value of 5408.62 on 10 March 2000 (which remains unsurpassed as this book goes to press). The era 1996–2000 is now known as the dot-com bubble because many of the firms that contributed to the rise in the Nasdaq belonged to the new internet sector.

In a speech before the American Enterprise Institute on 5 December 1996, Alan Greenspan, chairman of the Federal Reserve from 1987 to 2006, said, "But how do we know when *irrational exuberance* has unduly escalated assets, which then become subject to prolonged contractions as they have in Japan over the past decade?" The term irrational exuberance seemed to perfectly describe the times. The Dow Jones Industrial Average was also on a historic climb, breaking through the 10 000 barrier on 29 March 1999, and prompting books with titles like *Dow 40 000: Strategies for Profiting from the Greatest Bull Market in History*. It took four years for the bubble to burst, but from its March 2000 maximum the Nasdaq plummeted to half of its value within a year and tested the 1000 barrier in late 2002. Equity indices fell worldwide, although markets recovered and began to surge ahead again from 2004.

The dot-com bubble was in many respects a conventional asset bubble, but it was also during this period that the seeds of the next financial crisis were being sown. Financial engineers had discovered the magic of securitization: the bundling and repackaging of many risks into securities with defined risk profiles that could be sold to potential investors. While the idea of transferring so-called tranches of a pool of risks to other risk bearers was well known to the insurance world, it was now being applied on a massive scale to credit-risky assets, such as mortgages, bonds, credit card debt and even student loans (see Section 12.1.1 for a description of the tranching concept).

In the US, the subprime lending boom to borrowers with low credit ratings fuelled the supply of assets to securitize and a market was created in mortgage-backed securities (MBSs). These in turn belonged to the larger pool of assets that were available to be transformed into collateralized debt obligations (CDOs). The banks originating these credit derivative products had found a profitable business turning poor credit risks into securities. The volume of credit derivatives ballooned over a very short period; the CDO market accounted for almost $3 trillion in nominal terms by 2008 but this was dwarfed by the nominal value of the credit default swap (CDS) market, which stood at about $30 trillion.

Credit default swaps, another variety of credit derivative, were originally used as instruments for hedging large corporate bond exposures, but they were now increasingly being used by investors to speculate on the changing credit outlook of companies by adopting so-called naked positions (see Section 10.1.4 for more explanation). Although the actual economic value of CDS and CDO markets was actually smaller (when the netting of cash flows is considered), these are still huge figures when compared with world gross domestic product (GDP), which was of the order of $60 trillion at that time.

The consensus was that all this activity was a good thing. Consider the following remarks made by the then chairman of the Federal Reserve, Alan Greenspan, before the Council on Foreign Relations in Washington DC on 19 November 2002 (Greenspan 2002):

> More recently, instruments ... such as credit default swaps, collateralized debt obligations and credit-linked notes have been developed and their use has grown rapidly in recent years. The result? Improved credit risk management together with more and better risk-management tools appear to have significantly reduced loan concentrations in telecommunications and, indeed, other areas and the associated stress on banks and other financial institutions.... It is noteworthy that payouts in the still relatively small but rapidly growing market in credit derivatives have been proceeding smoothly for the most part. Obviously this market is still too new to have been tested in a widespread down-cycle for credit, but, to date, it appears to have functioned well.

As late as April 2006 the International Monetary Fund (IMF) wrote in its Global Financial Stability Report that:

> There is a growing recognition that the dispersion of credit risk by banks to a broader and more diverse group of investors, rather than warehousing such risks on their balance sheets, has helped to make the banking and overall financial system more resilient.... The improved

> resilience may be seen in fewer bank failures and more consistent credit provision. Consequently, the commercial banks, a core system of the financial system, may be less vulnerable today to credit or economic shocks.

It has to be said that the same IMF report also warned about possible vulnerabilities, and the potential for market disruption, if these credit instruments were not fully understood.

One of the problems was that not all of the risk from CDOs was being dispersed to outside investors as the IMF envisaged. As reported in Acharya et al. (2009), large banks were holding on to a lot of it themselves:

> These large, complex financial institutions ignored their own business models of securitization and chose not to transfer credit risk to other investors. Instead they employed securitization to manufacture and retain tail risk that was systemic in nature and inadequately capitalized.... Starting in 2006, the CDO group at UBS noticed that their risk-management systems treated AAA securities as essentially risk-free even though they yielded a premium (the proverbial free lunch). So they decided to hold onto them rather than sell them! After holding less than $5 billion of them in 02/06, the CDO desk was warehousing a staggering $50 billion in 09/07.... Similarly, by late summer of 2007, Citigroup had accumulated over $55 billion of AAA-rated CDOs.

On the eve of the crisis many in the financial industry seemed unconcerned. AIG, the US insurance giant, had become heavily involved in underwriting MBS and CDO risk by selling CDS protection through its AIG Financial Products arm. In August 2007 the chief executive officer of AIG Financial Products is quoted as saying:

> It is hard for us, without being flippant, to even see a scenario within any kind of realm of reason that would see us losing one dollar in any of these transactions.

The financial crisis of 2007–9. After a peak in early 2006, US house prices began to decline in 2006 and 2007. Subprime mortgage holders, experiencing difficulties in refinancing their loans at higher interest rates, defaulted on their payments in increasing numbers. Starting in late 2007 this led to a rapid reassessment of the riskiness of securitizations and to losses in the value of CDO securities. Banks were forced into a series of dramatic *write-downs* of the value of these assets on their balance sheets, and the severity of the impending crisis became apparent.

Reflecting on the crisis in his article "It doesn't take Nostradamus" in the 2008 issue of *Economists' Voice*, Nobel laureate Joseph E. Stiglitz recalled the views he expressed in 1992 on securitization and the housing market:

> The question is, has the growth of securitization been a result of more efficient transaction technologies or an unfounded reduction in concern about the importance of screening loan applicants? It is perhaps too

> early to tell, but we should at least entertain the possibility that it is the latter rather than the former.

He also wrote:

> At the very least, the banks have demonstrated ignorance of two very basic aspects of risk: (a) the importance of correlation ... [and] (b) the possibility of price declines.

These "basic aspects of risk", which would appear to belong in a Banking 101 class, plunged the world's economy into its most serious crisis since the late 1920s. Salient events included the demise of such illustrious names as Bear Stearns (which collapsed and was sold to JPMorgan Chase in March 2008) and Lehman Brothers (which filed for Chapter 11 bankruptcy on 15 September 2008). The latter event in particular led to worldwide panic. As markets tumbled and liquidity vanished it was clear that many banks were on the point of collapse. Governments had to bail them out by injecting capital or by acquiring their distressed assets in arrangements such as the US Troubled Asset Relief Program.

AIG, which had effectively been insuring the default risk in securitized products by selling CDS protection, got into difficulty when many of the underlying securities defaulted; the company that could not foresee itself "losing one dollar in any of these transactions" required an emergency loan facility of $85 billion from the Federal Reserve Bank of New York on 16 September 2008. In the view of George Soros (2009), CDSs were "instruments of destruction" that should be outlawed:

> Some derivatives ought not to be allowed to be traded at all. I have in mind credit default swaps. The more I've heard about them, the more I've realised they're truly toxic.

Much has been written about these events, and this chapter's Notes and Comments section contains a number of references. One strand of the commentary that is relevant for this book is the apportioning of a part of the blame to mathematicians (or financial engineers); the failure of valuation models for complex securitized products made them an easy target. Perhaps the most publicized attack came in a blog by Felix Salmon (*Wired Magazine*, 23 February 2009) under the telling title "Recipe for disaster: the formula that killed Wall Street". The formula in question was the *Gauss copula*, and its application to credit risk was attributed to David Li. Inspired by what he had learned on an actuarial degree, Li proposed that a tool for modelling dependent lifetimes in life insurance could be used to model correlated default times in bond portfolios, thus providing a framework for the valuation and risk management of CDOs, as we describe in Chapter 12.

While an obscure formula with a strange name was a gift for bloggers and newspaper headline writers, even serious regulators joined in the chorus of criticism of mathematics. The Turner Review of the global banking crisis (Lord Turner 2009) has a section entitled "Misplaced reliance on sophisticated mathematics" (see Section 1.3.3 for more on this theme). But this reliance on mathematics was only one factor in the crisis, and certainly not the most important. Mathematicians had also warned well beforehand that the world of securitization was being built on shaky model foundations that were difficult to calibrate (see, for example, Frey, McNeil and Nyfeler 2001). It was also abundantly clear that political shortsightedness, the greed of market participants and the slow reaction of regulators had all contributed in very large measure to the scale of the eventual calamity.

Recent developments and concerns. New threats to the financial system emerge all the time. The financial crisis of 2007–9 led to recession and sovereign debt crises. After the wave of bank bailouts, concerns about the solvency of banks were transformed into concerns about the abilities of countries to service their own debts. For a while doubts were cast on the viability of the eurozone, as it seemed that countries might elect to, or be forced to, exit the single currency.

On the more technical side, the world of high-frequency trading has raised concerns among regulators, triggered by such events as the Flash Crash of 6 May 2010. In this episode, due to "computer trading gone wild", the Dow Jones lost around 1000 points in a couple of minutes, only to be rapidly corrected. High-frequency trading is a form of algorithmic trading in which trades are executed by computers according to algorithms in fractions of a second. One notable casualty of algorithmic trading was Knight Capital, which lost $460 million due to trading errors on 1 August 2012. Going forward, it is clear that vigilance is required concerning the risks arising from the deployment of new technologies and their systemic implications.

Indeed, *systemic risk* is an ongoing concern to which we have been sensitized by the financial crisis. This is the risk of the collapse of the entire financial system due to the propagation of financial stress through a network of participants. When Lehman Brothers failed there was a moment when it seemed possible that there could be a catastrophic cascade of defaults of banks and other firms. The interbank lending market had become dysfunctional, asset prices had plummeted and the market for any form of debt was highly illiquid. Moreover, the complex chains of relationships in the CDS markets, in which the same credit-risky assets were referenced in a large volume of bilateral payment agreements, led to the fear that the default of a further large player could cause other banks to topple like dominoes.

The concerted efforts of many governments were successful in forestalling the Armageddon scenario. However, since the crisis, research into financial networks and their embedded systemic risks has been an important research topic. These networks are complex, and as well as banks and insurance companies they contain members of a "shadow banking system" of hedge funds and structured investment vehicles, which are largely unregulated. One important theme is the identification of so-called systemically important financial institutions (SIFI) whose failure might cause a systemic crisis.

1.2.2 The Road to Regulation

There is no doubt that regulation goes back a long way, at least to the time of the Venetian banks and the early insurance enterprises sprouting in London's coffee shops in the eighteenth century. In those days there was more reliance on self-regulation or local regulation, but rules were there. However, the key developments that led to the present prudential regulatory framework in financial services are a very much more recent story.

The main aim of modern prudential regulation has been to ensure that financial institutions have enough capital to withstand financial shocks and remain solvent. Robert Jenkins, a member of the Financial Policy Committee of the Bank of England, was quoted in the *Independent* on 27 April 2012 as saying:

> Capital is there to absorb losses from risks we understand and risks we may not understand. Evidence suggests that neither risk-takers nor their regulators fully understand the risks that banks sometimes take. That's why banks need an appropriate level of loss absorbing equity.

Much of the regulatory drive originated from the Basel Committee of Banking Supervision. This committee was established by the central-bank governors of the Group of Ten at the end of 1974. The Group of Ten is made up of (oddly) eleven industrial countries that consult and cooperate on economic, monetary and financial matters. The Basel Committee does not possess any formal supranational supervising authority, and hence its conclusions do not have legal force. Rather, it formulates broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements—statutory or otherwise—that are best suited to their own national system. The summary below is brief. Interested readers can consult, for example, Tarullo (2008) for further details, and should also see this chapter's Notes and Comments section.

The first Basel Accord. The first Basel Accord on Banking Supervision (Basel I, from 1988) took an important step towards an international minimum capital standard. Its main emphasis was on credit risk, by then clearly the most important source of risk in the banking industry. In hindsight, however, Basel I took an approach that was fairly coarse and measured risk in an insufficiently differentiated way. In measuring credit risk, claims were divided into three crude categories according to whether the counterparties were governments, regulated banks or others. For instance, the risk capital charge for a loan to a corporate borrower was five times higher than for a loan to an Organisation for Economic Co-operation and Development (OECD) bank. The risk weighting for all corporate borrowers was identical, independent of their credit rating. The treatment of derivatives was also considered unsatisfactory.

The birth of VaR. In 1993 the G-30 (an influential international body consisting of senior representatives of the private and public sectors and academia) published a seminal report addressing, for the first time, so-called off-balance-sheet products, like derivatives, in a systematic way. Around the same time, the banking industry clearly saw the need for proper measurement of the risks stemming from these new products. At JPMorgan, for instance, the famous Weatherstone 4.15 report asked for a one-day, one-page summary of the bank's market risk to be delivered to the chief executive officer in the late afternoon (hence "4.15"). Value-at-risk (VaR) as a market risk measure was born and the JPMorgan methodology, which became known as RiskMetrics, set an industry-wide standard.

In a highly dynamic world with round-the-clock market activity, the need for instant market valuation of trading positions (known as *marking-to-market*) became a necessity. Moreover, in markets where so many positions (both long and short) were written on the same underlyings, managing risks based on simple aggregation of nominal positions became unsatisfactory. Banks pushed to be allowed to consider *netting* effects, i.e. the compensation of long versus short positions on the same underlying.

In 1996 an important amendment to Basel I prescribed a so-called *standardized* model for market risk, but at the same time allowed the bigger (more sophisticated) banks to opt for an *internal* VaR-based model (i.e. a model developed in house). Legal implementation was to be achieved by the year 2000. The coarseness problem for credit risk remained unresolved and banks continued to claim that they were not given enough incentives to diversify credit portfolios and that the regulatory capital rules currently in place were far too risk insensitive. Because of overcharging on the regulatory capital side of certain credit positions, banks started shifting business away from certain market segments that they perceived as offering a less attractive risk–return profile.

The second Basel Accord. By 2001 a consultative process for a new Basel Accord (Basel II) had been initiated; the basic document was published in June 2004. An important aspect was the establishment of the three-pillar system of regulation: Pillar 1 concerns the quantification of regulatory capital; Pillar 2 imposes regulatory oversight of the modelling process, including risks not considered in Pillar 1; and Pillar 3 defines a comprehensive set of disclosure requirements.

Under Pillar 1 the main theme of Basel II was credit risk, where the aim was to allow banks to use a finer, more risk-sensitive approach to assessing the risk of their credit portfolios. Banks could opt for an *internal-ratings-based* approach, which permitted the use of internal or external credit-rating systems wherever appropriate.

The second important theme of Basel II at the level of Pillar 1 was the consideration of operational risk as a new risk class. A basic premise of Basel II was that the overall size of regulatory capital throughout the industry should stay unchanged under the new rules. Since the new rules for credit risk were likely to reduce the credit risk charge, this opened the door for operational risk, defined as the risk of losses resulting from inadequate or failed internal processes, people and systems or from external events; this definition included legal risk but excluded reputational and strategic risk.

Mainly due to the financial crisis of 2007–9, implementation of the Basel II guidelines across the globe met with delays and was rather spread out in time. Various further amendments and additions to the content of the original 2004 document were made. One important criticism of Basel II that emerged from the crisis was that it was inherently *procyclical*, in that it forced firms to take action to increase their capital ratios at exactly the wrong point in the business cycle, when their actions had a negative impact on the availability of liquidity and made the situation worse (see Section 1.3.3 for more discussion on this).

Basel 2.5. One clear lesson from the crisis was that modern products like CDOs had opened up opportunities for regulatory arbitrage by transferring credit risk from the capital-intensive banking book (or loan book) to the less-capitalized trading book. Some enhancements to Basel II were proposed in 2009 with the aim of addressing the build-up of risk in the trading book that was evident during the crisis. These enhancements, which have come to be known as Basel 2.5, include a *stressed VaR* charge, based on calculating VaR from data for a twelve-month period of market turmoil, and the so-called *incremental risk charge*, which seeks to capture some of the default risk in trading book positions; there were also specific new rules for certain securitizations.

The third Basel Accord. In view of the failure of the Basel rules to prevent the 2007–9 crisis, the recognized deficiencies of Basel II mentioned above, and the clamour from the public and from politicians for regulatory action to make banks and the banking system safer, it is no surprise that attention quickly shifted to Basel III.

In 2011 a series of measures was proposed that would extend Basel II (and 2.5) in five main areas:

(1) measures to increase the quality and amount of bank capital by changing the definition of key capital ratios and allowing countercyclical adjustments to these ratios in crises;

(2) a strengthening of the framework for counterparty credit risk in derivatives trading, with incentives to use central counterparties (exchanges);

(3) the introduction of a leverage ratio to prevent excessive leverage;

(4) the introduction of various ratios that ensure that banks have sufficient funding liquidity;

(5) measures to force systemically important banks to have even higher capacity to absorb losses.

Most of the new rules will be phased in progressively, with a target end date of 2019, although individual countries may impose stricter guidelines with respect to both schedule and content.

Parallel developments in insurance regulation. The insurance industry worldwide has also been subject to increasing risk regulation in recent times. However, here the story is more fragmented and there has been much less international coordination of efforts. The major exception has been the development of the Solvency II framework in the European Union, a process described in more detail below. As the most detailed and model intensive of the regulatory frameworks proposed, it serves as our main reference point for insurance regulation in this book. The development of the Solvency II framework is overseen by the European Insurance and Occupational Pensions Authority (EIOPA; formerly the Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS)), but the implementation in individual countries is a matter for national regulators, e.g. the Prudential Regulatory Authority in the UK.

In the US, insurance regulation has traditionally been a matter for state governments. The National Association of Insurance Commissioners (NAIC) provides support to insurance regulators from the individual states, and helps to promote the development of accepted regulatory standards and best practices; it is up to the individual states whether these are passed into law, and if so in what form. In the early 1990s the NAIC promoted the concept of risk-based capital for insurance companies as a response to a number of insolvencies in the preceding years; the NAIC describes risk-based capital as "a method of measuring the minimum amount of capital appropriate for a reporting entity to support its overall business operations in consideration of its size and profile". The method, which is a rules-based approach rather than a model-based approach, has become the main plank of insurance regulation in the US.

Federal encroachment on insurance supervision has generally been resisted, although this may change due to a number of measures enacted after the 2007–9 crisis in the wide-ranging 2010 Dodd–Frank Act. These include the creation of both the Federal Insurance Office, to "monitor all aspects of the insurance sector", and the Financial Stability Oversight Council, which is "charged with identifying risks to the financial stability of the United States" wherever they may arise in the world of financial services.

The International Association of Insurance Supervisors has been working to foster some degree of international convergence in the processes for regulating the capital adequacy of insurers. They have promoted the idea of the Own Risk and Solvency Assessment (ORSA). This has been incorporated into the Solvency II framework and has also been embraced by the NAIC in the US.

There are also ongoing initiatives that aim to bring about convergence of banking and insurance regulation, particularly with respect to financial conglomerates engaged in both banking and insurance business. The Joint Forum on Financial Conglomerates was established in early 1996 under the aegis of the Basel Committee, the International Association of Insurance Supervisors and the International Organization of Securities Commissions to take forward this work.

From Solvency I to Solvency II. Mirroring the progress in the banking sector, Solvency II is the latest stage in a process of regulatory evolution from simple and crude rules to a more risk-sensitive treatment of the capital requirements of insurance companies.

The first European Union non-life and life directives on solvency margins appeared around 1970. The solvency margin was defined as an extra capital buffer against unforeseen events such as higher than expected claims levels or unfavourable investment results. However, there were differences in the way that regulation was applied across Europe and there was a desire for more harmonization of regulation and mutual recognition.

Solvency I, which came into force in 2004, is a rather coarse rules-based framework calling for companies to have a minimum guarantee fund (minimal capital) of €3 million, and a solvency margin consisting of 16–18% of non-life premiums together with 4% of the technical provisions for life. This has led to a single robust system that is easy to understand and inexpensive to monitor. However, on the negative side, it is mainly volume based, not explicitly risk based; issues like guarantees, embedded options and the proper matching of assets and liabilities are largely neglected in many countries.

To address these shortcomings, Solvency II was initiated in 2001 with the publication of the influential Sharma Report. While the Solvency II directive was adopted by the Council of the European Union and the European Parliament in November 2009, implementation of the framework is not expected until 1 January 2016. The process of refinement of the framework is managed by EIOPA, and one of the features of this process has been a series of quantitative impact studies in which companies have effectively tried out aspects of the proposals and information has been gathered with respect to the impact and practicability of the new regulations.

The goal of the Solvency II process is that the new framework should strengthen the capital adequacy regime by reducing the possibilities of consumer loss or market disruption in insurance; Solvency II therefore has both policyholder-protection and financial-stability motives. Moreover, it is also an aim that the harmonization of regulation in Europe should promote deeper integration of the European Union insurance market and the increased competitiveness of European insurers. A high-level description of the Solvency II framework is given in Section 1.3.2.

The Swiss Solvency Test (SST). Special mention should be made of Switzerland, which has already developed and implemented its own principles-based risk capital regulation for the insurance industry. The SST has been in force since 1 January 2011. It follows similar principles to Solvency II but differs in some details of its treatment of different types of risk; it also places more emphasis on the development of internal models. The implementation of the SST falls under the remit of the Swiss Financial Markets Supervisory Authority, a body formed in 2007 from the merger of the banking and insurance supervisors, which has statutory authority over banks, insurers, stock exchanges, collective investment schemes and other entities.

1.3 The Regulatory Framework

This section describes in more detail the framework that has emerged from the Basel process and the European Union solvency process.

1.3.1 The Basel Framework

As indicated in Section 1.2.2, the Basel framework should be regarded as the product of an evolutionary process. As this book goes to press, the Basel II and Basel 2.5 proposals have been implemented in many developed countries (with some variations in detail), while the proposals of Basel III are still being debated and refined. We sketch the framework as currently implemented, before indicating some of the proposed changes and additions to the framework in Basel III.

The three-pillar concept. A key feature of the Basel framework is the three-pillar concept, as is apparent from the following statement summarizing the Basel philosophy, which accompanied the original Basel II publication (Basel Committee on Banking Supervision 2004):

> The Basel II Framework sets out the details for adopting more risk-sensitive minimum capital requirements [Pillar 1] for banking organizations. The new framework reinforces these risk-sensitive requirements by laying out principles for banks to assess the adequacy of their capital and for supervisors to review such assessments to ensure banks have adequate capital to support their risks [Pillar 2]. It also seeks to strengthen market discipline by enhancing transparency in banks' financial reporting [Pillar 3]. The text that has been released today reflects the results of extensive consultations with supervisors and bankers worldwide. It will serve as the basis for national rule-making and approval processes to continue and for banking organizations to complete their preparations for the new Framework's implementation.

Under *Pillar 1*, banks are required to calculate a *minimum capital charge*, referred to as regulatory capital. There are separate Pillar 1 capital charges for credit risk in the banking book, market risk in the trading book and operational risk, which are considered to be the main quantifiable risks. Most banks use internal models based on VaR methodology to compute the capital charge for market risk. For credit risk and operational risk banks may choose between several approaches of increasing risk sensitivity and complexity, some details of which are discussed below.

Pillar 2 recognizes that any quantitative approach to risk management should be embedded in a properly functioning corporate governance structure. Best-practice risk management imposes constraints on the organization of the institution, i.e. the board of directors, management, employees, and internal and external audit processes. In particular, the board of directors assumes the ultimate responsibility for oversight of the risk landscape and the formulation of the company's risk appetite. Through Pillar 2, also referred to as the *supervisory review process*, local regulators review the various checks and balances that have been put in place. Under Pillar 2, residual quantifiable risks that are not included in Pillar 1, such as interest-rate risk in the banking book, must be considered and *stress tests* of a bank's capital adequacy must be performed. The aim is to ensure that the bank holds capital in line with its true economic loss potential, a concept known as *economic capital*.

Finally, in order to fulfil its promise that increased regulation will increase transparency and diminish systemic risk, clear reporting guidelines on the risks carried by financial institutions are called for. *Pillar 3* seeks to establish *market discipline* through a better public disclosure of risk measures and other information relevant to risk management. In particular, banks will have to offer greater insight into the adequacy of their capitalization.

Credit and market risk; the banking and trading books. Historically, banking activities have been organized around the banking book and the trading book, a split that reflects different accounting practices for different kinds of assets. The banking book contains assets that are *held to maturity*, such as loans; these are typically valued at book value, based on the original cost of the asset. The trading book contains assets and instruments that are *available to trade*; these are generally valued by *marking-to-market* (i.e. using quoted market prices). From a regulatory point of view, credit risk is mainly identified with the banking book and market risk is mainly identified with the trading book.

We have already noted that there are problems with this simple dichotomy and that the Basel 2.5 rules were introduced (partly) to account for the neglect of credit risk (default and rating-migration risk) in the trading book. There are also forms of market risk in the banking book, such as interest-rate risk and foreign-exchange risk. However, the Basel framework continues to observe the distinction between banking book and trading book and we will describe the capital charges in terms of the two books. It is clear that the distinction is somewhat arbitrary and rests on the concept of "available to trade". Moreover, there can be incentives to "switch" or move instruments from one book to the other (particularly from the banking book to the trading book) to benefit from a more favourable capital treatment. This is acknowledged by the Basel Committee in its background discussion of the "Fundamental review of the trading book: a revised market risk framework" (Basel Committee on Banking Supervision 2013a):

> The Committee believes that the definition of the regulatory boundary between the trading book and the banking book has been a source of weakness in the design of the current regime. A key determinant of the boundary has been banks' self-determined intent to trade.... Coupled with large differences in capital requirements against similar types of risk on either side of the boundary, the overall capital framework proved susceptible to arbitrage before and during the crisis.... To reduce the incentives for arbitrage, the Committee is seeking a less permeable boundary with strict limits on switching between books and measures to prevent "capital benefit" in instances where switching is permitted.

The capital charge for the banking book. The credit risk of the banking book portfolio is assessed as the sum of *risk-weighted assets*: that is, the sum of notional exposures weighted by a coefficient reflecting the creditworthiness of the counterparty (the risk weight). To calculate risk weights, banks use either the *standardized* approach or one of the more advanced *internal-ratings-based* (IRB) approaches. The choice of method depends on the size and complexity of the bank, with the larger, international banks having to go for IRB approaches. The capital charge is determined as a fraction of the sum of risk-weighted assets in the portfolio. This fraction, known as the capital ratio, was 8% under Basel II but is already being increased ahead of the planned implementation of Basel III in 2019.

The standardized approach refers to a system that has been in place since Basel I, whereby the risk weights are prescribed by the regulator according to the nature and creditworthiness of the counterparty. For example, there are risk weights for retail loans secured on property (mortgages) and for unsecured retail loans (such as credit cards and overdrafts); there are also different risk weights for corporate and government bonds with different ratings.

Under the more advanced IRB approaches, banks may dispense with the system of fixed risk weights provided by the regulator. Instead, they may make an *internal* assessment of the riskiness of a credit exposure, expressing this in terms of an estimated annualized *probability of default* and an estimated *loss given default*, which are used as inputs in the calculation of risk-weighted assets. The total sum of risk-weighted assets is calculated using formulas specified by the Basel Committee; the formulas also take into account the fact that there is likely to be positive correlation (sometimes called systematic risk) between the credit risks in the portfolio. The use of internally estimated probabilities of default and losses given default allows for increased risk sensitivity in the IRB capital charges compared with the standardized approach. It should be noted, however, that the IRB approaches do not permit fully internal models of credit risk in the banking book; they only permit internal estimation of inputs to a model that has been specified by the regulator.

The capital charge for the trading book. For market risk in the trading book there is also the option of a standardized approach based on a system of risk weights and specific capital charges for different kinds of instrument. However, most major banks elect to use an *internal VaR model approach*, as permitted by the 1996 amendment to Basel I. In Sections 2.2 and 9.2 of this book we give a detailed description of the VaR approach to trading book risk measurement. The approach is based on the estimation of a P&L distribution for a ten-day holding period and the estimation of a particular percentile of this distribution: the 99th percentile of the losses.

A ten-day VaR at 99% of $20 million therefore means that it is estimated that our market portfolio will incur a loss of $20 million *or more* with probability 1% by the end of a ten-day holding period, if the composition remains fixed over this period. The conversion of VaR numbers into an actual capital charge is accomplished by a formula that we discuss in Section 2.3.3.

The VaR calculation is the main component of risk quantification for the trading book, but the 2009 Basel 2.5 revision added further elements (see Basel Committee on Banking Supervision 2012, p. 10), including the following.

Stressed VaR: banks are required to carry out a VaR calculation essentially using the standard VaR methodology but calibrating their models to a historical twelve-month period of significant financial stress.

Incremental risk charge: Since default and rating-migration risk are not generally considered in the standard VaR calculation, banks must calculate an additional charge based on an estimate of the 99.9th percentile of the one-year distribution of losses due to defaults and rating changes. In making this calculation they may use internal models for credit risk (in contrast to the banking book) but must also take into account the market liquidity of credit-risky instruments.

Securitizations: exposures to securitizations in the trading book are subject to a series of new capital charges that bring them more into line with equivalent exposures in the banking book.

The capital charge for operational risk. There are also options of increasing sophistication for assessing operational risk. Under the *basic-indicator* and *standardized* approaches, banks may calculate their operational risk charge using simple formulas based on gross annual income. Under the *advanced measurement approach*, banks may develop internal models. Basel is not prescriptive about the form of these models provided they capture the tail risk of extreme events; most such models are based on historical loss data (internal and external to the firm) and use techniques that are drawn from the actuarial modelling of general insurance losses. We provide more detail in Chapter 13.

New elements of Basel III. Under Basel III there will be a number of significant changes and additions to the Basel framework. While the detail of the new rules may change before final implementation in 2019, the main developments are now clear.

- Banks will need to hold both *more capital* and *better-quality capital* as a function of the risks taken. The "better quality" is achieved though a more restrictive definition of eligible capital (through more stringent definitions of Tier 1 and Tier 2 capital and the phasing out of Tier 3 capital); see Section 2.1.3 for more explanation of capital tiers. The "more" comes from the addition (on top of the minimum ratio of 8%) of a capital conservation buffer of 2.5% of risk-weighted assets, for building up capital in good times to absorb losses under stress, and a countercyclical buffer within the range 0–2.5%, in order to enhance the shock resilience of banks and limit expansion in periods of excessive credit growth. This leads to a total (Tier 1 plus Tier 2) ratio of up to 13%, compared with Basel II's 8%. There will be a gradual phasing in of all these new ratios, with a target date for full implementation of 1 January 2019.

- A *leverage ratio* will be imposed to put a floor under the build-up of excessive leverage in the banking system. Leverage will essentially be measured through the ratio of Tier 1 capital to total assets. A minimum ratio of 3% is currently being tested but the precise definitions may well change as a result of testing experience and bank lobbying. The leverage limit will restrain the size of bank assets, regardless of their riskiness.

- The risk coverage of the system of capital charges is being extended, in particular to include a charge for *counterparty credit risk*. When counterparty credit risk is taken into account in the valuation of over-the-counter derivatives contract, the default-risk-free value has to be adjusted by an amount known as the credit value adjustment (CVA); see Section 17.2 for more explanation. There will now be a charge for changes in CVA.

- Banks will become subject to *liquidity rules*; this is a completely new direction for the Basel framework, which has previously been concerned only with capital adequacy. A *liquidity coverage ratio* will be introduced to ensure that banks have enough highly liquid assets to withstand a period of net cash outflow lasting thirty days. A *net stable funding ratio* will ensure that sufficient funding is available in order to cover long-term commitments (exceeding one year).

It should also be mentioned that under an ongoing review of the trading book, the principle of risk quantification may change from one based on VaR (a percentile) to one based on *expected shortfall* (ES). For a given holding period, the ES at the 99% level, say, is the expected loss given that the loss is higher than the VaR at the 99% level over the same period. ES is a severity measure that always dominates the frequency measure VaR and gives information about the expected size of tail losses; it is also a measure with superior aggregation properties to VaR, as discussed in Section 2.3.5 and Chapter 8 (particularly Sections 8.1 and 8.4.4).

1.3.2 The Solvency II Framework

Below we give an outline of the Solvency II framework, which will come into force in the countries of the European Union on or before 1 January 2016.

Main features. In common with the Basel Accords, Solvency II adopts a three-pillar system, where the first pillar requires the quantification of regulatory capital requirements, the second pillar is concerned with governance and supervision, and the third pillar requires the disclosure of information to the public to improve market discipline by making it easier to compare the risk profiles of companies.

Under Pillar 1, a company calculates its *solvency capital requirement*, which is the amount of capital it should have to ensure that the probability of insolvency over a one-year period is no more than 0.5%—this is often referred to as a confidence level of 99.5%. The company also calculates a smaller *minimum capital requirement*, which is the minimum capital it should have to continue operating without supervisory intervention.

To calculate the capital requirements, companies may use either an *internal model* or a simpler *standard formula* approach. In either case the intention is that a *total balance sheet* approach is taken in which all risks and their interactions are considered. The insurer should have *own funds* (a surplus of assets over liabilities) that exceed both the solvency capital requirement and the minimum capital requirement. The assets and liabilities of the firm should be valued in a *market-consistent* manner.

The supervisory review of the company takes place under Pillar 2. The company must demonstrate that it has a risk-management system in place and that this system is integrated into decision-making processes, including the setting of risk appetite by the company's board, and the formulation of risk limits for different business units. An internal model must pass the "use test": it must be an integral part of the risk-management system and be actively used in the running of the firm. Moreover, a firm must undertake an ORSA as described below.

Market-consistent valuation. In Solvency II the valuation must be carried out according to market-consistent principles. Where possible it should be based on actual *market values*, in a process known as *marking-to-market*. In a Solvency II glossary provided by the Comité Européen des Assurances and the Groupe Consultatif in 2007, market value is defined as:

> The amount for which an asset could be exchanged or a liability settled, between knowledgeable, willing parties in an arm's length transaction, based on observable prices within an active, deep and liquid market which is available to and generally used by the entity.

The concept of market value is related to the concept of *fair value* in accounting, and the principles adopted in Solvency II valuation have been influenced by International Financial Reporting Standards (IFRS) accounting standards. When no relevant market values exist (or when they do not meet the quality criteria described by the concept of an "active, deep and liquid market"), then market-consistent valuation requires the use of models that are calibrated, as far as possible, to be consistent with financial market information, a process known as *marking-to-model*; we discuss these ideas in more detail in Section 2.2.2.

The market-consistent valuation of the liabilities of an insurer is possible when the cash flows paid to policyholders can be fully replicated by the cash flows generated by the so-called matching assets that are held for that purpose; the value of the liability is then given by the value of the *replicating portfolio* of matching assets. However, it is seldom the case that liabilities can be fully replicated and hedged; mortality risk is a good example of a risk factor that is difficult to hedge.

The valuation of the unhedgeable part of a firm's liabilities is carried out by computing the sum of a *best estimate* of these liabilities (basically an expected value) plus an extra *risk margin* to cover some of the uncertainty in the value of the liability. The idea of the risk margin is that a third party would not be willing to take over the unhedgeable liability for a price set at the best estimate but would have to be further compensated for absorbing the additional uncertainty about the true value of the liability.

Standard formula approach. Under this approach an insurer calculates capital charges for different kinds of risk within a series of *modules*. There are modules, for example, for market risk, counterparty default risk, life underwriting risk, non-life underwriting risk and health insurance risk. The risk charges arising from these modules are aggregated to obtain the solvency capital requirement using a formula that involves a set of prescribed correlation parameters (see Section 8.4.2).

Within each module, the approach drills down to fundamental risk factors; for example, within the market-risk module, there are sub-modules relating to interest-rate risk, equity risk, credit-spread risk and other typical market-risk factors. Capital charges are calculated with respect to each risk factor by considering the effect of a series of defined stress scenarios on the value of net assets (assets minus liabilities). The stress scenarios are intended to represent 1-in-200-year events (i.e. events with an annual probability of 0.5%).

The capital charges for each risk factor are aggregated to obtain the module risk charge using a similar kind of formula to the one used at the highest level. Once again, a set of correlations expresses the regulatory view of dependencies between the effects of the fundamental risk factors. The details are complex and run to many pages, but the approach is simple and highly prescriptive.

Internal-model approach. Under this approach firms can develop an internal model for the financial and underwriting risk factors that affect their business; they may then seek regulatory approval to use this model in place of the standard formula. The model often takes the form of a so-called *economic scenario generator* in which risk-factor scenarios for a one-year period are randomly generated and applied to the assets and liabilities to determine the solvency capital requirement. Economic scenario generators vary greatly in their detail, ranging from simple distributional models to more sophisticated dynamic models in discrete or continuous time.

ORSA. In a 2008 Issues Paper produced by CEIOPS, the ORSA is described as follows:

> The entirety of the processes and procedures employed to identify, assess, monitor, manage, and report the short and long term risks a (re)insurance undertaking faces or may face and to determine the own funds necessary to ensure that the undertaking's overall solvency needs are met at all times.

The concept of an ORSA is not unique to Solvency II and a useful alternative definition has been provided by the NAIC in the US on its website:

> In essence, an ORSA is an internal process undertaken by an insurer or insurance group to assess the adequacy of its risk management and current and prospective solvency positions under normal and severe stress scenarios. An ORSA will require insurers to analyze all reasonably foreseeable and relevant material risks (i.e., underwriting, credit, market, operational, liquidity risks, etc.) that could have an impact on an insurer's ability to meet its policyholder obligations.

The Pillar 2 ORSA is distinguished from the Pillar 1 capital calculations in a number of ways. First, the definition makes clear that the ORSA refers to a process, or set of processes, and not simply an exercise in regulatory compliance. Second, each firm's ORSA is its *own* process and is likely to be *unique*, since it is not bound by a common set of rules. In contrast, the standard-formula approach to Pillar 1 is clearly a uniform process for all companies; moreover, firms that seek internal-model approval for Pillar 1 are subject to very similar constraints.

Finally, the ORSA goes beyond the one-year time horizon (which is a limitation of Pillar 1) and forces firms to assess solvency over their business planning horizon, which can mean many years for typical long-term business lines, such as life insurance.

1.3.3 Criticism of Regulatory Frameworks

The benefits arising from the regulation of financial services are not generally in doubt. Customer-protection acts, responsible corporate governance, fair and comparable accounting rules, transparent information on risk, capital and solvency for shareholders and clients are all viewed as positive developments.

Very few would argue the extreme position that the prudential regulatory frameworks we have discussed are not needed; in general, after a crisis, the demand (at least from the public and politicians) is for more regulation. Nevertheless, there are aspects of the regulatory frameworks that have elicited criticism, as we now discuss.

Cost and complexity. The cost factor of setting up a well-functioning risk-management system compliant with the present regulatory framework is significant, especially (in relative terms) for smaller institutions. On 27 March 2013, the *Financial Times* quoted Andrew Bailey (head of the Prudential Regulatory Authority in the UK) as saying that Solvency II compliance was set to cost UK companies at least £3 billion, a "frankly indefensible" amount. Related to the issue of cost is the belief that regulation, in its attempt to become more risk sensitive, is becoming too complex; this theme is taken up by the Basel Committee in their 2013 discussion paper entitled "The regulatory framework: balancing risk sensitivity, simplicity and comparability" (Basel Committee on Banking Supervision 2013b).

Endogenous risk. In general terms, this refers to the risk that is generated within a system and amplified by the system due to feedback effects. Regulation, a feature of the system, may be one of the channels by which shocks are amplified.

Regulation can lead to *risk-management herding*, whereby institutions following similar (perhaps VaR-based) rules may all be "running for the same exit" in times of crisis, consequently destabilizing an already precarious situation even further. This herding phenomenon has been suggested in connection with the 1987 stock market crash and the events surrounding the 1998 LTCM crisis (Daníelsson et al. 2001b).

An even more compelling example was observed during the 2007–9 crisis; to comply with regulatory capital ratios in a market where asset values were falling and risks increasing, firms adjusted their balance sheets by selling assets, causing further asset value falls and vanishing market liquidity. This led to criticism of the inherently *procyclical* nature of the Basel II regulation, whereby capital requirements may rise in times of stress and fall in times of expansion; the Basel III proposals attempt to address this issue with a countercyclical capital buffer.

Consequences of fair-value accounting and market-consistent valuation. The issue of procyclicality is also related to the widespread use of fair-value accounting and market-consistent valuation, which are at the heart of both the Basel rules for the trading book and the Solvency II framework. The fact that capital requirements are so closely coupled to volatile financial markets has been another focus of criticism.

An example of this is the debate around the valuation of insurance liabilities in periods of market stress. A credit crisis, of the kind experienced in 2007–9, can impact the high-quality corporate bonds that insurance companies hold on the asset side of their balance sheets. The relative value of corporate bonds compared with safe government bonds can fall sharply as investors demand more compensation for taking on both the credit risk and, in particular, the liquidity risk of corporate bonds.

The effect for insurers is that the value of their assets falls relative to the value of their liabilities, since the latter are valued by comparing cash flows with safe government bonds. At a particular point in time, an insurer may appear to have insufficient capital to meet solvency capital requirements. However, if an insurer has matched its asset and liability cash flows and can continue to meet its contractual obligations to policyholders, the apparent depletion of capital may not be a problem; insurance is a long-term business and the insurer has no short-term need to sell assets or offload liabilities, so a loss of capital need not be realized unless some of the bonds actually default.

Regulation that paints an unflattering picture of an insurer's solvency position is not popular with regulated firms. Firms have argued that they should be able to value liabilities at a lower level, by comparing the cash flows not with expensive government bonds but instead with the corporate bonds that are actually used as matching assets, making allowance only for the credit risk in corporate bonds. This has given rise to the idea of discounting with an extra *illiquidity premium*, or *matching premium*, above a risk-free rate. There has been much debate about this issue between those who feel that such proposals undermine market-consistent valuation and those who believe that strict adherence to market-consistent valuation overstates risk and has potential systemic consequences (see, for example, Wüthrich 2011).

Limits to quantification. Further criticism has been levelled at the highly quantitative nature of regulation and the extensive use of mathematical and statistical methods. The section on "Misplaced reliance on sophisticated mathematics" in the Turner Review of the global banking crisis (Lord Turner 2009) states that:

> The very complexity of the mathematics used to measure and manage risk, moreover, made it increasingly difficult for top management and boards to assess and exercise judgement over the risk being taken. Mathematical sophistication ended up not containing risk, but providing false assurances that other prima facie indicators of increasing risk (e.g. rapid credit extension and balance sheet growth) could be safely ignored.

This idea that regulation can lead to overconfidence in the quality of statistical risk measures is related to the view that the essentially backward-looking nature of estimates derived from historical data is a weakness. The use of conventional VaR-based methods has been likened to driving a car while looking in the rear-view mirror, the idea being that this is of limited use in preparing for the shocks that lie ahead.

The extension of the quantitative approach to operational risk has been controversial. Whereas everyone agrees that risks such as people risk (e.g. incompetence, fraud), process risk (e.g. model, transaction and operational control risk), technology risk (e.g. system failure, programming error) and legal risk are important, there is much disagreement on the extent to which these risks can be measured.

Limits to the efficacy of regulation. Finally, there is some debate about whether or not tighter regulation can ever prevent the occurrence of crises like that of 2007–9. The sceptical views of central bankers and regulatory figures were reported in the *Economist* in an article entitled "The inevitability of instability" (25 January 2014) (see also Prates 2013). The article suggests that "rules are constantly overtaken by financial innovation" and refers to the economist J. K. Galbraith (1993), who wrote:

> All financial innovation involves, in one form or another, the creation of debt secured in greater or lesser adequacy by real assets. . . . All crises have involved debt that, in one fashion or another, has become dangerously out of scale in relation to the underlying means of payment.

Tightening up the capital treatment of securitizations may prevent a recurrence of the events surrounding the 2007–9 crisis, but, according to the sceptical view, it will not prevent different forms of debt-fuelled crisis in the future.

1.4 Why Manage Financial Risk?

An important issue that we have barely touched upon is the reason for investing in risk management in the first place. This question can be addressed from various perspectives, including those of the customer of a financial institution, its shareholders, its management, its board of directors, regulators, politicians, or the general public; each of these stakeholders may have a different view. In the selective account we give here, we focus on two viewpoints: that of society as a whole, and that of the shareholders (owners) of a firm.

1.4.1 A Societal View

Modern society relies on the smooth functioning of banking and insurance systems, and it has a collective interest in the stability of such systems. The regulatory process that has given us the Basel and Solvency II frameworks was initially motivated by the desire to prevent the insolvency of individual institutions, thus protecting customers and policyholders; this is sometimes referred to as a *microprudential* approach. However, the reduction of systemic risk—the danger that problems in a single financial institution may spill over and, in extreme situations, disrupt the normal functioning of the entire financial system—has become an important secondary focus, particularly since the 2007–9 crisis. Regulation therefore now also takes a *macroprudential* perspective.

Most members of society would probably agree that protection of customers against the failure of an individual firm is an important aim, and there would be widespread agreement that the promotion of financial stability is vital. However, it is not always clear that the two aims are well aligned. While there are clearly situations where the failure of one company may lead to spillover effects that result in a systemic crisis, there may also be situations where the long-term interests of financial stability are better served by allowing a company to fail: it may provide a lesson in the importance of better risk management for other companies. This issue is clearly related to the *systemic importance* of the company in question: in other words, to its size and the extent of its connectivity to other firms. But the recognition that there may be firms that are too important or are *too big to fail* creates a *moral hazard*, since the management of such a firm may take more risk in the knowledge that the company would be bailed out in a crisis. Of course, it may be the case that in some countries some institutions are also *too big to save*.

The 2007–9 crisis provided a case study that brought many of these issues to the fore. As we noted in our account of the crisis in Section 1.2, it was initially believed that the growth in securitization was dispersing credit risk throughout the system and was beneficial to financial stability. But the warehousing of vast amounts of inadequately capitalized credit risk (in the form of CDOs) in trading books, combined with the interconnectedness of banks through derivatives and interbank lending activities, meant that quite the opposite was true. The extent of the systemic risk that had been accumulating became apparent when Lehman Brothers filed for bankruptcy on 15 September 2008 and governments intervened to bail out the banks.

It was the following phase of the crisis during which society suffered. The world economy went into recession, households defaulted on their debts, and savings and pensions were hit hard. The crisis moved "from Wall Street to Main Street". Naturally, this led to resentment as banking remained a highly rewarded profession and it seemed that the government-sponsored bailouts had allowed banks "to privatize their gains and socialize their losses".

There has been much debate since the crisis on whether the US government could have intervened to save Lehman, as it did for other firms such as AIG. In the *Financial Times* on 14 September 2009, the historian Niall Ferguson wrote:

> Like the executed British admiral in Voltaire's famous phrase, Lehman had to die *pour encourager les autres*—to convince the other banks that they needed injections of public capital, and to convince the legislature to approve them. Not everything in history is inevitable; contingencies abound. Sometimes it is therefore right to say "if only". But an imagined rescue of Lehman Brothers is the wrong counterfactual. The right one goes like this. If only Lehman's failure and the passage of TARP had been followed—not immediately, but after six months—by a clear statement to the surviving banks that none of them was henceforth too big to fail, then we might actually have learnt something from this crisis.

While it is difficult to speak with authority for "society", the following conclusions do not seem unreasonable. The interests of society are served by enforcing the discipline of risk management in financial firms, through the use of regulation. Better risk management can reduce the risk of company failure and protect customers and policyholders who stand in a very unequal financial relationship with large firms. However, the regulation employed must be designed with care and should not promote herding, procyclical behaviour or other forms of endogenous risk that could result in a systemic crisis with far worse implications for society than the failure of a single firm. Individual firms need to be allowed to fail on occasion, provided customers can be shielded from the worst consequences through appropriate compensation schemes. A system that allows firms to become too big to fail creates moral hazard and should be avoided.

1.4.2 The Shareholder's View

It is widely believed that proper financial risk management can increase the value of a corporation and hence shareholder value. In fact, this is the main reason why corporations that are not subject to regulation by financial supervisory authorities engage in risk-management activities. Understanding the relationship between shareholder value and financial risk management also has important implications for the design of risk-management systems. Questions to be answered include the following.

- When does risk management increase the value of a firm, and which risks should be managed?
- How should risk-management concerns factor into investment policy and capital budgeting?

There is a rather extensive corporate-finance literature on the issue of "corporate risk management and shareholder value". We briefly discuss some of the main arguments. In this way we hope to alert the reader to the fact that there is more to risk management than the mainly technical questions related to the implementation of risk-management strategies dealt with in the core of this book.

The first thing to note is that from a corporate-finance perspective it is by no means obvious that in a world with perfect capital markets risk management enhances shareholder value: while *individual* investors are typically risk averse and should therefore manage the risk in their portfolios, it is not clear that risk management or risk reduction at the *corporate level*, such as hedging a foreign-currency exposure or holding a certain amount of risk capital, increases the value of a corporation. The rationale for this (at first surprising) observation is simple: if investors have access to perfect capital markets, they can do the risk-management transactions they deem necessary via their own trading and diversification. The following statement from the chief investment officer of an insurance company exemplifies this line of reasoning: "If our shareholders believe that our investment portfolio is too risky, they should short futures on major stock market indices."

The potential irrelevance of corporate risk management for the value of a corporation is an immediate consequence of the famous *Modigliani–Miller Theorem* (Modigliani and Miller 1958). This result, which marks the beginning of modern corporate-finance theory, states that, in an ideal world without taxes, bankruptcy costs and informational asymmetries, and with frictionless and arbitrage-free capital markets, the financial structure of a firm, and hence also its risk-management decisions, are irrelevant when assessing the firm's value. Hence, in order to find reasons for corporate risk management, one has to "turn the Modigliani–Miller Theorem upside down" and identify situations where risk management enhances the value of a firm by deviating from the unrealistically strong assumptions of the theorem. This leads to the following rationales for risk management.

- Risk management can reduce *tax costs*. Under a typical tax regime the amount of tax to be paid by a corporation is a *convex* function of its profits; by reducing the variability in a firm's cash flow, risk management can therefore lead to a higher expected after-tax profit.

- Risk management can be beneficial, since a company may (and usually will) have better access to capital markets than individual investors.

- Risk management can increase firm value in the presence of *bankruptcy costs*, as it makes bankruptcy less likely.

- Risk management can reduce the impact of *costly external financing* on the firm value, as it facilitates the achievement of optimal investment.

The last two points merit a more detailed discussion. Bankruptcy costs consist of direct bankruptcy costs, such as the cost of lawsuits, and the more important indirect bankruptcy costs. The latter may include liquidation costs, which can be substantial in the case of intangibles like research and development and knowhow. This is why high research and development spending appears to be positively correlated with the use of risk-management techniques. Moreover, increased likelihood of bankruptcy often has a negative effect on key employees, management and customer relations, in particular in areas where a client wants a long-term business relationship. For instance, few customers would want to enter into a life insurance contract with an insurance company that is known to be close to bankruptcy. On a related note, banks that are close to bankruptcy might be faced with the unpalatable prospect of a bank run, where depositors try to withdraw their money simultaneously. A further discussion of these issues is given in Altman (1993).

It is a "stylized fact" of corporate finance that for a corporation, external funds are more costly to obtain than internal funds, an observation which is usually attributed to problems of asymmetric information between the management of a corporation and bond and equity investors. For instance, raising external capital from outsiders by issuing new shares might be costly if the new investors, who have incomplete information about the economic prospects of a firm, interpret the share issue as a sign that the firm is overvalued. This can generate a rationale for risk management for the following reason: without risk management the increased variability of a company's cash flow will be translated either into an increased variability of the funds that need to be raised externally or to an increased variability in the amount of investment. With increasing marginal costs of raising external capital and decreasing marginal profits from new investment, we are left with a decrease in (expected) profits. Proper risk management, which amounts to a smoothing of the cash flow generated by a corporation, can therefore be beneficial. For references to the literature see Notes and Comments below.

1.5 Quantitative Risk Management

The aim of this chapter has been to place QRM in a larger historical, regulatory and even societal framework, since a study of QRM without a discussion of its proper setting and motivation makes little sense. In the remainder of the book we adopt a somewhat narrower view and treat QRM as a quantitative science that uses the language of mathematics in general, and of probability and statistics in particular.

In this section we discuss the relevance of the Q in QRM, describe the quantitative modelling challenge that we have attempted to meet in this book, and end with thoughts on where QRM may lead in the future.

1.5.1 The Q in QRM

In Section 1.2.1 we discussed the view that the use of advanced mathematical modelling and valuation techniques has been a contributory factor in financial crises, particularly those attributed to derivative products, such as CDOs in the 2007–9 crisis. We have also referred to criticism of the quantitative, statistical emphasis of the modern regulatory framework in Section 1.3.3. These arguments must be taken seriously, but we believe that it is neither possible nor desirable to remove the quantitative element from risk management.

Mathematics and statistics provide us with a suitable language and appropriate concepts for describing financial risk. This is clear for complex financial products such as derivatives, which cannot be valued and handled without mathematical models. But the need for quantitative modelling also arises for simpler products, such as a book of mortgages for retail clients. The main risk in managing such a book is the occurrence of disproportionately many defaults: a risk that is directly related to the dependence between defaults (see Chapter 11 for details). In order to describe this dependence, we need mathematical concepts from multivariate statistics, such as correlations or copulas; if we want to carry out a simulation study of the behaviour of the portfolio under different economic scenarios, we need a mathematical model that describes the joint distribution of default events; if the portfolio is large, we will also need advanced simulation techniques to generate the relevant scenarios efficiently.

Moreover, mathematical and statistical methods can do better than they did in the 2007–9 crisis. In fact, providing concepts, techniques and tools that address some of the weaker points of current methodology is a main theme of our text and we come back to this point in the next section.

There is a view that, instead of using mathematical models, there is more to be learned about risk management through a *qualitative* analysis of historical case studies and the formulation of narratives. What is often overlooked by the non-specialist is that mathematical models are themselves nothing more than narratives, albeit narratives couched in a precise symbolic language. Addressing the question "What is mathematics?", Gale and Shapley (1962) wrote: "Any argument which is carried out with sufficient precision is mathematical." Lloyd Shapley went on to win the 2012 Nobel Memorial Prize in Economic Science.

It is certainly true that mathematical methods can be misused. Mathematicians are very well aware that a mathematical result has not only a conclusion but, equally importantly, certain conditions under which it holds. Statisticians are well aware that inductive reasoning on the basis of models relies on the assumption that these conditions hold in the real world. This is especially true in economics, which as a social science is concerned with phenomena that are not easily described by clear mathematical or physical laws. By starting with questionable assumptions, models can be used (or manipulated) to deliver bad answers. In a talk on 20 March 2009, the economist Roger Guesnerie said, "For this crisis, mathematicians are innocent . . . and this in both meanings of the word." The implication is that quantitative risk managers must become more worldly about the ways in which models are used. But equally, the regulatory system needs to be more vigilant about the ways in which models can be gamed and the institutional pressures that can circumvent the best intentions of prudent quantitative risk managers.

We are firmly of the opinion—an opinion that has only been reinforced by our study of financial crises—that the Q in QRM is an essential part of the process. We reject the idea that the Q is part of the problem, and we believe that it remains (if applied correctly and honestly) a part of the solution to managing risk. In summary, we strongly agree with Shreve (2008), who said:

> Don't blame the quants. Hire good ones instead and listen to them.

1.5.2 The Nature of the Challenge

When we began this book project we set ourselves the task of defining a new discipline of QRM. Our approach to this task has had two main strands. On the one hand, we have attempted to put current practice onto a firmer mathematical footing, where, for example, concepts like P&L distributions, risk factors, risk measures, capital allocation and risk aggregation are given formal definitions and a consistent notation. In doing this we have been guided by the consideration of what topics should form the core of a course on QRM for a wide audience of students interested in risk-management issues; nonetheless, the list is far from complete and will continue to evolve as the discipline matures. On the other hand, the second strand of our endeavour has been to put together material on techniques and tools that go beyond current practice and address some of the deficiencies that have been repeatedly raised by critics. In the following paragraphs we elaborate on some of these issues.

Extremes matter. A very important challenge in QRM, and one that makes it particularly interesting as a field for probability and statistics, is the need to address unexpected, abnormal or extreme outcomes, rather than the expected, normal or average outcomes that are the focus of many classical applications. This is in tune with the regulatory view expressed by Alan Greenspan in 1995 at the Joint Central Bank Research Conference:

> From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be

> balanced against the significant advantage of simplification. From the central bank's corner, the consequences are even more serious because we often need to concentrate on the left tail of the distribution in formulating lender-of-last-resort policies. Improving the characterization of the distribution of extreme values is of paramount importance.

While the quote is older, the same concern about underestimation of extremes is raised in a passage in the Turner Review (Lord Turner 2009):

> Price movements during the crisis have often been of a size whose probability was calculated by models (even using longer-term inputs) to be almost infinitesimally small. This suggests that the models systematically underestimated the chances of small probability high impact events.... It is possible that financial market movements are inherently characterized by fat-tail distributions. VaR models need to be buttressed by the application of stress test techniques which consider the impact of extreme movements beyond those which the model suggests are at all probable.

Much space in our book is devoted to models for financial risk factors that go beyond the normal (or Gaussian) model and attempt to capture the related phenomena of heavy or fat tails, excess volatility and extreme values.

The interdependence and concentration of risks. A further important challenge is presented by the multivariate nature of risk. Whether we look at market risk or credit risk, or overall enterprise-wide risk, we are generally interested in some form of aggregate risk that depends on high-dimensional vectors of underlying risk factors, such as individual asset values in market risk or credit spreads and counterparty default indicators in credit risk.

A particular concern in our multivariate modelling is the phenomenon of dependence between extreme outcomes, when many risk factors move against us simultaneously. In connection with the LTCM case (see Section 1.2.1) we find the following quote in *Business Week* (September 1998):

> Extreme, synchronized rises and falls in financial markets occur infrequently but they do occur. The problem with the models is that they did not assign a high enough chance of occurrence to the scenario in which many things go wrong at the same time—the "perfect storm" scenario.

In a perfect storm scenario the risk manager discovers that portfolio diversification arguments break down and there is much more of a concentration of risk than had been imagined. This was very much the case with the 2007–9 crisis: when borrowing rates rose, bond markets fell sharply, liquidity disappeared and many other asset classes declined in value, with only a few exceptions (such as precious metals and agricultural land), a perfect storm was created.

We have mentioned (see Section 1.2.1) the notorious role of the Gauss copula in the 2007–9 financial crisis. An April 2009 article in the *Economist*, with the title

"In defence of the Gaussian copula", evokes the environment at the time of the securitization boom:

> By 2001, correlation was a big deal. A new fervour was gripping Wall Street—one almost as revolutionary as that which had struck when the Black–Scholes model brought about the explosion in stock options and derivatives in the early 1980s. This was structured finance, the culmination of two decades of quants on Wall Street.... The problem, however, was correlation. The one thing any off-balance-sheet securitisation could not properly capture was the interrelatedness of all the hundreds of thousands of different mortgage loans they owned.

The Gauss copula appeared to solve this problem by offering a model for the correlated times of default of the loans or other credit-risky assets; the perils of this approach later became clear. In fact, the Gauss copula is not an example of the use of oversophisticated mathematics; it is a relatively simple model that is difficult to calibrate reliably to available market information. The modelling of dependent credit risks, and the issue of model risk in that context, is a subject we look at in some detail in our treatment of credit risk.

The problem of scale. A further challenge in QRM is the typical scale of the portfolios under consideration; in the most general case, a portfolio may represent the entire position in risky assets of a financial institution. Calibration of detailed multivariate models for all risk factors is an almost impossible task, and any sensible strategy must involve dimension reduction; that is to say, the identification of key risk drivers and a concentration on modelling the main features of the overall risk landscape.

In short, we are forced to adopt a fairly broad-brush approach. Where we use econometric tools, such as models for financial return series, we are content with relatively simple descriptions of individual series that capture the main phenomenon of volatility, and which can be used in a parsimonious multivariate factor model. Similarly, in the context of portfolio credit risk, we are more concerned with finding suitable models for the default dependence of counterparties than with accurately describing the mechanism for the default of an individual, since it is our belief that the former is at least as important as the latter in determining the risk of a large diversified portfolio.

Interdisciplinarity. Another aspect of the challenge of QRM is the fact that ideas and techniques from several existing quantitative disciplines are drawn together. When one considers the ideal education for a quantitative risk manager of the future, then a combined quantitative skill set should undoubtedly include concepts, techniques and tools from such fields as mathematical finance, statistics, financial econometrics, financial economics and actuarial mathematics. Our choice of topics is strongly guided by a firm belief that the inclusion of modern statistical and econometric techniques and a well-chosen subset of actuarial methodology are essential for the establishment of best-practice QRM. QRM is certainly not just about financial mathematics and derivative pricing, important though these may be.

Communication and education. Of course, the quantitative risk manager operates in an environment where additional non-quantitative skills are equally important. Communication is certainly an important skill: risk professionals, by the definition of their duties, will have to interact with colleagues with diverse training and backgrounds, at all levels of their organization. Moreover, a quantitative risk manager has to familiarize him or herself quickly with all-important market practice and institutional details. A certain degree of humility will also be required to recognize the role of QRM in a much larger picture.

A lesson from the 2007–9 crisis is that improved education in QRM is essential; from the front office to the back office to the boardroom, the users of models and their outputs need to be better trained to understand model assumptions and limitations. This task of educating users is part of the role of a quantitative risk manager, who should ideally have (or develop) the pedagogical skills to explain methods and conclusions to audiences at different levels of mathematical sophistication.

1.5.3 QRM Beyond Finance

The use of QRM technology is not restricted to the financial services industry, and similar developments have taken place, or are taking place, in other sectors of industry. Some of the earliest applications of QRM are to be found in the manufacturing industry, where similar concepts and tools exist under names like reliability or total quality control. Industrial companies have long recognized the risks associated with bringing faulty products to the market. The car manufacturing industry in Japan, in particular, was an early driving force in this respect.

More recently, QRM techniques have been adopted in the transport and energy industries, to name but two. In the case of energy, there are obvious similarities with financial markets: electrical power is traded on energy exchanges; derivatives contracts are used to hedge future price uncertainty; companies optimize investment portfolios combining energy products with financial products; some Basel methodology can be applied to modelling risk in the energy sector. However, there are also important dissimilarities due to the specific nature of the industry; most importantly, there are the issues of the cost of storage and transport of electricity as an underlying commodity, and the necessity of modelling physical networks including the constraints imposed by the existence of national boundaries and quasi-monopolies.

There are also markets for environmental emission allowances. For example, the Chicago Climate Futures Exchange offers futures contracts on sulphur dioxide emissions. These are traded by industrial companies producing the pollutant in their manufacturing process, and they force such companies to consider the cost of pollution as a further risk in their risk landscape.

A natural consequence of the evolution of QRM thinking in different industries is an interest in the transfer of risks between industries; this process is known as alternative risk transfer. To date the best examples of risk transfer are between the insurance and banking industries, as illustrated by the establishment of catastrophe futures by the Chicago Board of Trade in 1992. These came about in the wake of Hurricane Andrew, which caused $20 billion of insured losses on the East Coast of the US. While this was a considerable event for the insurance industry in relation to overall reinsurance capacity, it represented only a drop in the ocean compared with the daily volumes traded worldwide on financial exchanges. This led to the recognition that losses could be covered in future by the issuance of appropriately structured bonds with coupon streams and principal repayments dependent on the occurrence or non-occurrence of well-defined natural catastrophe events, such as storms and earthquakes.

A speculative view of where these developments may lead is given by Shiller (2003), who argues that the proliferation of risk-management thinking coupled with the technological sophistication of the twenty-first century will allow any agent in society, from a company to a country to an individual, to apply QRM methodology to the risks they face. In the case of an individual this may be the risk of unemployment, depreciation in the housing market or investment in the education of children.

Notes and Comments

The language of probability and statistics plays a fundamental role throughout this book, and readers are expected to have a good knowledge of these subjects. At the elementary level, Rice (1995) gives a good first introduction to both. More advanced texts in probability and stochastic processes are Williams (1991), Resnick (1992) and Rogers and Williams (1994); the full depth of these texts is certainly not required for the understanding of this book, though they provide excellent reading material for more mathematically sophisticated readers who also have an interest in mathematical finance. Further recommended texts on statistical inference include Casella and Berger (2002), Bickel and Doksum (2001), Davison (2003) and Lindsey (1996).

In our discussion of risk and randomness in Section 1.1.1 we mentioned Knight (1921) and Keynes (1920), whose classic texts are very much worth revisiting. Knightian uncertainty refers to uncertainty that cannot be measured and is sometimes contrasted with risks that can be measured using probability. This relates to the more recent idea of a Black Swan event, a term popularized in Taleb (2007) but introduced in Taleb (2001). Black swans were believed to be imaginary creatures until the European exploration of Australia and the name is applied to unprecedented and unpredictable events that challenge conventional beliefs and models. Donald Rumsfeld, a former US Secretary of Defense, referred to “unknown unknowns” in a 2002 news briefing on the evidence for the presence of weapons of mass destruction in Iraq.

An excellent text on the history of risk and probability with financial applications in mind is Bernstein (1998). We also recommend Shiller (2012) for more on the societal context of financial risk management. A thought-provoking text addressing risk on Wall Street from a historical perspective is Brown (2012).

For the mathematical reader looking to acquire more knowledge about the relevant economics we recommend Mas-Colell, Whinston and Green (1995) for microeconomics, Campbell, Lo and MacKinlay (1997) or Gouriéroux and Jasiak (2001) for econometrics, and Brealey and Myers (2000) for corporate finance. From the vast literature on options, an entry-level text for the general reader is Hull (2014). At a more mathematical level we like Bingham and Kiesel (2004), Musiela and Rutkowski (1997), Shreve (2004a) and Shreve (2004b). One of the most readable texts on the basic notion of options is Cox and Rubinstein (1985). For a rather extensive list of the kind of animals to be found in the zoological garden of derivatives, see, for example, Haug (1998).

There are several texts on the spectacular losses that occurred as the result of speculative trading and the careless use of derivatives. For a historical overview of financial crises, see Reinhart and Rogoff (2009), as well as the much earlier Galbraith (1993) and Kindleberger (2000). Several texts exist on more recent crises; we list only a few. The LTCM case is well documented in Dunbar (2000), Lowenstein (2000) and Jorion (2000), the latter particularly focusing on the technical risk-measurement issues involved. Boyle and Boyle (2001) give a very readable account of the Orange County, Barings and LTCM stories (see also Jacque 2010). For the Equitable Life case see the original Penrose Report, published by the UK government (Lord Penrose 2004), or an interesting paper by Roberts (2012). Many books have emerged on the 2007–9 crisis; early warnings are well summarized, under Greenspan's memorable "irrational exuberance" phrase, in a pre-crisis book by Shiller (2000), and the post-mortem by the same author is also recommended (Shiller 2008).

An overview of options embedded in life insurance products is given in Dillmann (2002), guarantees are discussed in detail in Hardy (2003), and Briys and de Varenne (2001) contains an excellent account of risk-management issues facing the (life) insurance industry. For risk-management and valuation issues underlying life insurance, see Koller (2011) and Møller and Steffensen (2007). Market-consistent actuarial valuation is discussed in Wüthrich, Bühlmann and Furrer (2010).

The historical development of banking regulation is well described in Crouhy, Galai and Mark (2001) and Steinherr (1998). For details of the current rules and regulations coming from the Basel Committee, see its website at www.bis.org/bcbs. Besides copies of the various accords, one can also find useful working papers, publications and comments written by stakeholders on the various consultative packages. For Solvency II and the Swiss Solvency Test, many documents are to be found on the web. Comprehensive textbook accounts are Sandström (2006) and Sandström (2011), and a more technical treatment is found in Wüthrich and Merz (2013). The complexity of risk-management methodology in the wake of Basel II is critically addressed by Hawke (2003), from his perspective as US Comptroller of the Currency. Among the numerous texts written after the 2007–9 crisis, we found all of Rochet (2008), Shin (2010), Dewatripont, Rochet and Tirole (2010) and Bénéplanc and Rochet (2011) useful. For a discussion of issues related to the use of fair-value accounting during the financial crisis, see Ryan (2008).

For a very detailed overview of relevant practical issues underlying risk management, we again strongly recommend Crouhy, Galai and Mark (2001). A text stressing the use of VaR as a risk measure and containing several worked examples is Jorion (2007), whose author also has a useful teaching manual on the same subject

(Jorion 2002b). Insurance-related issues in risk management are nicely presented in Doherty (2000).

For a comprehensive discussion of the management of bank capital given regulatory constraints, see Matten (2000), Klaassen and van Eeghen (2009) and Admati and Hellwig (2013). Graham and Rogers (2002) contains a discussion of risk management and tax incentives. A formal account of the Modigliani–Miller Theorem and its implications can be found in many textbooks on corporate finance: a standard reference is Brealey and Myers (2000), and de Matos (2001) gives a more theoretical account from the perspective of modern financial economics. Both texts also discuss the implications of informational asymmetries between the various stakeholders in a corporation. Formal models looking at risk management from a corporate-finance angle are to be found in Froot and Stein (1998), Froot, Scharfstein and Stein (1993) and Stulz (1996, 2002). For a specific discussion on corporate-finance issues in insurance, see Froot (2007) and Hancock, Huber and Koch (2001).

There are several studies on the use of risk-management techniques for non-financial firms (see, for example, Bodnar, Hayt and Marston 1998; Geman 2005, 2009). Two references in the area of the reliability of industrial processes are Bedford and Cooke (2001) and Does, Roes and Trip (1999). Interesting edited volumes on alternative risk transfer are Shimpi (2001), Barrieu and Albertini (2009) and Kiesel, Scherer and Zagst (2010); a detailed study of model risk in the alternative risk transfer context is Schmock (1999). An area we have not mentioned so far in our discussion of QRM in the future is that of real options. A real option is the right, but not the obligation, to take an action (e.g. deferring, expanding, contracting or abandoning) at a predetermined cost called the exercise price. The right holds for a predetermined period of time—the life of the option. This definition is taken from Copeland and Antikarov (2001). Examples of real options discussed in the latter are the valuation of an internet project and of a pharmaceutical research and development project. A further useful reference is Brennan and Trigeorgis (1999).

A well-written critical view of the failings of the standard approach to risk management is given in Rebonato (2007). And finally, for an entertaining text on the biology of the much criticized “homo economicus”, we like Coates (2012).

2
Basic Concepts in Risk Management

In this chapter we define or explain a number of fundamental concepts used in the measurement and management of financial risk. Beginning in Section 2.1 with the simplified balance sheet of a bank and an insurer, we discuss the risks faced by such firms, the nature of capital, and the need for a firm to have sufficient capital to withstand financial shocks and remain solvent.

In Section 2.2 we establish a mathematical framework for describing changes in the value of portfolios and deriving loss distributions. We provide a number of examples to show how this framework applies to different kinds of asset and liability portfolios. The examples are also used to discuss the meaning of value in more detail with reference to fair-value accounting and risk-neutral valuation.

Section 2.3 is devoted to the subject of using risk measures to determine risk or solvency capital. We present different quantitative approaches to measuring risk, with a particular focus on risk measures that are calculated from loss distributions, like value-at-risk and expected shortfall.

2.1 Risk Management for a Financial Firm

2.1.1 Assets, Liabilities and the Balance Sheet

A good way to understand the risks faced by a modern financial institution is to look at the stylized balance sheet of a typical bank or insurance company. A balance sheet is a financial statement showing *assets and liabilities*; roughly speaking, the assets describe the financial institution's investments, whereas liabilities refer to the way in which funds have been raised and the obligations that ensue from that fundraising.

A typical bank raises funds by taking in customer deposits, by issuing bonds and by borrowing from other banks or from central banks. Collectively these form the *debt capital* of the bank, which is invested in a number of ways. Most importantly, it is used for loans to retail, corporate and sovereign customers, invested in traded securities, lent out to other banks or invested in property or in other companies. A small fraction is also held as cash.

A typical insurance company sells insurance contracts, collecting premiums in return, and raises additional funds by issuing bonds. The liabilities of an insurance company thus consist of its obligations to policyholders, which take the form of a *technical reserve against future claims*, and its obligations to bondholders. The funds raised are then invested in traded securities, particularly bonds, as well as other assets such as real estate.

Table 2.1. The stylized balance sheet of a typical bank.

Bank ABC (31 December 2015)			
Assets		Liabilities	
Cash (and central bank balance)	£10M	Customer deposits	£80M
Securities	£50M	Bonds issued	
– bonds		– senior bond issues	£25M
– stocks		– subordinated bond issues	£15M
– derivatives		Short-term borrowing	£30M
Loans and mortgages	£100M	Reserves (for losses on loans)	£20M
– corporates			
– retail and smaller clients		*Debt (sum of above)*	£170M
– government			
Other assets	£20M		
– property			
– investments in companies		*Equity*	£30M
Short-term lending	£20M		
Total	£200M	Total	£200M

In both cases a small amount of extra funding stems from occasional *share issues*, which form the share capital of the bank or insurer. This form of funding is crucial as it entails no obligation towards outside parties.

These simplified banking and insurance *business models* are reflected in the stylized balance sheets shown in Tables 2.1 and 2.2. In these financial statements, assets and liabilities are valued on a given date. The position marked *equity* on the liability side of the balance sheet is the residual value defined in the balance sheet equation

$$\text{value of assets} = \text{value of liabilities} = \text{debt} + \text{equity}. \tag{2.1}$$

A company is *solvent* at a given point in time if the equity is nonnegative; otherwise it is insolvent. Insolvency should be distinguished from the notion of default, which occurs if a firm misses a payment to its debtholders or other creditors. In particular, an otherwise-solvent company can default because of *liquidity* problems, as discussed in more detail in the next section.

It should be noted that assigning values to the items on the balance sheet of a bank or insurance company is a non-trivial task. Broadly speaking, two different approaches can be distinguished. The practice of *fair-value accounting* attempts to value assets at the prices that would be received if they were sold and to value liabilities at the prices that would have to be paid if they were transferred to another party. Fair-value accounting is relatively straightforward for positions that are close to securities traded on liquid markets, since these are simply valued by (an estimate of) their market price. It is more challenging to apply fair-value principles to non-traded or illiquid assets and liabilities.

The more traditional practice of *amortized cost accounting* is still applied to many kinds of financial asset and liability. Under this practice the position is assigned a

Table 2.2. The stylized balance sheet of a typical insurer.

Insurer XYZ (31 December 2015)			
Assets		Liabilities	
Investments		Reserves for policies written (technical provisions)	£80M
– bonds	£50M		
– stocks	£5M	Bonds issued	£10M
– real estate	£5M		
Investments for unit-linked contracts	£30M	*Debt (sum of above)*	£90M
Other assets – property	£10M		
		Equity	£10M
Total	£100M	Total	£100M

book value at its inception and this is carried forward over time. In some cases the value is progressively reduced or impaired to account for the aging of the position or the effect of adverse events. An example of assets valued at book value are the loans on the balance sheet of the bank. The book value would typically be an estimate of the present value (at the time the loans were made) of promised future interest and principal payments minus a provision for losses due to default.

In the European insurance industry the practice of *market-consistent valuation* has been promoted under the Solvency II framework. As described in Section 1.3.2, the rationale is very similar to that of fair-value accounting: namely, to value positions by "the amount for which an asset could be exchanged or a liability settled, between knowledgeable, willing parties in an arm's length transaction, based on observable prices within an active, deep and liquid market". However, there are some differences between market-consistent valuation and fair-value accounting for specific kinds of position. A European insurer will typically have two versions of the balance sheet in order to comply with accounting rules, on the one hand, and Solvency II rules for capital adequacy, on the other. The accounting balance sheet may mix fair-value and book-value approaches, but the Solvency II balance sheet will apply market-consistent principles throughout.

Overall, across the financial industry, there is a tendency for the accounting standard to move towards fair-value accounting, even if the financial crisis of 2007–9 demonstrated that this approach is not without problems during periods when trading activity and market liquidity suddenly vanish (see Section 1.3.3 for more discussion of this issue). Fair-value accounting for financial products will be discussed in more detail in Section 2.2.2.

2.1.2 Risks Faced by a Financial Firm

An obvious source of risk for a bank is a decrease in the value of its investments on the asset side of the balance sheet. This includes market risk, such as losses from securities trading, and credit risk. Another important risk is related to funding and so-called *maturity mismatch*: for a typical bank, large parts of the asset side consist of relatively illiquid, long-term investments such as loans or property, whereas a large part of the liabilities side consists of short-term obligations such as funds borrowed from money markets and most customer deposits. This may lead to problems when the cost of short-term refinancing increases due to rising short-term interest rates, because the banks may have difficulties selling long-term assets to raise funds. This can lead to the default of a bank that is technically solvent; in extreme cases there might even be a *bank run*, as was witnessed during the 2007–9 financial crisis. This clearly shows that risk is found on both sides of the balance sheet and that risk managers should not focus exclusively on the asset side.

The primary risk for an insurance company is clearly insolvency, i.e. the risk that the claims of policyholders cannot be met. This can happen due to adverse events affecting the asset side or the liability side of the balance sheet. On the asset side, the risks are similar to those for a bank. On the liability side, the main risk is that reserves are insufficient to cover future claim payments. It is important to bear in mind that the liabilities of a life insurer are of a long-term nature (due to the sale of products such as annuities) and are subject to many categories of risk including interest-rate risk, inflation risk and longevity risk, some of which also affect the asset side. An important aspect of the risk-management strategy of an insurance company is, therefore, to hedge parts of these risks by proper investment of the premium income (so-called liability-driven investment).

It should be clear from this discussion that a sound approach to risk management cannot look at one side of the balance sheet in isolation from the other.

2.1.3 Capital

There are many different notions of bank *capital*, and three broad concepts can be distinguished: *equity (or book) capital*, *regulatory capital* and *economic capital*. All of these notions of capital refer to items on the liability side of the balance sheet that entail no (or very limited) obligations to outside creditors and that can thus serve as a buffer against losses.

The equity capital can be read from the balance sheet according to the balance sheet equation in (2.1). It is therefore a measure of the value of the company to the shareholders. The balance sheet usually gives a more detailed breakdown of the equity capital by listing separate positions for *shareholder capital*, *retained earnings* and other items of lesser importance. Shareholder capital is the initial capital invested in the company by purchasers of equity. For companies financed by a single share issue, this is given by the numbers of shares issued multiplied by their price at the issuance date. Shareholder capital is therefore different from market capitalization, which is given by the number of shares issued multiplied by their current market price. Retained earnings are the accumulated earnings that have not been paid out in the form of dividends to shareholders; these can in principle be negative if the company has made losses.

Regulatory capital is the amount of capital that a company should have according to regulatory rules. For a bank, the rules are set out in the Basel framework, as described in more detail in Section 1.3.1. For European insurance companies, regulatory capital takes the form of a minimum capital requirement and a solvency capital requirement as set out in the Solvency II framework (see Section 1.3.2).

A regulatory capital framework generally specifies the amount of capital necessary for a financial institution to continue its operations, taking into account the size and the riskiness of its positions. Moreover, it specifies the quality of the capital and hence the form it should take on the balance sheet. In this context one usually distinguishes between different numbered capital *tiers*.

For example, in the Basel framework, Tier 1 capital is the sum of shareholder capital and retained earnings; in other words, the main constituents of the equity capital. This capital can act in full as a buffer against losses as there are no other claims on it. Tier 2 capital includes other positions of the balance sheet, in particular subordinated debt. Holders of this debt would effectively be the last to be paid before the shareholders in the event of the liquidation of the company, so subordinated debt can be viewed as an extra layer of protection for depositors and other senior debtholders. For illustration, the bank in Table 2.1 has Tier 1 capital of £30 million (assuming the equity capital consists of shareholder capital and retained earnings only) and Tier 2 capital of £45 million.

Economic capital is an estimate of the amount of capital that a financial institution needs in order to control the probability of becoming insolvent, typically over a one-year horizon. It is an internal assessment of risk capital that is guided by economic modelling principles. In particular, an economic capital framework attempts to take a holistic view that looks at assets and liabilities simultaneously, and works, where possible, with fair or market-consistent values of balance sheet items. Although, historically, regulatory capital frameworks have been based more on relatively simple rules and on book values for balance sheet items, there is increasing convergence between the economic and regulatory capital concepts, particularly in the insurance world, where Solvency II emphasizes market-consistent valuation of liabilities.

Note that the various notions of capital refer to the way in which a financial firm finances itself and not to the assets it invests in. In particular, capital requirements do not require the setting aside of funds that cannot be invested productively, e.g. by issuing new loans. There are other forms of financial regulation that refer to the asset side of the balance sheet and restrict the investment possibilities, such as obligatory cash reserves for banks and constraints on the proportion of insurance assets that may be invested in stocks.

Notes and Comments

A good introduction to the business of banking and the risks affecting banks is Choudhry (2012), while Thoyts (2010) provides a very readable overview of theory and practice in the insurance industry, with a focus on the UK. Readers wanting to go deeper into the subject of balance sheets have many financial accounting textbooks to choose from, a popular one being Elliott and Elliott (2013). A paper that gives more explanation of fair-value accounting and also discusses issues raised by the financial crisis is Ryan (2008).

Regulatory capital in the banking industry is covered in many of the documents produced by the Basel Committee, in particular the papers covering the Basel II and Basel III capital frameworks (Basel Committee on Banking Supervision 2006, 2011). For regulatory capital under Solvency II, see Sandström (2011). Textbook treatments of the management of bank capital given regulatory constraints are found in Matten (2000) and Klaassen and van Eeghen (2009), while Admati et al. (2013) provides a strong argument for capital regulation that ensures banks have a high level of equity capital. This issue is discussed at a slightly less technical level in the book by Admati and Hellwig (2013). A good explanation of the concept of economic capital may be found in the relevant entry in the *Encyclopedia of Quantitative Finance* (Rosen and Saunders 2010).

2.2 Modelling Value and Value Change

We have seen in Section 2.1.1 that an analysis of the risks faced by a financial institution requires us to consider the change in the value of its assets and liabilities. In Section 2.2.1 we set up a formal framework for modelling value and value change and illustrate this framework with stylized asset and liability portfolios. With the help of these examples we take a closer look at valuation methods in Section 2.2.2. Finally, in Section 2.2.3 we discuss the different approaches that are used to construct loss distributions for portfolios over given time horizons.

2.2.1 Mapping Risks

In our general mathematical model for describing financial risks we represent the uncertainty about future states of the world by a probability space $(\Omega, \mathcal{F}, P)$, which is the domain of all random variables (rvs) we introduce below.

We consider a given portfolio of assets and, in some cases, liabilities. At the simplest level, this could be a collection of stocks or bonds, a book of derivatives or a collection of risky loans. More generally, it could be a portfolio of life insurance contracts (liabilities) backed by investments in securities such as bonds, or even a financial institution's overall balance sheet. We denote the *value* of the portfolio at time t by V_t and assume that the rv V_t is known, or can be determined from information available, at time t. Of course, the valuation of many positions on a financial firm's balance sheet is a challenging task; we return to this issue in more detail in Section 2.2.2.

We consider a given risk-management time horizon Δt, which might be one day or ten days in market risk, or one year in credit, insurance or enterprise-wide risk management. To develop a simple formalism for talking about value, value change and the role of risk factors, we will make two simplifying assumptions:

- the portfolio composition remains fixed over the time horizon; and
- there are no intermediate payments of income during the time period.

While these assumptions may hold approximately for a one-day or ten-day horizon, they are unlikely to hold over one year, where items in the portfolio may mature and be replaced by other investments and where dividend or interest income may accumulate. In specific situations it would be possible to relax these assumptions, e.g. by specifying simple rebalancing rules for portfolios or by taking intermediate income into account.

Using a time-series notation (with time recorded in multiples of the time horizon Δt) we write the value of the portfolio at the end of the time period as V_{t+1} and the change in value of the portfolio as $\Delta V_{t+1} = V_{t+1} - V_t$. We define the *loss* to be $L_{t+1} := -\Delta V_{t+1}$, which is natural for short time intervals. For longer time intervals, on the other hand, this definition neglects the time value of money, and an alternative would be to define the loss to be $V_t - V_{t+1}/(1 + r_{t,1})$, where $r_{t,1}$ is the simple risk-free interest rate that applies between times t and $t + 1$; this measures the loss in units of money at time t. The rv L_{t+1} is typically random from the viewpoint of time t, and its distribution is termed the *loss distribution*. Practitioners in risk management are often concerned with the so-called P&L distribution. This is the distribution of the change in portfolio value ΔV_{t+1}. In this text we will often focus on L_{t+1} as this simplifies the application of many statistical methods and is in keeping with conventions in actuarial risk theory.

The value V_t is typically modelled as a function of time and a d-dimensional random vector $\boldsymbol{Z}_t = (Z_{t,1}, \dots, Z_{t,d})'$ of *risk factors*, i.e. we have the representation

$$V_t = f(t, \boldsymbol{Z}_t) \tag{2.2}$$

for some measurable function $f: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. Risk factors are usually assumed to be observable, so the random vector $\boldsymbol{Z}_t$ takes some known realized value z_t at time t and the portfolio value V_t has realized value $f(t, z_t)$. The choice of the risk factors and of f is of course a modelling issue and depends on the portfolio at hand, on the data available and on the desired level of precision (see also Section 2.2.2). A representation of the portfolio value in the form (2.2) is termed a *mapping* of risks. Some examples of the mapping procedure are provided below.

We define the random vector of *risk-factor changes* over the time horizon to be $\boldsymbol{X}_{t+1} := \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t$. Assuming that the current time is t and using the mapping (2.2), the portfolio loss is given by

$$L_{t+1} = -(f(t+1, z_t + \boldsymbol{X}_{t+1}) - f(t, z_t)), \tag{2.3}$$

which shows that the loss distribution is determined by the distribution of the risk-factor change $\boldsymbol{X}_{t+1}$.

If f is differentiable, we may also use a first-order approximation L^{Δ}_{t+1} of the loss in (2.3) of the form

$$L^{\Delta}_{t+1} := -\biggl(f_t(t, z_t) + \sum_{i=1}^{d} f_{z_i}(t, z_t) X_{t+1,i}\biggr), \tag{2.4}$$

where the subscripts on f denote partial derivatives. The notation L^{Δ} stems from the standard *delta* terminology in the hedging of derivatives (see Example 2.2 below). The first-order approximation is convenient as it allows us to represent the loss as a *linear* function of the risk-factor changes. The quality of the approximation (2.4) is obviously best if the risk-factor changes are likely to be small (i.e. if we are measuring risk over a short horizon) and if the portfolio value is almost linear in the risk factors (i.e. if the function f has small second derivatives).

We now consider a number of examples from the areas of market, credit and insurance risk, illustrating how typical risk-management problems fit into this framework.

Example 2.1 (stock portfolio). Consider a fixed portfolio of d stocks and denote by λ_i the number of shares of stock i in the portfolio at time t. The price process of stock i is denoted by $(S_{t,i})_{t\in\mathbb{N}}$. Following standard practice in finance and risk management we use logarithmic prices as risk factors, i.e. we take $Z_{t,i} := \ln S_{t,i}$, $1 \leqslant i \leqslant d$, and we get $V_t = \sum_{i=1}^{d} \lambda_i \mathrm{e}^{Z_{t,i}}$. The risk-factor changes $X_{t+1,i} = \ln S_{t+1,i} - \ln S_{t,i}$ then correspond to the log-returns of the stocks in the portfolio. The portfolio loss from time t to $t+1$ is given by

$$L_{t+1} = -(V_{t+1} - V_t) = -\sum_{i=1}^{d} \lambda_i S_{t,i}(\mathrm{e}^{X_{t+1,i}} - 1),$$

and the linearized loss L^{Δ}_{t+1} is given by

$$L^{\Delta}_{t+1} = -\sum_{i=1}^{d} \lambda_i S_{t,i} X_{t+1,i} = -V_t \sum_{i=1}^{d} w_{t,i} X_{t+1,i}, \tag{2.5}$$

where the weight $w_{t,i} := (\lambda_i S_{t,i})/V_t$ gives the proportion of the portfolio value invested in stock i at time t. Given the mean vector and covariance matrix of the distribution of the risk-factor changes, it is very easy to compute the first two moments of the distribution of the linearized loss L^{Δ}. Suppose that the random vector $\boldsymbol{X}_{t+1}$ has a distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Using general rules for the mean and variance of linear combinations of a random vector (see also equations (6.7) and (6.8)), we immediately get

$$E(L^{\Delta}_{t+1}) = -V_t \boldsymbol{w}'\boldsymbol{\mu} \quad \text{and} \quad \operatorname{var}(L^{\Delta}_{t+1}) = V_t^2 \boldsymbol{w}' \Sigma \boldsymbol{w}. \tag{2.6}$$

Example 2.2 (European call option). We now consider a simple example of a portfolio of derivative securities: namely, a standard European call on a non-dividend-paying stock with maturity time T and exercise price K. We use the *Black–Scholes option-pricing formula* for the valuation of our portfolio. The value of a call option on a stock with price S at time t is given by

$$C^{\mathrm{BS}}(t, S; r, \sigma, K, T) := S\Phi(d_1) - K\mathrm{e}^{-r(T-t)}\Phi(d_2), \tag{2.7}$$

where Φ denotes the standard normal distribution function (df), r represents the continuously compounded risk-free interest rate, σ denotes the volatility of the underlying stock, and where

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}. \tag{2.8}$$

For notational simplicity we assume that the time to maturity of the option, $T - t$, is measured in units of the time horizon, and that the parameters r and σ are expressed in terms of those units; for example, if the time horizon is one day, then r and σ are the daily interest rate and volatility. This differs from standard market practice where time is measured in years and r and σ are expressed in annualized terms.

To map the portfolio at time t, let S_t denote the stock price at time t and let r_t and σ_t denote the values that a practitioner chooses to use at that time for the interest rate and volatility. The log-price of the stock ($\ln S_t$) is an obvious risk factor for changes in value of the portfolio. While in the Black–Scholes option-pricing model the interest rate and volatility are assumed to be constant, in real markets interest rates change constantly, as do the *implied volatilities* that practitioners tend to use as inputs for the volatility parameter. Hence, we take $\boldsymbol{Z}_t = (\ln S_t, r_t, \sigma_t)'$ as the vector of risk factors.

According to the Black–Scholes formula the value of the call option at time t equals $C^{\mathrm{BS}}(t, S_t; r_t, \sigma_t, K, T)$, which is of the form (2.2). The risk-factor changes are given by

$$\boldsymbol{X}_{t+1} = (\ln S_{t+1} - \ln S_t, r_{t+1} - r_t, \sigma_{t+1} - \sigma_t)',$$

and the linearized loss can be calculated to be

$$L^{\Delta}_{t+1} = -(C^{\mathrm{BS}}_t + C^{\mathrm{BS}}_S S_t X_{t+1,1} + C^{\mathrm{BS}}_r X_{t+1,2} + C^{\mathrm{BS}}_\sigma X_{t+1,3}), \tag{2.9}$$

where the subscripts denote partial derivatives of the Black–Scholes formula (2.7). Note that we have omitted the arguments of C^{BS} to simplify the notation. Note also that an S_t term appears because we take the equity risk factor to be the log-price of the stock rather than the price; applying the chain rule with $S = \mathrm{e}^{z_1}$ we have

$$C^{\mathrm{BS}}_{z_1} = C^{\mathrm{BS}}_S \left.\frac{\mathrm{d}S}{\mathrm{d}z_1}\right|_{z_1 = \ln S_t} = C^{\mathrm{BS}}_S S_t.$$

In Section 9.1.2 and Example 9.1 we give more detail concerning the derivation of mapping formulas similar to (2.9) and pay more attention to the choice of timescale in the mapping function.

The derivatives of the Black–Scholes option-pricing function are often referred to as the *Greeks*: C^{BS}_S (the partial derivative with respect to the stock price S) is called the *delta* of the option; C^{BS}_t (the partial derivative with respect to time) is called the *theta* of the option; C^{BS}_r (the partial derivative with respect to the interest rate r) is called the *rho* of the option; and, in a slight abuse of the Greek language, C^{BS}_σ (the partial derivative with respect to volatility σ) is called the *vega* of the option. The Greeks play an important role in the risk management of derivative portfolios.

The reader should keep in mind that for portfolios containing derivatives, the linearized loss can be a rather poor approximation of the true loss, since the portfolio value is often a highly nonlinear function of the risk factors. This has led to the development of higher-order approximations such as the *delta–gamma approximation*, where first- and second-order derivatives are used (see Section 9.1.2).

Example 2.3 (stylized loan portfolio). In this example we show how losses from a portfolio of short-term loans fit into our general framework; a detailed discussion of models for loan portfolios will be presented in Chapter 11.

Following standard practice in credit risk management, the risk-management horizon Δt is taken to be one year. We consider a portfolio of loans to m different borrowers or obligors that have been made at time t and valued using a book-value approach. To keep the example simple we assume that all loans have to be repaid at time $t+1$. We denote the amount to be repaid by obligor i by k_i; this term comprises the interest payment at $t+1$ and the repayment of the loan principal. The *exposure* to obligor i is defined to be the present value of the promised interest and principal cash flows, and it is therefore given by $e_i = k_i/(1+r_{t,1})$.

In order to take the possibility of default into account we introduce a series of random variables $(Y_{t,i})_{t\in\mathbb{N}}$ that represent the default state of obligor i at t, and we let $Y_{t,i} = 1$ if obligor i has defaulted by time t, with $Y_{t,i} = 0$ otherwise. These variables are known as *default indicators*. For simplicity we assume that all obligors are in a non-default state at time t, so $Y_{t,i} = 0$ for all $1 \leqslant i \leqslant m$.

In keeping with valuation conventions, in practice we define the book value of a loan to be the exposure of the loan reduced by the discounted expected loss due to default; in this way the valuation includes a provision for default risk. We assume that in the case of a default of borrower i, the lender can recover an amount $(1-\delta_i)k_i$ at the maturity date $t+1$, where $\delta_i \in (0, 1]$ describes the so-called *loss given default* of the loan, which is the percentage of the exposure that is lost in the event of default. Moreover, we denote by p_i the probability that obligor i defaults in the period $(t, t+1]$. In this introductory example we suppose that δ_i and p_i are known constants. In practice, p_i could be estimated using a credit scoring model (see Section 10.2 for more discussion). The discounted expected loss due to a default of obligor i is thus given by

$$\frac{1}{1+r_{t,1}}\delta_i p_i k_i = \delta_i p_i e_i.$$

The book value of loan i is therefore equal to $e_i(1-\delta_i p_i)$, the discounted expected pay-off of the loan. Note that in practice, one would make further provisions for administrative, refinancing and capital costs, but we ignore these issues for the sake of simplicity. Moreover, one should keep in mind that the book value is not an estimate for the fair value of the loan (an estimate for the amount for which the loan could be sold in a securitization deal); the latter is usually lower than the discounted expected pay-off of the loan, as investors demand a premium for bearing the default risk (see also our discussion of risk-neutral valuation in the next section). The book value of the loan portfolio at time t is thus given by

$$V_t = \sum_{i=1}^{m} e_i(1-\delta_i p_i).$$

The value of a loan to obligor i at the maturity date $t+1$ equals the size of the repayment and is therefore equal to k_i if $Y_{t+1,i} = 0$ (no default of obligor i) and equal to $(1-\delta_i)k_i$ if $Y_{t+1,i} = 1$ (default of obligor i). Hence, V_{t+1}, the value of the portfolio at time $t+1$, equals

$$V_{t+1} = \sum_{i=1}^{m}((1-Y_{t+1,i})k_i + Y_{t+1,i}(1-\delta_i)k_i) = \sum_{i=1}^{m} k_i(1-\delta_i Y_{t+1,i}).$$

Since we use a relatively long risk-management horizon of one year, it is natural to discount V_{t+1} in computing the portfolio loss. Again using the fact that $e_i = k_i/(1+r_{t,1})$, we obtain

$$\begin{aligned} L_{t+1} = V_t - \frac{V_{t+1}}{1+r_{t,1}} &= \sum_{i=1}^{m} e_i(1-\delta_i p_i) - \sum_{i=1}^{m} e_i(1-\delta_i Y_{t+1,i}) \\ &= \sum_{i=1}^{m} \delta_i e_i Y_{t+1,i} - \sum_{i=1}^{m} \delta_i e_i p_i, \end{aligned}$$

which gives a simple formula for the portfolio loss involving exposures, default probabilities, losses given default and default indicators.

Finally, we explain how this example fits into the mapping framework given by (2.2) and (2.3). In this case the risk factors are the default indicator variables $\mathbf{Z}_t = (Y_{t,1}, \dots, Y_{t,m})'$. If we write the mapping formula as

$$f(s, \mathbf{Z}_s) = \sum_{i=1}^{m}(1-Y_{s,i})\frac{k_i}{(1+r_{s,1})^{t+1-s}}(1-(t+1-s)\delta_i p_i) + \sum_{i=1}^{m} Y_{s,i}k_i(1-\delta_i),$$

we see that this gives the correct portfolio values at times $s = t$ and $s = t+1$. The issue of finding and calibrating a good model for the joint distribution of the risk-factor changes $\mathbf{Z}_{t+1} - \mathbf{Z}_t$ is taken up in Chapter 11.

Example 2.4 (insurance example). We consider a simple whole-life annuity product in which a policyholder (known as an annuitant) has purchased the right to receive a series of payments as long as he or she remains alive. Although realistic products would typically make monthly payments, we assume annual payments for simplicity and consider a risk-management horizon of one year.

At time t we assume that an insurer has a portfolio of n annuitants with current ages x_i, $i = 1, \dots, n$. Annuitant i receives a fixed annual annuity payment κ_i, and the time of their death is represented by the random variable τ_i. The annuity payments are made in arrears at times $t+1, t+2, \dots$, a form of product known as a whole-life immediate annuity.

At time t there is uncertainty about the value of the cash flow to any individual annuitant stemming from the uncertainty about their time of death. The liability due to a single annuitant takes the form

$$\sum_{h=1}^{\infty} I_{\{\tau_i > t+h\}}\kappa_i D(t, t+h),$$

where $D(t, t+h)$ is a discount factor that gives the time-t value of one unit paid out at time $t+h$. Following standard discrete-time actuarial practice we set $D(t, t+h) =$